BEHAVIORAL RESEARCH ON SUSTAINABLE AND SOCIALLY/ENVIRONMENTALLY RESPONSIBLE OPERATIONS

A Dissertation
Presented to
The Academic Faculty

By

Mahdi Mahmoudzadeh

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
Scheller College of Business

Georgia Institute of Technology

August 2020

BEHAVIORAL RESEARCH ON SUSTAINABLE AND SOCIALLY/ENVIRONMENTALLY RESPONSIBLE OPERATIONS

Approved by:

Dr. Peter Thompson
Committee Co-Chair
Scheller College of Business
Georgia Institute of Technology

Dr. Karthik Ramachandran
Committee Co-Chair
Scheller College of Business
Georgia Institute of Technology

Dr. Manpreet Hora
Scheller College of Business
Georgia Institute of Technology

Dr. Jeffrey Hales
McCombs School of Business
University of Texas at Austin

Dr. Enno Siemsen
Wisconsin School of Business
University of Wisconsin-Madison

Date Approved: June 08, 2020

TABLE OF CONTENTS

List of Tables vi

List of Figures ix

Summary x

Chapter 1: Introduction 1

Chapter 2: Framing Effects and the Non-Equivalence of Trade-ins and Upgrades 5

2.1 Introduction 5

2.2 Hypotheses 9

2.3 Experimental Design 14

2.3.1 Market Settings 16

2.3.2 Procedure 16

2.4 Results and Discussions 20

2.4.1 Framing Effects and Influential Anchors 21

2.4.2 A Reference-Dependence Model of Trade-ins and Upgrades 28

2.5 Concluding Remarks 37

Chapter 3: Structuring Supply Chains for Socially Responsible Behavior 39

3.1 Introduction 39

3.2 Literature Review 44

3.2.1 Sourcing Decisions in Supply Chains 44

3.2.2 Market Segmentation and Price Discrimination 45

3.2.3 Socially and Environmentally Responsible Operations 46

3.2.4 Behavioral Operations Management 46

3.3 Theory Development 47

3.3.1 Problem Setting 47

3.3.2 Normative Theory 48

3.3.3 Behavioral Theory 52

3.4 Behavioral Experiment 57

3.4.1 Experimental Procedure 59

3.4.2 Preliminary Results 61

3.4.3 Behavioral Predictions 63

3.4.4 Robustness Checks 64

3.5 Conclusion 69

Chapter 4: Charging for Recycling and Consumer Adoption of Green Electronics 72

4.1 Introduction 72

4.2 Hypotheses 77

4.2.1 Charging for Recycling and Adoption of Green Versions 77

4.2.2 Consumer Heterogeneity 80

4.3 Experimental Study 83

4.3.1 Design 83

4.3.2 Incentive Compatibility, Social Desirability, and Attention Check . . 85

4.3.3 Descriptions 88

4.3.4 Procedure 91

4.4 Results and Analyses 91

4.4.1 Analysis 1: Charging for Recycling and Adoption of the Green Version in Regular/New and Emergency/Replacement Purchase Situations 95

4.4.2 Analysis 2: Robustness of the Effect to Common Consumer Heterogeneity beyond the Purchase Situation 98

4.4.3 Analysis 3: Characterizing the Effect in Terms of Green Market Expansion and Green Market Intensification 102

4.5 General Discussion and Concluding Remarks 107

Appendix A: 111

Appendix B: 116

Appendix C: 124

References 127

LIST OF TABLES

2.1 Market settings and potential anchors. 17

2.2 Price intervals (in dollars). 19

2.3 Total number of responses in each experimental setting. 19

2.4 Influential anchors on the reference point with the alternative frames. 22

2.5 Influential anchors on the reference point with the alternative frames (effect of the existence/absence of an external secondary market price). 23

2.6 Influential anchors on the reference point with the alternative frames (effect of the manufacturer's (dis)continuing the sale of the original version). . . . 24

2.7 Effect of the secondary market price on anchoring with the upgrade frame. . 25

2.8 Effect of overlapping production on anchoring with the upgrade frame under a low innovation. 27

3.1 The buyer's profit with four possible sourcing policies. 49

3.2 The buyer's profit with four possible sourcing policies in the case of salvage value = 0. 50

3.3 Experimental groups. 58

3.4 Number of participants. 60

3.5 Observed order quantities and predictions of the classical model. 62

3.6 Parameter estimation of the behavioral model. 65

3.7 Observed order quantities and predictions of the behavioral model. 65

3.8 Robustness checks on the nonuniformity of the dual-sourcing bias and the market induced dual-sourcing. 67

3.9 Robustness check on the classification of the parameter β. 69

4.1 Experimental framework. 86

4.2 Mixed-model analysis. 93

4.3a Mixed-model analysis with charging for recycling scenario vs. other recycling scenarios. 94

4.3b Mixed-model analysis with recycling tax scenario vs. other recycling scenarios. 94

4.3c Mixed-model analysis with free recycling scenario vs. other recycling scenarios. 95

4.5 Mixed-model analysis including participants' demographics. 98

4.6 The observed effect in Analysis 1 considering participants' demographics in the analysis. 100

4.7 The effect of starting WTP on the adoption of the Green version. 101

4.8 Mixed-model analysis for participants with (extra WTP$_G$)%>0. 104

4.9 Mixed-model analysis including participants' demographics for participants with (extra WTP$_G$)%>0. 105

4.10 The observed effect in Analysis 1 considering participants' demographics for participants with (extra WTP$_G$)%>0. 106

4.11 The effect of starting WTP on the adoption of the Green version for participants with (extra WTP$_G$)%>0. 106

A1 Robustness check on using the same participants for both innovation level. . 112

A2 Regression (3) for when the original purchase price is the influential anchor. 113

A3 Regression (3) for when the new version's price is the influential anchor. . . 113

B1 Parameter ranges in the experiment. 122

LIST OF FIGURES

2.1 Market segmentation in the semi-anonymous case. 30

2.2 Manufacturer's optimal profit in the semi-anonymous case. 33

2.3 Manufacturer's optimal solution in the semi-anonymous case. 35

3.1 Nonresponsibly sourced products—observed versus predicted by the classical model. *Note: Error bars represent 95% confidence intervals.* 62

3.2 Nonresponsibly sourced products—observed versus predicted by the classical and behavioral models. *Note: Error bars represent 95% confidence intervals.* 66

4.1 Average of (extra WTP$_G$)% for the two purchase situations with the three recycling policies. *Note: Error bars are standard errors of means (SEM).* . . 96

4.2 The effect of charging for recycling on the average of (extra WTP$_G$)% for the two purchase situations. *Note: Error bars are standard errors of means (SEM).* 96

4.3 Portion of participants with (extra WTP$_G$)%¿0 for the two purchase situations with the three recycling policies. 103

4.4 Average of (extra WTP$_G$)% for the two purchase situations with the three recycling policies for participants with (extra WTP$_G$)%¿0. *Note: Error bars are standard errors of means (SEM).* 103

4.5 The effect of charging for recycling on the average of (extra WTP$_G$)% for the two purchase situations for participants with (extra WTP$_G$)%¿0. *Note: Error bars are standard errors of means (SEM).* 104

B1 Four possible sourcing policies. 116

SUMMARY

With the growth of global interest in sustainability and social responsibility, more companies and manufacturers have started practicing sustainable operations and social and environmental responsibility in their supply chains. Recent advancements in the field of behavioral economics have uncovered many relevant insights that can be of help in understanding interests and motives among different entities in supply chains including suppliers, manufacturers, policy makers, and customers. Utilizing both experimental and analytical methods, this dissertation's focus is to incorporate some of the relevant insights from behavioral economics into topics related to sustainable operations, circular economy, and social responsibility in supply chains. The first chapter looks at replacement purchases and buyback schemes by durable goods manufacturers. In contrast to the classical model and conventional wisdom that ignore the relevance of framing effects in difference schemes, this chapter explores the framing difference between trade-ins and upgrades and studies how relaxing the equivalence assumption modifies predictions of the classical model and provides predictions more in line with today's durable goods markets. The second chapter looks at social/environmental responsibility in supply chains and examines what type of consumer reactions—encouraging ones that highlight the value of responsible sourcing or discouraging ones that highlight the possibility of a consumer boycott—can lead supply chains towards more responsible sourcing. Our results enrich the normative model's insights and lead to a straightforward recommendation for NGOs that is also in line with what can be expected from consumers. This third chapter, motivated by Best Buy's recent recycling program, studies the potential of a charging for recycling program from a circular economy perspective. We find evidence that, in contrast to the long-standing practice of free recycling, charging for recycling can increase adoption of green electronics among consumers. This chapter suggests that current environmental laws that prohibit retailers from charging for recycling may be counterproductive to circular economy.

CHAPTER 1
INTRODUCTION

Over the past decades, more companies have become interested in pursuing sustainability and social responsibility in their operations and supply chains. Creating a discipline of sustainability is a task of aligning interests, policies, and behavior among suppliers, manufacturers, policy makers, and customers. Recent findings in behavioral economics can expand our knowledge and understanding of dynamics between different entities and help us obtain results more applicable to reality. Using both experimental and analytical methods, this dissertation aims at incorporating insights from behavioral economics into topics related to sustainability, circular economy, and social/environmental responsibility in supply chains. Behavioral experiments sharpen our assumptions in mathematical models and our discussions, while analytical methods allow us to capture these behavioral assumptions in mathematical models, perform analyses, and come up recommendations for practice. In three chapters, this dissertation studies three dimensions related to managing product take back in closed-loop supply chains, improving social/environmental responsibility in supply chains, and creating more efficient circular economy. The general theme of all three chapters concerns bringing a behavioral lens into the problem in order to obtain insights more in line with reality.

The first chapter looks at closed-loop supply chains and product take back by durable goods manufacturers. Manufacturers of durable goods often offer replacement purchases to customers to buy back older versions of their products from them and induce them to switch to improved versions. Besides expanding the market for new versions of their products, replacement purchases are useful mechanisms for manufacturers in managing the quantity, quality, and timing of their product returns. Recent market studies show that replacement purchases have become extensive for many durable goods (for example, more than 50% of digital camera sales, and more than 80% of iPhone sales are replacement purchases). Classical model and conventional wisdom have long ignored that the framing of these buyback schemes, whether through trade-ins or upgrades, can be relevant for consumer behavior and theory. We explore the framing difference between trade-ins and upgrades, and study how relaxing the equivalence assumption modifies predictions of the classical model. Our study presents a stylized way of capturing behavioral influences in trade-ins and upgrades to further incorporate in analytical models. Covering the common market settings, our study provides insights for manufacturers practicing trade-ins and upgrades under different market settings. Through controlled experiments, we capture the framing effect in trade-ins and upgrades in terms of a shift in customers' reference points. We then use the experimental findings to extend a reference-dependence version of the classical model of trade-ins and upgrades, to incorporate the framing difference between them. We find that the alternative frames are not isomorphic, and that the framing effect induces consumers to switch which prices they anchor to as their reference points when deciding a price for their current version. Trade-ins (respectively, upgrades) put customers in a selling (respectively, buying) position and result in anchoring to prices relevant to that. We also show that the behavioral extension overturns key predictions of the classical model and provides predictions more in line with reality. Our research highlights the importance of considering behavioral influences in modeling consumer decision-making in replacement purchases (i.e., trade-ins and upgrades) and that it can help obtain accurate outcomes from analytical models more in line with reality.

Taking the responsibility of their used products and returning them back into reuse or recycling is a valuable step for manufacturers towards more responsibility in their supply chains. Meanwhile, social/environmental responsibility in supply chains can go above and beyond this by aiming to ensure responsibility from the starting point of supply chains. Numerous examples of labor and building code violations in global supply chains have made it clear that nonresponsible sourcing is an ongoing issue in globally expanded supply chains.

The second chapter of this dissertation looks at socially/environmentally responsible sourcing in global supply chains. This chapter examines what type of consumer reactions—encouraging ones that emphasize the value of responsible sourcing or discouraging ones that emphasize the possibility of a consumer boycott in the face of violations—can lead supply chain partners to engage in more responsible sourcing. This question is important for NGOs promoting socially responsible behavior among consumers to combat child labor and labor code violations in global supply chains. The normative model in this context predicts that only discouraging consumer reactions reliably incentivize firms to source responsibly, as long as the salvage value of the product is not too low. Our analysis provides a more comprehensive perspective. We develop a behavioral model that incorporates the dual sourcing bias in firms' responsible sourcing decisions. Including this bias alters some of the normative model's predictions and results in more applicable insights. We validate predictions of our behavioral model against those of the classical model in an experimental study. Our behavioral model predicts that an encouraging reaction, irrespective of its magnitude, always increases the firm's responsible sourcing when the discouraging reaction from the market is weak, and further refines that a discouraging reaction always increases responsible sourcing irrespective of the type of product considered. Our results lead to a straightforward recommendation for NGOs to consider in different industries and supply chains: in supply chains of products with no (respectively, many) substitutes, the encouraging (respectively, discouraging) reaction is the most effective reaction to reduce nonresponsible sourcing—a recommendation that is also in line with what one can expect from customers based on the nature of the products.

Managing product return streams helps with the first step in circular economy, that is, collecting end-of-life products and avoiding landfills. Yet, companies and recyclers face serious challenges, e.g., not being able to recycle collected products in a cost-effective way, in taking next steps towards full circular economy. The high operational costs of recycling, especially for e-waste, have kept recycling rates far below an ideal point. Overcoming this challenge to create more efficient circular economy is, as emphasized by both academics and practitioners, contingent on more recyclable products entering the economy in the first place. Evidence has shown, however, that direct green-marketing to increase consumer adoption of such products has not been a successful strategy, and as a result, retailers usually do not find investing in direct green marketing financially attractive. Motivated by Best Buy's recent decision to charge a flat fee to accept some used consumer-electronics for recycling, the third chapter of this dissertations studies the potential of this counterintuitive recycling program from a circular economy perspective. Specifically, we postulate that charging for recycling, in contrast to the long-standing practice of free recycling, can pique consumers' attention to the fact that recycling is costly, and that recyclability is a valuable attribute in the product. Building on two behavioral insights, *nudging* and the theory of planned behavior, through experimental studies we explore whether charging for recycling can increase consumer adoption of green electronics, compared with current recycling policies as the baseline. Our results show that charging for recycling, besides helping with operational costs of recycling, can in fact increase adoption of green electronics by many customers in regular purchase situations. Our findings have clear implications for green marketing, recycling operations, and environmental policy. In particular, our findings expand previous research on tying recycling operations and green marketing and contradict the conventional wisdom that favors free recycling as the beneficial bridge to coordinate the two. Our findings also suggest that, from a bigger picture to circular economy, current environmental laws that prohibit retailers from charging for recycling may be counterproductive.

CHAPTER 2

FRAMING EFFECTS AND THE NON-EQUIVALENCE OF TRADE-INS AND UPGRADES

2.1 Introduction

In 2004, 82% of computer sales, 63% of cell phone sales, and 55% of digital camera sales were replacement purchases (Gordon 2009). Moreover, 77% of iPhone 4 buyers and 83% of iPhone 5 buyers upgraded from previous iPhone versions (Statisticbrain 2016). These high rates of replacement purchases are mainly due to market saturation, and product introduction cycles that are shorter than the useful life of durable goods. To expand the market for newer versions of their products, manufacturers of durable goods have increasingly managed replacement purchases separately from their sales channels for first-time buyers and engaged in price discrimination between repeat and new customers, typically by buying back older versions of their products to offer price reductions to repeat customers (Golson 2016, Kohler 2016). Indeed, Ray et al. (2005) showed that by offering different prices for repeat and new customers, manufacturers could increase their profit by up to 40%. Replacement purchases have also become useful mechanisms for manufacturers in managing their closed-loop operations through controlling their product returns. The success of any closed-loop system depends on the efficiency of its product return stream (Li et al. 2011, Tang and Zhou 2012, Dutta et al. 2016), and replacement purchases can particularly help manufacturers manage the quantity, quality, and timing of their product returns (Guide and Wassenhove 2001, Souza 2013).

Trade-ins and upgrades are the two common mechanisms for replacement purchases. With a trade-in, replacement buyers are quoted a price for their current model when buying its newer version at its regular sale price. With an upgrade, replacement buyers are offered the new version at a discounted price upon giving their current model to the dealer. In standard economic analysis, the two mechanisms are isomorphic when the trade-in value on the old model is equal to the discount value on the upgrade, and so they have often been used interchangeably in the long-standing literature of trade-ins and upgrades (see, e.g., Levinthal and Purohit 1989, Van Ackere and Reyniers 1993, 1995, Fudenberg and Tirole 1998, Ray et al. 2005, Rao et al. 2009, Zhao and Jagpal 2009, Li et al. 2011, Agrawal et al. 2016).

There is ample evidence, however, that customers' decision making in replacement purchases is driven mainly by behavioral influences (see Guiltinan 2010, for a good review). In particular, the non-equivalence of trade-ins and upgrades can no longer be expected to hold once one considers the framing difference between them: trade-ins separate the selling transaction of the replacement purchase explicitly, while upgrades embeds that in the buying transaction. This framing difference resembles some of the seminal examples of the framing effect provided by Tversky and Kahneman (1981, 1986), and is particularly important to study since observations have shown that, although a replacement purchase comprises two simultaneous selling-buying transactions for customers (Purohit 1995, Okada 2001), customers are more sensitive to the selling transaction than to the buying transaction when making a simultaneous selling-buying (Purohit 1995, Hoyer et al. 2002, Zhu et al. 2008, Kim et al. 2011), which is due to the behavioral anomalies in selling/giving-up a product (Guiltinan 2010).

The framing effect, i.e., influencing people's decisions through alternative representations of decision problems, has been established in many different contexts, from financial investments (see, e.g., Kumar and Lim 2008, Cheng and Chiou 2008) to health decisions (Gallagher and Updegraff 2012), and in both laboratory and field experiments (Hossain and List 2012), and it has been shown that even experimental economists are prone to it (Gächter et al. 2009). In this paper, we study the framing effect in trade-ins and upgrades, and postulate that it would result in different evaluations of the selling transaction in a replacement purchase. We use shifts in the reference point to capture this framing effect and the induced difference in evaluating the same transaction. The reference point is the hinge of evaluating any transaction (Tversky and Kahneman 1991), and the reference-point shift has been established as a proper mechanism to capture the framing effect (see, e.g., Heath et al. 1995, Lehner 2000, Hossain and List 2012). The main logic is that two alternative frames would induce different utilities if and only if they create different reference points, through which they would result in different evaluations of the same outcome (see, for instance, Hossain and List 2012, for a reference-point shift between base-salary and enhanced-salary in reward *vs* penalty salary frames). Our first research question is thus whether trade-ins and upgrades are different in terms of customers' reference points for the price for their current product.[1]

Similar to reference-point studies with multiple price-anchors (see, e.g., Baucells et al. 2011), through an experimental study (via Amazon Mechanical Turk)[2], we extract the influential price-anchors, among the set of all potential price-anchors, on the reference point with both the trade-in and the upgrade frames. Understanding what influences replacement buyers' reference points, and how it shifts with the alternative frames, reveals the leverage and the proper frame that manufacturers can use to manage their customers' reference points and thus their replacement decision. The straightforward application of this finding would be in developing reference-dependence models of trade-ins and upgrades to investigate how the framing difference between them would matter for manufacturers' policies. Our next question is thus on exploring this through a reference-dependence extension of the classical model of trade-ins and upgrades and revisiting the predictions of the classical model in this matter.

[1]It is noted that unsystematic approaches, such as comparison of willingness-to-accepts with the alternative frames without a knowledge on the reference point, may lead to misleading results. For instance, for a given pricing in the market, it is possible to observe comparable levels of willingness-to-accept with the two frames, while overlooking the very possibility of customers' use of different reference points (e.g., the secondary market price or the new version's price) to arrive at that willingness-to-accept.

[2]Amazon Mechanical Turk (AMT) has recently become a popular avenue for subject recruitment, and several studies have endorsed AMT as a reliable source of experimental data (see Lee et al. 2018 and references therein).

We find that since the upgrade frame embeds the selling transaction in a net buying transaction, it directs customers' attention toward price anchors relevant to the buying position (i.e., the manufacturer's sale prices), while the trade-in frame places customers in an explicit selling position and results in anchoring to prices relevant to that (i.e., the secondary market price). Therefore, the reference point shifts with the trade-in and the upgrade frame as a result of anchoring to different prices regarding the induced positions with them. In addition, different market settings influence the reference point with a given frame if they affect the set of price anchors relevant to the induced position.

Furthermore, through a behavioral extension of the classical model of trade-ins and upgrades by Fudenberg and Tirole (1998), we find that their two key predictions, that under a high innovation level in the new version, the manufacturer will not continue producing the original version, and that the manufacturer is better off in an anonymous market than in a semi-anonymous market, do not hold anymore. Their second prediction, laid out through a comparison of trade-ins (for the anonymous case) and upgrades (for the semi-anonymous case), is a result of their first prediction, which is, of course, at odds with the numerous instances we see nowadays wherein manufacturers continue production and sale of older versions even under a high innovation level in the new version. For instance, Apple continued production of the iPhone 4s until September 2014, a full year after the iPhone 5s and 5c were introduced to the market. Similarly, the iPhone 5s remained available until March 2016, over a year after the iPhone 6 and 6 plus were introduced to the market.[3] Market segmentation and technology adoption are the common explanations come to mind with overlapping production. However, the aforementioned overlapping productions exist even though, driven by Apple's very large R&D investments (Appleinsider 2016), new iPhone series are produced with fairly the same production costs yet obtaining high consumer valuations and sale prices (Time 2014, 2016). This falls under the high innovative category by

[3]Sony continued production of the PS3 till September 2015, over a year after the PS4 was released. The PS3 was not compatible with the PS4, which came with major (and unique) innovations, such as integrated social gaming services.

Fudenberg and Tirole (1998) (see Section 2.4.2) and rules out those common rational explanations. Our finding uncovers an explanation, which has been overlooked for these cases, that manufacturers' another objective in continuing production and sale of older versions is to manage the upgraders' reference points through the sale price of the older version. This explanation, which is salient in nowadays durable-goods markets with extensively grown segment of replacement buyers, can only be obtained through the behavioral extension.

The remainder of the paper is organized as follows: Section 2.2 covers the related literature and the research hypotheses; Section 2.3 describes the experimental design; results and discussions are presented in Section 2.4, along with a comparison between predictions of the classical model and the behavioral model in which the framing effect matters; and finally, Section 2.5 concludes the paper.

2.2 Hypotheses

Since its introduction by Tversky and Kahneman (1981), the framing effect has been applied in a variety of contexts to manage people's decisions, e.g., pricing contracts (Ho and Zhang 2008), investment decisions (Kumar and Lim 2008), health decisions (Gallagher and Updegraff 2012), and etc. Related to our context, Monga and Zhu (2005), and Yang et al. (2013) found that buyers and sellers responded differently to the same frame. In this vein, Simonson and Drolet (2004) showed that different anchors had different effects on buyers' and sellers' decisions. Here, reasoning that trade-ins and upgrades differ in framing the selling transaction for replacement buyers, through which they induce different (selling *vs* buying) positions for them, and based on the common reference-point shift mechanism (see, e.g., Heath et al. 1995, Lehner 2000), we postulate that these alternative frames result in customers' anchoring to different prices as their reference points for a price for their current product. As a result, we expect that replacement buyers' reference points will shift from prices relevant to the selling position (with a trade-in frame) to prices relevant to the buying position (with an upgrade frame).

To narrow down this general expectation, we fit our problem setting with that of the classical model of trade-ins and upgrades by Fudenberg and Tirole (1998), which has been used quite often in the literature: A manufacturer (she) produces two successive versions of a product over two periods. In period one, only the original version is produced and sold. In period two, the manufacturer introduces the new version with a high or low innovation, and offers trade-ins or upgrades to her former customers (i.e., replacement buyers) in addition to selling to new buyers. Upon introducing the new version in period two, the manufacturer may or may not decide to continue production and sale of the original version. This setting creates four potential price-anchors for replacement buyers at the replacement purchase point: the price they have paid in period one to buy the original version; its current sale price; its secondary market price; and the new version's sale price. The secondary market price is the only price anchor relevant to the selling position, while the other three, coming from the manufacturer's sale channel, are relevant to the buying position, among which the new version's price is the one most directly relatable from the buying position. Therefore, building on our general expectation, we expect that as their reference points, with trade-ins customers will anchor to the secondary market price, and with upgrades they will anchor to the new version's price. Furthermore, with the separation mentioned before, we expect that a change in the market setting will influence the reference point only if it is changing the set of price anchors relevant with the frame (trade-in or upgrade). We thus present the following Hypothesis on a comparison between trade-ins and upgrades:

HYPOTHESIS 1. *With a trade-in (upgrade), customers anchor to prices relevant to the selling (buying) position as reference prices for their current product.*

With the problem setting of Fudenberg and Tirole (1998) as laid out above, changes in three dimensions can impact the dynamics of the problem setting: the innovation level in the new version (which determines whether the two successive versions, and their prices, are comparable to each other or not); the existence or absence of an external secondary market (which determines the existence/absence of a potential price-anchor for replacement buyers); and whether the manufacturer continues production and sale of the original version alongside the new version (which determines the existence/absence of another potential price-anchor). The latter two are changing the structure of the market as they add/remove a potential price-anchor to/from the setting. The first dimension, on the other hand, alters an existing setting through shifting an existing price-anchor along a dimension. Nonetheless, we refer to all three dimensions as creators of different market settings. We next examine the effect of each dimension in detail, as they change the market setting, on the reference point with each frame.

Exploring the effect of the first dimension, i.e., the new version's innovation level, on customers' reference points directly relates to previous research on manufacturers' new product introductions. Most of the research in this area is in line with our setting: a two-period model with discrete product quality/innovation levels (see, e.g., Waldman 1996, Kornish 2001). The need for studying novel directions in new product introductions was first highlighted by Waldman (2003), and there have been extensions since then. Most relatable to our paper, via an experimental study Okada (2006) found that there was a significant difference in repeat buyers' interest in a new product, relative to new buyers' interest, when the new product was similar to their current product and when it was dissimilar to their current product. This relative difference comes from a change in how repeat buyers' close their mental account for their current product (Thaler 1980, 1985) influenced by the perceived (dis)similarity of the new product to their current product, which is rooted in the general theory of similarity by Tversky (1977). Here, in our studying replacement purchases, we relate this (closing the mental account for the current version) to a shift in replacement buyers' reference points for a price for their current version, driven by the new version's innovation level and its price (as the representative of (dis)similarity). Thus, expanding on Hypothesis 1, we expect that customers' anchoring to the new version's price as their reference point with the upgrade frame is the case as long as the new version is not a substantial improvement over the original version, that is, only when it is similar and its price is comparable to customers' current version. If the new version is high innovative and thus cannot provide a comparable point of reference, customers will refer to their current version and its current sale price, as the next most relatable price-anchor from the buying position in period two. In addition, based on Hypothesis 1, we expect that this shift in the reference point will only happen with the upgrade frame, not with the trade-in frame, as it is driven by a change in price anchors relevant to the buying position, not the selling position.

HYPOTHESIS 2. *The innovation level of the new version shifts customers' reference points with the upgrade frame, while it will not change the reference point with the trade-in frame.*

The second dimension, the existence or absence of external secondary markets, is usually out of manufacturers' direct control, but can be manipulated by indirect strategies (see Hendel and Lizzeri 1999, for a good discussion on this). Studying the latent benefits or harms for manufacturers in doing so is beyond the scope of this paper (on this topic, see Benjamin and Kormendi 1974, Liebowitz 1985, Rust 1986, Levinthal and Purohit 1989, Oraiopoulos et al. 2012). What we explore here is how the existence/absence of an external secondary market will influence customers' reference points with the two replacement purchase frames. From what has been discussed so far, and in particular based on Hypothesis 1, we expect that it will have no effect on customers' reference points with the upgrade frame, as it does not change the price anchors relevant to the buying position; however, it will change the reference point with the trade-in frame, wherein the secondary market price is the only price anchor directly relevant to customers' selling position. For the latter, with the absence of the external secondary market price, we postulate that customers would anchor to a price that can provide them an approximate of what their current product would sell for, that is, the current price of their product. This can be the case, however, only when the manufacturer continues production and sale of the original version. With the otherwise, none of the remaining price anchors can provide customers any direct (or indirect) information on the sale potential of their current product, and hence we expect no clear anchoring to any of the remaining prices. We present these predictions in the following Hypothesis:

HYPOTHESIS 3. *The existence/absence of an external secondary market shifts customers' reference points with the trade-in frame, while it will not change the reference point with the upgrade frame.*

The remaining dimension, whether the manufacturer continues production and sale of the original version after introducing the new version, is particularly important with the upgrade frame under a high innovation level in the new version, wherein we expected that customers would anchor to the current sale price of their product as their reference points (see Hypothesis 2). In this case, the absence of the current sale price of their product leaves customers with the price they have paid for it, as the only price anchor relevant to the buying position, and hence we expect customers' anchoring to that. When the new version is a low innovation, on the other hand, we do not expect that changes in this dimension would affect the reference point as customers are anchoring to the new version's price (see Hypothesis 2). Finally, based on Hypothesis 1, we expect that this shift in the reference point, driven by a change in price anchors relevant to the buying position, not the selling position, will not happen with the trade-in frame, wherein customers anchor to the secondary market price. With the trade-in frame, it only matters in market settings without an external secondary market, wherein, as expected in Hypothesis 3, customers are likely to use the current sale price of their product as an approximate of what their product would sell for. Similar to what expected in Hypothesis 3, in these market settings, with the absence of the current sale price of their product, we predict no clear anchoring to any of the remaining prices. We present these predictions in the following Hypothesis:

HYPOTHESIS 4. *The manufacturer's sale of the original version at the replacement purchase point will shift customers' reference points with the upgrade frame only when the new version is high innovative, while it will not in uence the reference point with the trade-in frame.*

2.3 Experimental Design

To increase external validity of our results, we match the experimental setting with a real-life situation: trading in or upgrading an electronic tablet. In addition, in order to capture the time point of decision making in replacement purchases, we need to use an imaginary situation: currently using a tablet and thinking about replacing it with its newer version. This is a common approach in previous experimental studies on replacement purchases (see, e.g., Purohit 1995, Zhu et al. 2008, Kim et al. 2011, Srivastava and Chakravarti 2011), and has been used in other decision-making problems, as well (see, e.g., Baucells et al. 2011 for reference points in selling stocks, and Hardisty and Pfeffer 2017 for choosing between present and future payoffs).[4] Since we study reference points, we are looking for the price that would make participants 'indifferent' about making the replacement purchase transaction. Our approach is similar to that of Baucells et al. (2011) in their extraction of reference points in selling a stock, where they asked participants for the price that would make them "neither happy nor unhappy about the sale." As Baucells et al. (2011) noted, "Because of the pure psychological nature of the reference point, no 'incentive-compatible' variable payment could be used." in these settings. Yet, since our experimental situation concerns a decision about a physical product, to make sure about eliminating potential attachment effects (and the price inflation it may induce), we apply the approach successfully implemented by Purohit (1995), Zhu et al. (2008), and Srivastava and Chakravarti (2011), and define the experimental situation for a third-party. That is, we describe a third-party situation and ask participants about the price for that person's current product that would make him "indifferent" about the replacement purchase.

We start with the experiment for the trade-in frame. To reach to the reference-point value in the selling transaction, first we need to keep the buying transaction's utility (i.e., the difference between participants' willingness-to-pay to the new version and its consump-

[4]There is also ample evidence that respondents' decisions in imaginary and real situations do not differ significantly (see, e.g., Carmon and Ariely 2000, and references therein).

tion utility) zero. The otherwise would interfere with the procedure of reference-point elicitation. Thus, we provide the participants two purchase options that our third-party 'Jack' is offered: a pure purchase and a trade-in. The difference between them is our variable of interest. With this setting, the price that participants think of making Jack indifferent between the two options is the price they think would bring him zero utility in selling his product, i.e., the reference-point value in the selling transaction. The experimental task is described to participants as follows (the prices shown here are just an example):

Jack has a fully functioning Tablet 4.2 (a brief description is provided below), which he had purchased for $190. *A Tablet 4.2 is currently sold at* $165. *The average price of a used Tablet 4.2 of the same condition as Jack's is* $60.
[Tablet 4.2 and Tablet 5.0 specifications were shown here, and are available in Appendix A]
Jack is thinking about buying a Tablet 5.0, and there are two purchase options available to him:
Option 1) Store A sells the Tablet 5.0 for $270.
Option 2) The same store has a 'trade-in' program. That is, the store will pay Jack cash for his current Tablet if he gives it to that store when purchasing the Tablet 5.0 at the same price as in Option 1 ($270).
How much cash do you think Jack should receive to be indifferent between these two options?

To study the framing effect, we change the experimental question to the 'upgrade' frame as follows:

Option 1) Store A sells the Tablet 5.0 for $270.
Option 2) The same store has an 'upgrade' program. That is, the store will sell Jack the Tablet 5.0 at a lower price (p) if he gives his current Tablet to that store.
At what price p do you think Jack would be indifferent between these two options?

The reference-point value in the selling transaction here is calculated by subtracting the indicated lower price (p) from the new version's price. The objective of our experiment is to extract which one of the price anchors participants would refer to (as their reference points) in deciding their indicated values under different market settings. Having enough data points for each market setting, we can do so through investigating changes in which price anchor has the most influence on the change in the reference-point value (see Section 2.4).

2.3.1 Market Settings

With the problem setting we study, the original version's purchase price and the new version's price are potential price-anchors that are always available regardless of the market setting. In contrast, the current sale price of the original version and the secondary market price may or may not exist (depending on the manufacturer's decision on continuing the production and sale of the original version at the replacement purchase point, and existence/absence of an external secondary market, respectively). These create four possible settings with different sets of potentially influential price-anchors (as shown in Table 2.1). The additional dimension, i.e., the discrete innovation level of the new version (low or high), creates two versions of each of the four settings. For them, we consider two new versions with distinct low and high innovation levels and sale prices: Tablet 5.0 and Tablet 5.2, with the improved features and the sale price relatively high for the latter. Therefore, in total we have eight market settings coming from a $2 \times 2 \times 2$ design. We adjust the experimental question for each setting by removing the part about the current sale of the original version, removing the part about the secondary market, and offering Tablet 5.0 or Tablet 5.2 as the new version.

2.3.2 Procedure

We recruited 1,195 participants (45.2% female; M_{age}=36.46, SD=11.62) in Amazon Mechanical Turk (AMT) and paid a flat fee ($0.60) for their participation. Participation was

Table 2.1: Market settings and potential anchors.

Setting	Potential anchors
With sale of the original version With external secondary market	Original purchase price Average secondary marlet price Current sale price of the original version New version's price
With sale of the original version Without external secondary market	Original purchase price Current sale price of the original version New version's price
Without sale of the original version With external secondary market	Original purchase price Average secondary marlet price New version's price
Without sale of the original version Without external secondary market	Original purchase price New version's price

Sale of the original version refers to the manufacturer's continuing production and sale of the original version in period two.

restricted by location (the United States only) and acceptance rate (above 97% with more than 5000 hits). The main advantage to using AMT is to access at reasonable cost a large population of average people, and thereby to obtain responses likely to be closer to typical market behavior. The majority of our sample had education beyond high school: 31.2% had some college credit, while another 45.2%, reported having earned a Bachelor's degree. Nonetheless, the average annual income was in the range $25,000-$49,999.

After agreeing to participate in the study, participants were referred to an external platform (Qualtrics) for the experimental tasks. Following some demographic questions, participants were randomly assigned to either the trade-in or the upgrade frame under one of the four market settings (created regarding *sale of the original version* $\times$ *external secondary market*, as shown in Table 1). Each participant answered two questions in random order, one for the low innovation level and one for the high innovation level, with different price sets.[5] In indicating their responses, each participant sees a different set of prices randomly drawn from the predefined intervals given in Table 2.2. The price intervals are

[5]The ordering was random and counterbalanced. Nevertheless, to ensure the reliability of the results from using the same participants for both innovation levels, we performed a robustness check, by considering only the first question that the participants answered, and obtained the same results (see Appendix A).

chosen in a way that all possible ratios between the four potential prices anchors are realistic, based on real market products and prices. For instance, the average secondary market price is kept between 14% and 53% of the original purchase price and between 16% and 67% of the current sale price. Furthermore, the price range for the new version under the high innovation is set to create relative prices big enough to signal the high innovation nature of the new version. In contrast, the price range under the low innovation makes sure of having prices close enough to that of the original version to resemble a low innovation in the new version.

To help with the regressions we run for each experimental group (see Section 2.4), while keeping the assignment process random, we assign more participants to the groups representing market settings with more potentially influential price-anchors (e.g., more participants to the experimental group representing the first setting in Table 2.1 than to the one representing the fourth setting). This helps make sure that we have enough data points for the regressions given the number of the independent variables in each experimental group. Table 2.3 shows the total number of responses in each experimental group (we have relatively more data in the first market setting in the interest of performing some robustness checks; see Appendix A). We ensure the reliability of responses by applying filtering criteria including an attention-check question at the end of the experimental task, filtering obviously false data, and ruling out outliers using Tukey's outlier criterion, i.e., responses below $Q_1 - 1.5 \times [Q_3 - Q_1]$ and above $Q_3 + 1.5 \times [Q_3 - Q_1]$, with Q_1 and Q_3 being the first and the third quantiles, respectively (Seo 2006). However, unless we have a strong reason to filter false data, such as participants' misunderstanding or responses of 0 because participants were not able to enter a valid number, we leave the data cleaning to the outlier criterion (e.g., we do not exclude any data because of the participant's response time). The final data loss is just under 19.8 percent, together with the attention-check failures.[6] Similar

[6]It is worth noting that our attention-check filtering is merely to ensure if participants paid attention to the descriptions, similar to pre-filtering of participants in lab experiments with test rounds. However, we report the whole collected data since our attention-check question was set at the end of the experiment after participants took the main experimental task. Pre-exclusion of these participants makes a data loss of just

to reference-point elicitation by Baucells et al. (2011), followed by the main experimental task, participants were also asked directly how important, on a seven-point Likert scale, each of the prices were in arriving at their response. Answers to this question were mainly confirmatory with no new insights, and hence we do not include them here. Therefore, our analysis, similar to that of Baucells et al. (2011), is only based on the regression results. We also asked participants about the clarity of the explained situations, on a 1-5 scale (5 being 'very clear'), and the average scale value was 4.2.

Table 2.2: Price intervals (in dollars).

Price	Interval*
Original purchase price	[170-230]
Average secondary marlet price	Original purchase price - [110-145]
Current sale price of the original version	Original purchase price - [15-50]
New version's price with a low innovation	Original purchase price + [50-105]
New version's price with a high innovation	Original purchase price + [210-265]

*All random numbers are generated in increments of 5.

Table 2.3: Total number of responses in each experimental setting.

Setting	Trade-in frame		Upgrade frame	
	Low innovation	High innovation	Low innovation	High innovation
With sale of the original version With external secondary market	240	240	240	240
With sale of the original version Without external secondary market	131	131	131	131
Without sale of the original version With external secondary market	133	133	129	129
Without sale of the original version Without external secondary market	96	96	95	95

under 4.7 percent through filtering the outlier data. Our main results are robust to including the outlier data in the sample; however, to ensure quantitative reliability of the results, we exclude them from the analyses.

2.4 Results and Discussions

Referring to Baucells et al.'s (2011) regression methodology in extracting the price anchors with statistically significant influence on the reference point, we use a regression model with all potentially influential price-anchors as independent variables, and the reference point as the dependent variable. We use Hierarchical Regression (HR), which, among a set of independent variables finds the ones with significant influence on the dependent variable, to identify the price anchor(s) with significant influence on the reference-point value. Regression models for the trade-in and the upgrade frame are as follows, respectively:

$$ref_i^T = \alpha + \alpha_o Opp_i + \alpha_s Sec_i + \alpha_c Curr_i + \alpha_n New_i + \varepsilon_i \tag{1}$$

$$ref_j^U = New_j - p_j = \beta + \beta_o Opp_j + \beta_s Sec_j + \beta_c Curr_j + \beta_n New_j + \varepsilon_j \tag{2}$$

The ref_i^T and ref_i^U are the extracted reference-point values from participants i and j under the trade-in and the upgrade frame, respectively. The independent variables Opp, Sec, $Curr$ and New indicate the original purchase price, the average secondary market, the current sale price of the original version, and the new version's price, respectively. Because of the nature of the dependent variable, the intercept will not have a meaningful interpretation (that is, when all prices are zero in the market, i.e., everything is free, customers' reference points is zero as well) (see Baucells et al.'s 2011 note on the intercept in their regression model for reference points). It helps, however, extract the influential anchors more reliably and is particularly important when the intercept is the only significant part of the regression's outcome, which would mean none of the anchors explains the reference point properly. Thus, we start running Hierarchical Regressions with an intercept, and if only the intercept is significant we conclude that no anchor explains the reference point. If the intercept is significant along with one or more prices, we take those price(s) to be the influential one(s). If the intercept is not significant, in line with the logic of Hierarchical Regression, we remove it from the regression model and take the outcome price(s) to be the influential one(s). Finally, if the regression yields more than one significant price, we statistically compare their coefficients (using a z-test) to find the most influential price-anchor. As the comparison will be between coefficients of different price-anchors, we will use standardized coefficients to take into account the standard deviations as well.

In running the HR for each group, only the independent variables corresponding to the potential anchors in that group (as shown in Table 2.1) are kept in the regression model. If two groups turn out to have different influential anchors, we conclude that they differ in terms of the reference point. If, on the other hand, two or more groups have the same influential anchor(s), we statistically compare the influential anchors' coefficients across them to conclude their difference or equality in terms of the reference point. In the between-group comparisons, we limit the regressions only to the extracted most influential anchor(s) since only those anchors would appear in the manufacturer's reference-dependence model. We use a z-test to statistically compare coefficients' magnitudes of the same price anchor across different regressions (see Clogg et al. 1995).

2.4.1 Framing Effects and Influential Anchors

If framing matters, all else being the same, we would see different influential anchors with the trade-in and the upgrade frame. Table 2.4 shows that this is in fact the case. As predicted by Hypothesis 1, with the trade-in frame, participants anchor to the secondary market price, and with the upgrade frame, participants anchor to the new version's price, as their reference points. In addition, as Hypothesis 2 predicted, the reference point with the upgrade frame shifts with the new version's innovation level: when it is low innovative and its price can provide a comparable reference-point for the current version, participants anchor to the new versions price; in the high innovation case, however, the new version's price is not a comparable anchor, and hence participants anchor to the current sale price of the current version, as the next price-anchor relevant to the buying position. It is also noted that in line with our predictions in Hypothesis 1, regarding the influence of the three dimensions with the alternative frames, the new version's innovation level does not influence the reference point with the trade-in frame as it does not influence the price anchor relevant to the selling position.

Table 2.4: Influential anchors on the reference point with the alternative frames.

Setting	Trade-in frame		Upgrade frame	
	Low innovation	High innovation	Low innovation	High innovation
With sale of the original version With external secondary market	Sec (< 0.001) 0.90	Sec (< 0.001) 0.95	New (< 0.001) 0.27	Curr (< 0.001) 0.60

Parentheses contain p-values. In all cells, HR yields only the indicated price with a significant influence on the reference point.
Numbers in third rows are the coefficients used in between-group comparisons as explained in the text.

Table 2.5 shows supporting results for Hypothesis 3 through illustrating the effect of the second dimension, i.e., existence/absence of the external secondary market, on the reference point with the two frames. As predicted, with the trade-in frame, absence of the secondary market price, removes the dominant (and the only obviously relevant) price-anchor to the selling position. Thus, participants anchor to the price that, to some extent, can provide a fair estimate of what the current model would sell for, that is, its current sale price. When this price is also removed (see the two bottom rows of Table 2.5), none of the remaining prices provides a proper reference-point for participants to anchor to. In addition, as expected in Hypothesis 1, with the upgrade frame, the reference point remains unchanged regardless of the existence/absence of the external secondary market, as it does not change the set of price anchors relevant to the buying position.

In a similar manner to Table 2.5, Table 2.6 presents supporting results for Hypothesis 4, which is on the effect of the manufacturer's sale of the original version at the replacement-purchase point. The most notable is that, as predicted, with the upgrade frame under a high innovation in the new version, when the manufacturer does not continue production and sale of the original version any longer, the next price anchor relevant to the buying position is the original purchase price of the current version, and hence participants anchor to that. Furthermore, as expected, it does not influence the reference point in the low

Table 2.5: Influential anchors on the reference point with the alternative frames (effect of the existence/absence of an external secondary market price).

Setting	Trade-in frame		Upgrade frame	
	Low innovation	High innovation	Low innovation	High innovation
With sale of the original version With external secondary market	Sec (<0.001) 0.90	Sec (<0.001) 0.95	New (<0.001) 0.27	Curr (<0.001) 0.60
With sale of the original version Without external secondary market	Curr (<0.001) 0.61	Curr (<0.001) 0.66	New (<0.001) 0.38	Curr (<0.001) 0.91
Without sale of the original version With external secondary market	Sec (<0.001) 0.94	Sec (<0.001) 1.04	New (<0.001) 0.32	Opp (<0.001) 0.49
Without sale of the original version Without external secondary market	*const* (<0.001) 91.48	*const* (<0.001) 106.54	New (<0.001) 0.38	Opp (<0.001) 0.74

Parentheses contain p-values. In all cells, HR yields only the indicated price with a significant influence on the reference point.
Numbers in third rows are the coefficients used in between-group comparisons as explained in the text.

innovation case since participants anchor to the new version's price. Finally, in line with predictions of Hypothesis 1, the reference point remains unchanged with the trade-in frame as long as there exists a secondary market to create a price anchor relevant to the selling position. In the absence of the secondary market price, the manufacturer's not selling the original version removes the price that was providing the participants an estimate of what the current version could sell for. As discussed, removing this price leaves the participants with no proper point of reference from the selling position to anchor to.

Discussion:

As seen in Tables 2.4-2.6, the effects of the secondary market, the innovation level, and overlapping production on customers' reference points depend on the frame of the replacement purchase offer. They may shift the reference point with one frame, while having no effect on that with the other frame. From the manufacturer's point of view, the framing effect shifts her control leverage over the replacement buyers' reference points. If the manufacturer goes with the trade-in frame, she deliberately outsources that control to the external secondary market. By choosing the upgrade frame, on the other hand, she takes over the control and can manage her former customers' reference points through her own

Table 2.6: Influential anchors on the reference point with the alternative frames (effect of the manufacturer's (dis)continuing the sale of the original version).

Setting	Trade-in frame		Upgrade frame	
	Low innovation	High innovation	Low innovation	High innovation
With sale of the original version With external secondary market	Sec (<0.001) 0.90	Sec (<0.001) 0.95	New (<0.001) 0.27	Curr (<0.001) 0.60
Without sale of the original version With external secondary market	Sec (<0.001) 0.94	Sec (<0.001) 1.04	New (<0.001) 0.32	Opp (<0.001) 0.49
With sale of the original version Without external secondary market	Curr (<0.001) 0.61	Curr (<0.001) 0.66	New (<0.001) 0.38	Curr (<0.001) 0.91
Without sale of the original version Without external secondary market	*const* (<0.001) 91.48	*const* (<0.001) 106.54	New (<0.001) 0.38	Opp (<0.001) 0.74

Parentheses contain p-values. In all cells, HR yields only the indicated price with a significant influence on the reference point.
Numbers in third rows are the coefficients used in between-group comparisons as explained in the text.

pricing policy. Whether the manufacturer is better off with one frame or the other depends on the circumstances. If the secondary market price is low enough and/or the cost of interfering with that is not too high for the manufacture, the manufacturer may be better off with the trade-in frame, while the otherwise can make the upgrade frame more attractive as long as the manufacturer has the possibility of offering upgrades and adjusting her sale prices accordingly.

It is worth noting here that although the secondary market price does not shift the reference point with the upgrade frame, its elimination seems to increase the reference point's value (through increasing the anchoring coefficient), as seen in Tables 2.5. The secondary market price is the smallest number that participants observe in the experiment, and hence its presence/absence may impact the range of values reported by the participants after anchoring to the chosen reference-point. To examine this secondary role of the secondary market price more carefully, we define a dummy variable for the existence/absence of the secondary market and run the following regressions over the pooled data for when the current sale price of the original version is the reference point (in Appendix A, we repeat the same analysis for when the original purchase price is the reference point and for when the new version's price is the reference point, and obtain similar results):

$$ref_j^U = \beta_c Curr_j + \beta_d dummy\_Sec_j + \beta_{cd} Curr \times dummy\_Sec_j + \varepsilon_j \quad (3)$$

The dummy variable $dummy\_Sec_j$ equals one if participant j had the secondary market price in her/his experimental question and zero otherwise. The $Curr \times dummy\_Sec_j$ is to capture any variable effect driven by the presence of the secondary market price. Table 2.7 shows the outcome of this regression. We see a negative, but statistically insignificant, fixed effect from the presence of the secondary market price. There is no statistically significant variable effect either, which is in line with the result of the Hierarchical Regression.

Table 2.7: Effect of the secondary market price on anchoring with the upgrade frame.

Variable	Coefficient
$Curr$	0.91
	(< 0.001)
$dummy\_Sec$	-12.26
	(> 0.807)
$Curr \times dummy\_Sec$	-0.24
	(> 0.428)

Parentheses contain p-values. $R^2 = 0.691, Adj.R^2 = 0.688.$

On the role of the new version's innovation level, first we note its exact opposite role compared to that of the secondary market: while shifting customers' reference points with the upgrade frame, it does not change the reference point with the trade-in frame (see the first two columns in Tables 2.4-2.6). Moreover, we find no statistically significant difference between the coefficients' magnitudes in column-wise comparisons of the coefficients (all p-values>0.080 in two-tailed z-tests).[7] Therefore, when the manufacturer wants to avoid any interaction between the new version's innovation level and customers' choice of reference point, she can do so by switching to the trade-in frame. The main discussion on the role of the innovation level is to highlight when it shifts the reference point with

[7]It is noted that, since column-wise comparisons are over the same participants, in order to account for possible correlations in error terms, we also used Seemingly Unrelated Estimations for column-wise comparisons and found the same results.

the upgrade frame. With low innovative new versions, the manufacturer is able to keep the replacement buyers' decisions only tied to the new version's price, and hence has a single control variable to influence both new buyers' and replacement buyers' decisions. With a high innovative new version, in contrast, the manufacturer has an advantage of controlling replacement buyers' decisions by sale prices other than the new version's price, which, depending on the circumstances, can provide more flexibility for the manufacturer in adjusting her sale prices.

On the role of overlapping production, the first part of Tables 2.6 shows that, as long as there is an external secondary market, with the trade-in frame, overlapping production has no effect on the reference point (nor does it make a significant difference to the anchoring coefficients; all p-values>0.098 in two-tailed z-tests for row-wise comparisons of the coefficients' magnitudes). With the upgrade frame, the maximum role of overlapping production for the manufacturer realizes when the new version is a high innovation, wherein it shifts customers' reference points from the original purchase price of their current version to its current price. To examine if it has any influence on the magnitude of the reference point under the low innovation case, wherein the reference point is always shaped by anchoring to the new version's price, we perform the same kind of analysis we did on the secondary market price in Equation (3), using the following regression model:

$$ref_j^U = \beta_n New_j + \beta_d dummy\_Curr_j + \beta_{nc} New \times dummy\_Curr_j + \varepsilon_j \qquad (4)$$

The variables $dummy\_Curr_j$ and $New \times dummy\_Curr_j$ have similar interpretations to those of the variables introduced in Equation (3). Table 2.8 shows the outcome of this regression. We see neither a variable nor a fixed effect on the reference point value driven by the existence/absence of the $Curr$. This means that when upgrading to a new version that is an incremental improvement, customers anchor to its sale price and are not affected by the presence/absence of another, less relevant, price-anchor from the manufacturer's sale channel.

To summarize, the experimental results show that trade-ins and upgrades induce differ-

Table 2.8: Effect of overlapping production on anchoring with the upgrade frame under a low innovation.

Variable	Coefficient
New	0.34
	(< 0.001)
$dummy\_Curr$	-11.52
	(> 0.726)
$New \times dummy\_Curr$	0.01
	(> 0.902)

Parentheses contain p-values. $R^2 = 0.740, Adj.R^2 = 0.739.$

ent reference points, and that this shift in reference points breaks down the isomorphism of the two frames. In particular, we found that among multiple potential price-anchors, with a trade-in (respectively, an upgrade) frame customers anchor to the price anchors relevant to the selling (respectively, buying) position as their reference points for the price for their current version. The secondary market price is the reference point with the trade-in frame unless the secondary market does not exist, in which case the current sale price of the original version becomes the reference point (if the manufacturer continues its production and sale at the replacement purchase point). With the upgrade frame, the new version's sale price is the reference point as long as it is not a substantial improvement over their current version; otherwise, either the manufacturer's current sale price of their version or its original purchase price becomes the reference point (depending on the manufacturer's decision on continuing the production and sale of the original version at the replacement purchase point). In addition, the existence/absence of the secondary market does not shift the reference point with the upgrade frame, because it does not change the set of price anchors relevant to the buying position. Similarly, as long as there exists an external secondary market, neither the new version's innovation level nor the manufacturer's decision on overlapping production shifts the reference point with the trade-in frame, as they do not change the price anchor relevant to the selling position.

2.4.2 A Reference-Dependence Model of Trade-ins and Upgrades

Building on the experimental results, we incorporate the framing effect into the classical model of trade-ins and upgrades by Fudenberg and Tirole (1998) to illustrate implications of our results for theory. We do so through extending a reference-dependence version of their classical model and showing that, in the presence of the framing effect and shifts in the reference point driven by that, their key results no longer hold. Here, we only consider the high innovation case as the results we revisit from Fudenberg and Tirole (1998) are made under that case. Hence, the analytical analysis here utilizes only part of the experimental groups among the overall 16 groups we had.

Our approach is to follow Fudenberg and Tirole's (1998) assumptions and model structure closely, adding in the reference dependence in a simple way in line with PEEMs (Portable Extensions of Existing Models) laid out by Rabin (2013a,b) (that is, keeping the same structure of the exciting model in all the ways that are not the focus of the behavioral modification). It is also noted that to create a coherent connection, we stay with the terminology used by Fudenberg and Tirole (1998) and explain the connection with our terminology over previous sections whenever necessary. The problem setting is the same two-period framework explained earlier. The original and the new version have fixed per-unit production costs of c_L and c_H, respectively, where $c_H = c_L + c_\Delta$, and c_Δ indicates the cost of incremental innovation. Also, they have valuations of V_L and V_H, respectively, where $V_H = V_L + V_\Delta$, and V_Δ shows the value of incremental innovation. To avoid trivial cases, it is assumed that $V_L > c_L$ and $V_\Delta > c_\Delta$. The new version's innovation is high if $\frac{V_\Delta}{V_L} > \frac{c_H}{c_L}$; it is moderate if $\frac{c_\Delta}{c_L} < \frac{V_\Delta}{V_L} \leq \frac{c_H}{c_L}$; and it is low otherwise. By normalizing the costs and the valuations, we can obtain simpler representations of the innovation levels, as in Fudenberg and Tirole (1998): $c_L = c_\Delta = 0$ resembles a high innovation level; and $c_L = 0$ and $c_\Delta > 0$ represents a low innovation. Thus, for the rest of the paper, we set $c_L = 0$ and $c_\Delta \geq 0$, where $c_\Delta = 0$ will represent the high innovation case.

There is a continuum of customers indexed by type $\theta \in [0, 1]$. Customers receive utility from the product based on their type, e.g., a customer type $\theta = 0.5$ receives $0.5V_L$ and $0.5V_H$ utility from the original and the new version, respectively. We assume customer types are distributed uniformly on the unit interval. Customer segments are determined by type cut-offs; for example, selling to customers with type $\theta \geq \theta_1$ in period one leaves a volume of $x = (1 - \theta_1)$ new buyers in period two. There is no depreciation in the product, so the manufacturer's sale price of the original version in period two (if she continues its production) is the same as its price in the secondary market.[8] The manufacturer and customers have discount factor δ for their second period payoffs. In line with Fudenberg and Tirole (1998), we also assume $V_L > \delta V_\Delta$. This is a restrictive (and simplifying) assumption; while it does not rule out all high innovation levels, it restricts the very large ones, and meanwhile is very helpful with some of the mathematical simplifications.

There are two cases regarding the market information: anonymous and semi-anonymous. In the anonymous case, where the manufacturer does not keep track of former customers, customers can buy the new version by selling their current version in a frictionless external secondary market or to the manufacturer (when she buys back). In the semi-anonymous case, the manufacturer is able to track the identity of product owners, thus she offers the new version to her former customers at a discounted price, called the upgrade price, less than its regular sale price.[9] As it appears, the anonymous case, wherein the manufacturer keeps buying back of the customers' old versions and selling them of the new version separate, resembles a trade-in frame. The semi-anonymous case, on the other hand, wherein the manufacturer combines the two transactions into a net buying transaction for the customers,

[8] This assumption of Fudenberg and Tirole (1998), though not being realistic, does not bring any qualitative shortfalls in the model and its results, while simplifying the analysis significantly. The otherwise is just a matter of depreciation parameter and making qualitative changes. Hence, we keep this assumption in our behavioral extension as well and assume no depreciation (or desperation rate equal to one) in our model and analysis. It is noted that since our experimental design was with the secondary market price less than the sale price, our extension does allow depreciation (through a depuration rate less than one) if one wishes to expand Fudenberg and Tirole's (1998) model in that direction in the interest of quantitative results.

[9] Here, we ignore Fudenberg and Tirole's (1998) "identified" case, where replacement buyers cannot hide that they are repeat buyers to buy the new version as a new buyer, and where the manufacturer is therefore able to charge an upgrade price higher than the regular sale price (this case may apply to services, such as internet and cable TV, but rarely to durable goods).

resembles an upgrade frame.

We start with the semi-anonymous case (i.e., the upgrade frame). The manufacturer's optimization begins with her production-pricing decision in period two, given the cut-off θ_1 from period one. Figure 2.1 shows the market segmentation. The θ_u is the cut-off for the former customers' type who decide to upgrade in period two. The upgrade price for these customers is $p_u = \theta_u V_\Delta$, coming from $\theta_u V_L - p_u = \theta_u V_H$. The θ_H and θ_L, on the other hand, are the cut-offs for the new customers' types who buy the new and the original version, respectively, in period two. It is easy to show that $p_H = \theta_L V_L + \theta_H V_\Delta$ and $p_L = \theta_L V_L$.

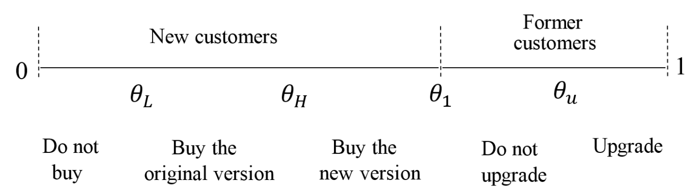

Figure 2.1: Market segmentation in the semi-anonymous case.

With the cut-offs illustrated in Figure 1, the manufacturer's profit function in period two is as follows (note that $c_L = 0$):

$$\Pi_2^S = [1-\theta_u][\theta_u V_\Delta + \lambda(\theta_L V_L + \theta_H V_\Delta - \theta_u V_\Delta - \alpha R) - c_H] + [\theta_1 - \theta_H][\theta_L V_L + \theta_H V_\Delta - c_H] + [\theta_H - \theta_L]\theta_L V_L \quad (5)$$

Where, $\theta_1 \leq \theta_u$.

The constraint $\theta_1 \leq \theta_u$ ensures that the volume of upgraders cannot be more than the volume of sale in period one. The $\lambda \geq 0$ is the reference-dependence weight, and with the special case of $\lambda = 0$, which implies no reference-dependence, Equation (5) is identical to the classical model presented in Fudenberg and Tirole (1998). Therefore, with $\lambda = 0$, and for $c_\Delta = 0$ (representing the high innovation case), we get their main result that the manufacturer will not continue production of the original version in period two (and hence, $p_H = \theta_H V_H$). The general intuition behind this result lies in the perception that producing the original version in the presence of a high innovative new version is useless. The behavioral model introduces an additional term, $\lambda(\theta_L V_L + \theta_H V_\Delta - \theta_u V_\Delta - \alpha R)$, or equivalently $\lambda(p_H - p_u - \alpha R)$, embedded in upgraders' utility function, which represents their reference-dependence behavior: $p_H - p_u$ is what the upgraders receive for their current version, $\alpha \geq 0$ is the strength of anchoring to the reference point R ($\alpha = 0$ means customers expect zero price for their current product and their reference-point value is zero), and λ is the reference-dependence weight. It is obvious that adding reference-dependence changes the optimal cut-offs and thus the manufacturer's optimal profit. Form the manufacturer's point of view, the reference dependence in Equation (5) has a simple interpretation: if the discount amount (i.e., $p_H - p_u$) is bigger than what the customers expect for their current version (i.e., $p_H - p_u - \alpha R > 0$), it brings a gain for them. Thus, the manufacturer can charge them more, up to the amount $p_u + \lambda(p_H - p_u - \alpha R)$, and still make the transaction happen. On the other hand, if the discount amount is lower than the customers' reference-point value and yields a loss (i.e., $p_H - p_u - \alpha R < 0$), the manufacturer has to charge them less (i.e., $p_u + \lambda(p_H - p_u - \alpha R) < p_u$) in order to make the upgrade happen. In light of what the reference dependence adds, there could be value for the manufacturer in sale of the original version in period two. That is, if the manufacturer can manage the former customers' reference points via the sale price of the original version, continuing its production and sale in period two is now not useless.

Based on the experimental results for Hypotheses 1 and 4, with an upgrade under a high innovation in the new version, two situations can happen regarding the reference point: I) customers anchor to the current sale price of their product, when the manufacturer sells the original version in period two; and II) customers anchor to the price they have paid to purchase their product in period one, when the manufacturer does not continue its production and sale in period two. Given that the sale price of the original version in period two is significantly lower than its price in period one, we can assume that customers' anchoring parameter (α) to the original purchase price of their model would be lower than that to the current sale price of that model (which can also be inferred from Tables 2.5 and 2.6). With this, we normalize $\alpha = 1$ in customers' anchoring to the current sale price of their product, and $0 < \alpha < 1$ in their anchoring to its original purchase price. This will remarkably simplify the analysis and later the comparison between the semi-anonymous and the anonymous cases, too. In addition, in the interest of simplicity, we assume $\lambda = 1$ regardless of loss/gain in selling the current model. These simplifications make no qualitative changes in the outcome of the model, while making the behavioral extension possible by adding only one new parameter ($0 < \alpha < 1$) to the existing model.

From the first part of Table 2.6, and with $\alpha = 1$, we replace αR in Equation (5) with $\theta_L V_L$ (i.e., the current sale price of the current version) that yields the manufacturer's second-period profit function (for when she continues the production and sale of the original version in period two) as follows:

$$\Pi_2^{Sp} = [1-\theta_u][\theta_H V_\Delta - c_H] + [\theta_1 - \theta_H][\theta_L V_L + \theta_H V_\Delta - c_H] + [\theta_H - \theta_L]\theta_L V_L \tag{6}$$

Where, $\theta_1 \leq \theta_u$.

Similarly, from the second part of Table 2.6, and with $0 < \alpha < 1$, replacing αR with αp_1 in Equation (5), we have the manufacturer's second-period profit function (for when she does not continue production and sale of the original version in period two) as follows:

$$\Pi_2^{Sn} = [1-\theta_u][\theta_H V_H - \alpha p_1 - c_H] + [\theta_1 - \theta_H][\theta_H V_H - c_H] \tag{7}$$

Where, $\theta_1 \leq \theta_u$.

The manufacturer's first-period profit comes from selling the original version to new buyers, adding the second-period profit to which shapes the manufacturer's total profit function, that is, $\Pi^S = \Pi_1^S + \delta\Pi_2^S$ (see Appendix A for deriving its mathematics). Comparing the manufacturer's total optimal profit with and without overlapping production in period two, we present the following proposition on the manufacturer's optimal strategy:

PROPOSITION 1. *In the semi-anonymous case, under a high innovation in the new version, for $\alpha > \hat{\alpha}$ the manufacturer is better off with overlapping production of the original version along with the new version in period two.*

The proof is available in Appendix A. Proposition 1 provides an opposite prediction to that of the classical model. The intuition behind Proposition 1 is straightforward and has roots in the rationale that sale of the original version in period two can be useful with managing the upgraders' reference points, that is, the manufacturer continues producing the original version and drops its price significantly to reduce the upgraders' expectation as to the price for their current version. Hence, for α of sufficient magnitude, the reference-point value under overlapping production ($\theta_L V_L$) is lower than that under no overlapping production (αp_1), and hence, driven by the reference dependence, the manufacturer can charge a higher upgrade price under overlapping production than under no overlapping production. In addition, with overlapping production in period two, the manufacturer is not limited in charging higher sale prices for the original version in period one, as she has the control over the upgrades' reference points through her second-period sale prices. This would in turn increases the first-period profit for the manufacturer. Therefore, there are α's for which the manufacturer is better off with overlapping production in period two. Figures 2.2 illustrates this by comparing the manufacturer's optimal profits under overlapping production and no overlapping production for all ranges of $0 \leq \alpha \leq 1$.

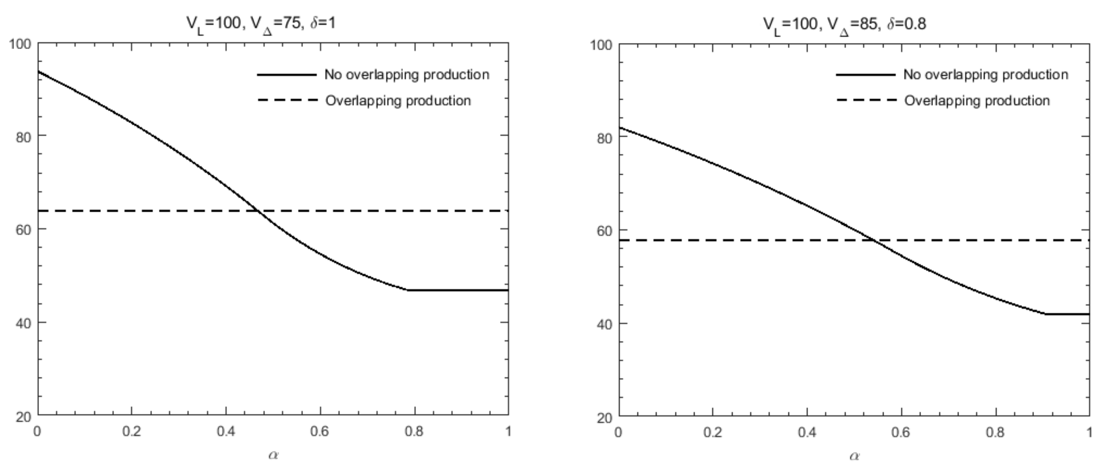

Figure 2.2: Manufacturer's optimal profit in the semi-anonymous case.

Figures 2.3 depicts the optimal solution, i.e., the optimal cut-offs, under no overlapping production for $0 \leq \alpha \leq 1$. In the extreme case of $\alpha = 0$, where former customers have no expectation as to a price for their current version and behave like new buyers in period two, the second period sale becomes independent from the first period sale, and as a result, all cut-offs reach the optimal point of $\theta^* = \frac{1}{2}$. This is very similar to a situation where the product's life cycle is only one period. When α increases above zero, the manufacturer faces the upgraders' expectation in period two. To manage that, she reduces the sale price in period one to keep the upgraders' reference point low in period two. In addition, she keeps the new version's sale price high to induce a high discount price (i.e., $p_H - p_u$) to the upgraders. The former increases the manufacturer's sale in period one, and the latter hinders selling to new buyers in period two. Thus, the manufacturer only relies on the upgraders and does not sell to new buyers in period two. After a point, the manufacturer cannot keep the upgraders' reference point at low levels, because α is too high. Nor can she handle the high reference-point in period two by increasing the new version's sale price (to increase the induced discount price). Therefore, she starts relying on new buyers in period two; the bigger is α, the smaller are sales in period one and upgrades in period two, but the bigger is the sale to new buyers in period two. This is reached by setting a high cut-off in period one (θ_1) and a low cut-off in period two (θ_H). For very high α's, the manufacturer is not able to handle the upgraders' reference point, and hence starts relying only on new buyers in each period, where no upgrade happens in period two, and the manufacturer's optimal profit reaches a steady point.

We next analyze the model of the anonymous case (i.e., the trade-in frame) in Fudenberg and Tirole (1998). Similar to the semi-anonymous case (i.e., the upgrade frame), the manufacturer's optimization begins with her production-pricing decision in period two, given the cut-off θ_1 from period one:

$$\Pi_2^A = [1-\theta_1][\theta_L V_L + \theta_H V_\Delta - \theta_L V_L + \lambda(\theta_L V_L - \alpha R) - c_H] + [\theta_1 - \theta_H][\theta_L V_L + \theta_H V_\Delta - c_H] + [\theta_H - \theta_L]\theta_L V_L \tag{8}$$

The terms $p_H = \theta_L V_L + \theta_H V_\Delta$ and $p_L = \theta_L V_L$ are the second-period prices for the

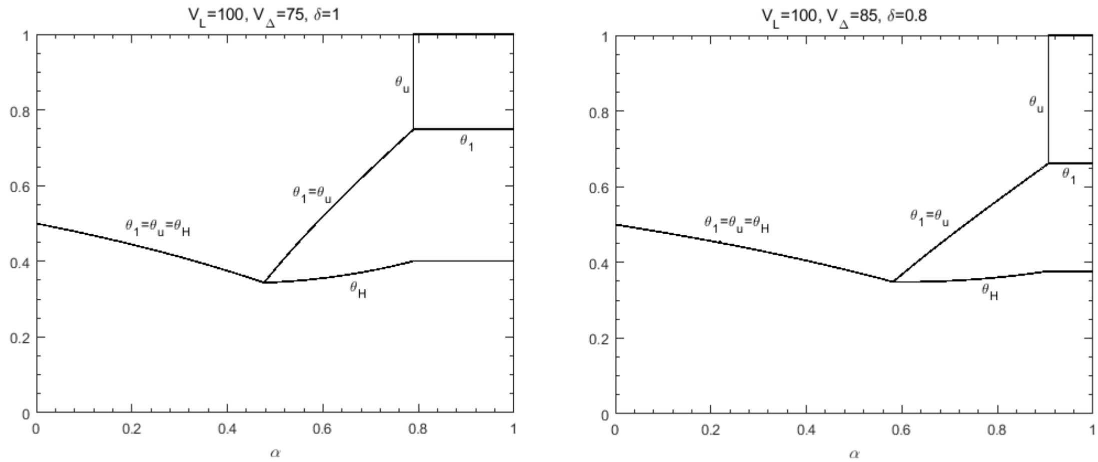

Figure 2.3: Manufacturer's optimal solution in the semi-anonymous case.

new and the original versions, respectively. The manufacturer sells the new version to all customers (i.e., replacement buyers and new buyers) at the regular sale price $\theta_L V_L + \theta_H V_\Delta$, she sells the original version to new buyers at price $\theta_L V_L$ and, because of no depreciation in the product, buys it back from the replacement buyers at the same price $\theta_L V_L$. The $\lambda \geq 0$ and $\alpha \geq 0$ have the same meanings as in Equation (5)). With $\lambda = 0$, Equation (8) is identical to the classical model in Fudenberg and Tirole (1998), and $c_\Delta = 0$ (representing the high innovation case) yields their main result, that the manufacturer does not produce the original version in period two.[10] The reference-dependence term has the same effect as in Equation (5)); it modifies the repeat customers' utility function, and changes the optimal cut-offs and the manufacturer's optimal profit. From the manufacturer's point of view, it has the same interpretation: if the price paid for the customers' current version is higher than what they expect it to be (i.e., $\theta_L V_L > \alpha R$), they feel a gain in the amount of $\lambda[\theta_L V_L - \alpha R]$. Thus, the manufacturer can reduce the price she pays to replacement buyers (i.e., $\theta_L V_L - \lambda[\theta_L V_L - \alpha R]$) and still make the trade-in happen. Note that the manufacturer can also increase the new version's price (i.e., $\theta_L V_L + \theta_H V_\Delta$) in taking advantage of $\theta_L V_L > \alpha R$;

[10]Note that although the manufacturer does not produce the original version in period two, new customers can still buy it from the secondary market, and hence there always exists a cut-off θ_L for those customers, and the new version's price always takes the form '$p_H = \theta_L V_L + \theta_H V_\Delta$' not '$p_H = \theta_H V_H$'.

however, she is limited in doing so because it causes her to lose the new buyers segment.

Based on the experimental results, with the trade-in frame, customers anchor to the secondary market price as their reference points, regardless of the new version's innovation level or the manufacturer's decision on production and sale of the original version in period two. Hence, from the first part of Table 2.6, and with $\alpha = 1$ (driven by Fudenberg and Tirole's 1998 assumption of no depreciation in the product and that the secondary market price of the original version is the same as its sale price in period two (when the manufacturer continues its production and sale)), we replace αR in Equation (7) with $\theta_L V_L$. This results in $\lambda(\theta_L V_L - \alpha R) = \lambda(\theta_L V_L - \theta_L V_L) = 0$. Hence, we have the manufacturer's profit function in period two as follows:

$$\Pi_2^A = [1 - \theta_1][\theta_H V_\Delta - c_H] + [\theta_1 - \theta_H][\theta_L V_L + \theta_H V_\Delta - c_H] + [\theta_H - \theta_L]\theta_L V_L \tag{9}$$

The manufacturer's first-period profit comes from selling the original version to new buyers, adding the second-period profit to which shapes the manufacturer's total profit function, that is, $\Pi^A = \Pi_1^A + \delta\Pi_2^A$ (see Appendix A for deriving its mathematics). It is worth noting that, based on what we obtained in Equation (9), with $\alpha = 1$ the reference-dependence model coincides with the classical model. Thus, Fudenberg and Tirole's (1998) main result in the anonymous case, i.e., not producing the original version, holds with the reference-dependence model as well. In comparing the anonymous case (i.e., the trade-in frame) and the semi-anonymous case (i.e., the upgrade frame), however, the anomaly between the predictions of the classical model and the reference-dependence model for the latter brings us to the following proposition:

PROPOSITION 2. *Under a high innovation in the new version, for $\alpha < \hat{\alpha}$ the manufacturer is better off with the semi-anonymous case than with the anonymous case. For $\alpha \geq \hat{\alpha}$, both yield the same optimal profit through the same optimal pricing.*

Proposition 2 is in stark contrast with the classical model, which favors the anonymous case.[11] The intuition behind Proposition 2 roots back in the previous result presented in

[11] The intuition behind this result of the classical model is that because of not pricing the original version in period two of the semi-anonymous case, the new version's sale price falls below the commitment price,

Proposition 1, where the overlapping production becomes the optimal strategy for $\alpha \geq \hat{\alpha}$. The overlapping production in the semi-anonymous case (i.e., the upgrade frame) pushes the optimal solution closer to the commitment solution by creating the cut-off θ_L in period two. Furthermore, since for $\alpha \geq \hat{\alpha}$, the reference point is $\theta_L V_L$ in both cases (i.e., with both frames), they yield the same optimal pricing and hence the same optimal profit (the mathematical proof of this is provided in Appendix A). That being said, the baseline profit in Figure 2.2 presents the optimal profit in the anonymous case (i.e., the trade-in frame) as well. As seen, for $\alpha < \hat{\alpha}$, the semi-anonymous case (i.e., the upgrade frame) yields more profit than the baseline, because of its reference point being very low; and for $\alpha \geq \hat{\alpha}$, the manufacturer can always obtain the baseline profit by switching to overlapping production.

2.5 Concluding Remarks

This study examined whether the conventional wisdom that trade-ins and upgrades are isomorphic was robust to behavioral influences in customers' decision-making. The analyses yielded two main results. First, experimental methods established that people change which prices they anchor to as their reference points when offered the alternative framings of a trade-in or an upgrade, and that different market settings have different effects with the two frames, in terms of shifting customers' reference points, depending on which price anchors they change. Second, the paper shows that framing effect of the sort revealed in the experiments, through reference-dependence mechanism, can overturn key predictions of the classical model of trade-ins and upgrades. In particular, the paper shows that manufacturers may prefer to produce older versions of their product concurrently with newer improved versions, even in settings where the classical model predicts they would not. The addition of reference dependence to the classical model therefore can explain observed coproduction of successive versions of products that cannot be explained by the classical model.

---

and hence the manufacturer never reaches the commitment solution. However, the secondary market in the anonymous case, which creates the cut-off θ_L for the original version in period two, pushes the cut-off θ_H upward, which in turn helps the manufacturer get closer to the commitment solution.

This modification further contradicts the classical model's outcome in comparing semi-anonymous and anonymous markets, through a comparison of trade-ins and upgrades, and shows that the dominant profitability of the anonymous case is not the case.

Our findings have obvious implications for manufacturers of durable goods who offer replacement purchase for their former customers. The choice between offering a trade-in or an upgrade influences customer's choice of the reference point for the price for their current version. Hence, manufacturers can choose the frame to direct their customers' focus away from low potential reference-points and towards high profitable ones. Because switching between frames is likely to be a low-cost tactic for most manufacturers, reference dependence offers a flexible tool in managing the former consumers in a profitable way. The extent to which manufacturers can interfere with external secondary markets and their flexibility in adjusting their own production-pricing policies would determine which frame they would be better off with. Finding the optimal frame and the optimal production-pricing policy in each market setting, which better be done through internalizing the choice of innovation level and the option of switching between the market settings for manufacturers, was beyond the scope of this paper. Nonetheless, we designed our experiments to cover all possible market settings in line with seminal studies of trade-ins and upgrades in order to provide inputs for future analytical research on this area.

CHAPTER 3

STRUCTURING SUPPLY CHAINS FOR SOCIALLY RESPONSIBLE BEHAVIOR

3.1 Introduction

Recent labor and building code violations, e.g., child labor by Nike's suppliers in the 1990s (Nisen 2013), environmental violations by Mattel's suppliers in the early 2000s (Roosevelt 2011), child labor by GAP's suppliers in 2007 (Brown 2007), workers' suicides in 2010 at Foxconn due to poor working conditions (Barboza 2010), and most tragically, the Bangladesh factory collapse in 2013 that killed more than 1,100 workers and injured more than 2,500 (Al-Mahmood 2013, Yardley 2013), have increased the salience of socially responsible sourcing in global supply chains. Unfortunately, sourcing from socially or environmentally nonresponsible suppliers is an ongoing issue in global supply chains (e.g., possible child labor in Apple's supply chain of the iPhone X; Vega 2017). Based on a recent report by the International Labor Organization, there are more than 152 million children currently working in child labor (ILO 2017).[1] The challenge of monitoring and controlling suppliers' compliance with social and environmental obligations adds to the complexity of managing these supply chains. Consumers wanting their products to be made in a socially responsible way participate in markets located far away from the factory workers and managers responsible for the process that makes these products. Firms interacting with the end consumers are, through manufacturers, suppliers, and subcontractors, removed from the daily decision making that determines the conduct in such processes. Creating a discipline of social responsibility in a supply chain is thus a task of aligning interests, policies, and behavior among many globally distributed partners.

[1] According to the International Labor Organization, not all work done by children is classified as child labor, and child labor is often defined as work that deprives children of their childhood, potential and dignity, and is harmful to their physical and mental development.

Operations Management researchers have started studying responsible sourcing in supply chains (see, e.g., Plambeck and Taylor 2015, Porteous et al. 2015, Guo et al. 2016, Chen and Lee 2017, Huang et al. 2017, Agrawal and Lee 2019, Lee and Tang 2018, Bondareva and Pinker 2019). The social and environmental impact of improving responsible sourcing in a firm's operation is discussed extensively in the academic literature as well as among practitioners (see, e.g., Guo et al. 2016, Kippenberg 2018). The focus of academic research in this area lies in studying the consequences of sourcing from socially/environmentally deviant suppliers on buying firms' profits and analyzing their optimal sourcing strategies, mainly by investigating how socially conscious consumers' reactions can increase the interest that firms have in sourcing their products responsibly.

Based on the academic literature and current practices such as Fair Trade certification, GoodWeave label, etc., we can differentiate between encouraging and discouraging consumer reactions in the market: customers' extra willingness-to-pay for responsibly sourced products is an encouraging reaction that values responsible sourcing; and, in contrast, customers' potential to boycott the firm if a breach of responsible sourcing becomes apparent is a discouraging reaction that denounces nonresponsible sourcing. Which consumer reaction—an encouraging or a discouraging one—is more effective in terms of promoting responsible sourcing? Answering this question is of practical value, particularly to the NGOs active in promoting socially responsible behavior among consumers. For example, GoodWeave is an NGO "dedicated to ending child labor, forced labor, and bonded labor in global supply chains." GoodWeave has created a "GoodWeave Label" that indicates a labeled product had no child labor in its production process. To make this initiative effective, GoodWeave needs consumers to react to this label. Learning about the effects of different consumer reactions in different industries and supply chains, an NGO such as GoodWeave can in turn emphasize the right reaction in their communication with consumers to help promote a target behavior (McKenzie-Mohr 2011).

For a product with many substitutes, the extra willingness-to-pay from consumers for a firm's responsible product may be limited since consumers can easily buy the product from other manufacturers. Thus, in supply chains for products with many substitutes, an NGO like GoodWeave may find it easier to promote discouraging reactions among consumers. This applies for products like T-shirts, and everyday essentials, which are available in comparable quality from many brands and manufacturers in the market. However, for products with little substitutes, consumers may have limited opportunity to boycott the product, since they will not be able to easily find an alternative product from other manufacturers. In this case, NGOs can more easily promote encouraging reactions among consumers. This applies for products like coffee, rice, hand-woven rugs, etc., where unique taste and characteristics of the product, often rooting back in its place of origin, makes customers stick with the brand/manufacturer while willing to pay extra for the responsibly sourced version of the product.

One of the seminal models of responsible sourcing in supply chains that analyzes the effect of different consumer reactions on firms' sourcing policy is that of Guo et al. (2016) (hereafter, GLS). Through a stylized model of a "buyer," a "responsible supplier," and a "nonresponsible supplier," the GLS model provides a simple and insightful framework to study the rational motivations of supply chain participants under different sourcing policies. This framework has been used by other researchers as well, see, e.g., Agrawal and Lee (2019). The possible sourcing policies highlighted by GLS are low-cost sourcing (i.e., sourcing from the nonresponsible supplier), dual sourcing (i.e., sourcing from both suppliers), and responsible sourcing (i.e., sourcing from the responsible supplier). GLS discuss these policies extensively and provide examples for each sourcing policy (e.g., dual sourcing by Whole Foods Market that sells multiple versions of meats graded on 'responsibility scale', or dual sourcing by apparel manufacturers that sell 'made in the US' versions together with regular versions). The framework allows studying, for example, when a buyer may prefer relatively more responsible policies (dual sourcing or responsible sourcing) over low-cost sourcing. In general, firms may prefer sourcing from nonresponsible suppliers over responsible sourcing because they are cheaper (which results in low-cost sourcing); further, they may prefer dual sourcing over single-sourcing due to the potential for price discrimination and market differentiation (e.g., selling the ‘made in the US’ version at a premium price in parallel to a regular version) (Guo et al. 2016). These two motivations can work against responsible sourcing.

The GLS model and analyses, which we refer to as the classical model in the interest of comparison with our behavioral model, show that among the factors influencing the buyer’s sourcing strategies, only boycotting behavior by socially conscious customers reliably decreases the buyer’s sourcing from the nonresponsible supplier. In contrast, the socially conscious customers’ extra willingness-to-pay (and the portion of these customers in the market that determines the scale of this extra revenue for the buyer) can backfire and increase nonresponsible sourcing. This result casts doubt on the effectiveness of an encouraging reaction by the market. In addition, this insight can be of limited practical value for NGOs, governments, and policymakers. Asking consumers to, for example, boycott products that have no substitutes may not be feasible. Moreover, the classical model provides conflicting advice on the effect of a discouraging consumer reaction in the context of products with low salvage value. Specifically, the model predicts that increasing the socially conscious customers’ willingness to boycott will push the buyer toward sourcing from the nonresponsible supplier in an effort to avoid excess stock of an expensive, responsibly sourced product in case of a boycott. The model thus does not provide a clear recommendation for products with low salvage value, e.g., perishables and fashion products. Understanding how to promote responsible sourcing in the fashion industry is particularly important given the prevalence of nonresponsible labor practices among suppliers in this industry.

In practice, we find some evidence that dual sourcing is abundant beyond the rational price discrimination and market differentiation that GLS assume. For example, with Fair Trade certification, which emphasizes paying fair wages to labors at the sourcing farms/factories, dual sourcing is the dominant policy among the firms selling Fair Trade certified products (see fairtradecertified.org). This dual sourcing, and hence selling both regular and Fair Trade versions of products, is being practiced by many firms listed in fairtradecertified.org and for many products (such as coffee, tea, chocolate, cloth, furniture, seafood, cereal, candies/cookies, snacks, etc.) even when the price difference between variants is very low. In fact, customers' extra willingness to pay for Fair Trade versions can be low and as little as 8% of the product's sale price (Hainmueller et al. 2015). It is thus possible that this level of dual sourcing is not entirely rational, but is also driven by dual sourcing bias.

The dual sourcing bias, which is documented in the context of sourcing decisions (Gurnani et. al 2014), implies that firms tend to source from multiple suppliers, even when single sourcing is the rational optimal policy. The behavioral roots of the dual sourcing bias are in diversification and variety-seeking biases (Simonson 1990, Read and Loewenstein 1995, Benartzi and Thaler 2001) and partition-dependence decision making (Fox and Clemen 2005, Fox et al. 2005). Therefore, we develop a behavioral model to incorporate this dual-sourcing bias into the buyer's sourcing decision with the GLS model. Such a bias may be generally present with the buyer facing two different suppliers. In our context, this bias could be particularly salient since dual sourcing allows the buyer to segment the market. In other words, we expect that the market segmentation potential created through an encouraging consumer reaction will lead to the dual-sourcing bias. The GLS model, by separating different sourcing policies for the buyer, provides us a perfect platform to incorporate this dual-sourcing bias.

We find that the behavioral model we propose mostly agrees with the classical model in terms of its predictions but does provide clearer advice in those contexts where the recommendations given by GLS appear impractical or ambiguous. For these scenarios, which we will outline in more detail below, we compare predictions from our behavioral model with those from the classical model via a behavioral experiment, to test our behavioral model.

Based on our analyses, we find that the encouraging reaction always (never) increases the firm's responsible sourcing when the discouraging reaction in the market is weak (strong), and that the discouraging reaction always increases responsible sourcing regardless of the product's salvage value. These results allow us to make a clear recommendation for NGOs: in supply chains of products with no (many) substitutes, the encouraging (discouraging) reaction is the most effective reaction from the market to support responsible sourcing in the supply chain. In other words, NGOs can fit their communication strategy with consumers to the nature of the product.

The remainder of the paper is organized as follows. Section 3.2 reviews the related literature. Section 3.3 develops our theory and builds the behavioral model. The experiment we use to validate our behavioral model is summarized in Section 3.4, together with an analysis and comparison of the classical and behavioral models' predictions. Finally, Section 3.5 contains our concluding remarks. All proofs are presented in Appendix B.

3.2 Literature Review

Our research questions contribute to four streams of literature: sourcing decisions in supply chains, socially and environmentally responsible operations management, market segmentation, and behavioral operations management. We discuss these different literature streams below.

3.2.1 Sourcing Decisions in Supply Chains

We study a firm's sourcing decision. The firm decides how much inventory to procure from a responsible (and more expensive) supplier and a nonresponsible (and cheaper) supplier. Thus, our research is related to the literature of supply chain sourcing with multiple suppliers. Because one of the suppliers in our case has an inherent chance of violating social or environmental obligations, our research relates to the supply disruption literature (see, e.g., Tomlin 2006). However, as GLS mention, the type of risk in the responsible sourcing decision differs slightly from the traditional risk of supply chain disruption. For a buyer, the negative consequences of violating social or environmental obligations lie in potentially losing a portion of the market. The disruption is thus effectively on the demand side (i.e., loss of customers) instead of the supply side (i.e., insufficient supplies). Hence, the common mitigation strategies in supply chain disruption, e.g., supplier diversification and improvement (see Wang et al. 2010 and Kalkanci 2017, for a literature review), are inapplicable in this context. Furthermore, driven by the socially conscious customers' extra willingness-to-pay for the responsibly sourced product, there is a potential for market segmentation tied directly to the firm's sourcing decision. This segmentation aspect, which contrasts with the traditional supply disruption problems in the literature, brings further nuances to the firm's sourcing decision. We study potential behavioral biases in this context, specifically dual-sourcing and oversourcing biases (Gurnani et al. 2014, Goldschmidt et al. 2014, Csermely and Minner 2015, Kalkanci 2017). Our paper also extends the work on supplier diversification in procurement for reasons other than supply uncertainty (e.g., Chod et al. 2019).

3.2.2 Market Segmentation and Price Discrimination

As noted above, the socially conscious customers' extra willingness-to-pay for a responsibly sourced product creates an interdependence between the buying firm's sourcing strategy and its marketing strategy: The firm can practice market segmentation and price discrimination only if it sources from both responsible and nonresponsible suppliers. A certain population is willing to pay a premium for responsible conduct. A large body of the work in the marketing literature is devoted to examining pricing and market segmentation strategies (see, e.g., Moorthy 1984, Moorthy and Png 1992, Desai 2001). Broadly related to our study, Chen (2001) studied product development and market segmentation of green and regular products where a green consumer segment is willing to pay a premium for the green attribute in the product. In the area of responsible sourcing, except for GLS, no work has studied both sourcing and market segmentation for responsibly sourced products. Inspired by the seminal papers in the marketing literature that look at behavioral biases related to diversification bias and variety-seeking behavior (Simonson 1990, Read and Loewenstein 1995, Benartzi and Thaler 2001) and partition-dependence decision making (Fox and Clemen 2005, Fox et al. 2005), we study how the existence of potential market segmentation, through explicitly partitioning the market between socially conscious and regular customers, would influence the firm's sourcing decision.

3.2.3 Socially and Environmentally Responsible Operations

The academic literature on sustainable and environmental operations and supply chain management is vast. Our study falls under the growing research stream of responsible sourcing. Chen and Lee (2017) and Huang et al. (2017) reviewed the academic work in this area. This stream of research is mostly concerned with strategies and policies to induce more responsibility among suppliers (see, e.g., Plambeck and Taylor 2014, Porteous et al. 2015, Chen and Lee 2017, Agrawal and Lee 2019, Bondareva and Pinker 2019). Like GLS, our study focuses on a firm's supplier selection and sourcing decisions and how consumers' reaction to the firm's (non)responsible sourcing can influence these decisions. Building on GLS, we bring a behavioral lens to the firm's decision making and explore how relevant behavioral biases affect the firm's sourcing decision. Following the lead of Lee and Tang (2018), who recommended applying a behavioral lens to the growing area of socially and environmentally responsible operations, we develop a behavioral model of responsible sourcing. We examine how this behavioral approach in modeling a firm's sourcing strategy can modify some of the insights derived from the classical model.

3.2.4 Behavioral Operations Management

Considerable research in behavioral operations management applies behavioral insights to topics in operations and supply chain management. For example, many researchers have examined the behavioral biases in newsvendor decision making with the goal of developing new models that can better explain human decision making and provide practical insights (see, e.g., Schweitzer and Cachon 2000, Su 2008, Ho et al. 2010). We apply behavioral insights to the same purpose, using concepts such as the diversification bias (Simonson 1990, Read and Loewenstein 1995, Benartzi and Thaler 2001) and the dual sourcing bias (Gurnani et al. 2014) in the context of responsible sourcing. Our study extends the recent stream of work in the behavioral operations management literature on firms' sourcing decisions (see, e.g., Gurnani et al. 2014, Goldschmidt et al. 2014, Csermely and Minner 2015, Kalkanci 2017). In contrast to the existing work, our study does not focus on supply uncertainty, but explores the impact of market segmentation potential through responsible sourcing on a firm's sourcing decision.

3.3 Theory Development

An important objective of our study is to document and explain behavioral deviations from the normative benchmark, which in our case corresponds to the model proposed by GLS. We will first review this existing model, then make predictions based on it and finally introduce behavioral forces that lead to deviations from these rational predictions.

3.3.1 Problem Setting

The problem setting is the same as that of GLS: A buyer needs to procure a product to sell to consumers in a market. The buyer will sell the product at a fixed retail price (V) to fulfill a constant demand (e.g., the market consists of $10,000$ customers each of whom will buy only one product). The product is available from two different suppliers: one that is socially responsible and abides by social and environmental obligations and another that is not responsible and might violate these obligations with a known chance of ϕ. The purchasing cost from the responsible supplier (C_R) is higher than that from the nonresponsible supplier ($C_{NR} < C_R$). Both suppliers will meet their purchase commitments; however, if a violation occurs on the suppliers' side, it may cause the buyer to lose a portion of the market. Specifically, a certain portion of the customers (i.e., θ of the $10,000$ customers) are socially conscious, and are willing to pay r more for the product if it is sourced from the responsible supplier. Meantime, α of these socially conscious customers will refuse to buy the product if a violation occurs. Two cases can be assumed regarding unsold products: the buyer can return the unsold products to the supplier and fully recover the purchase cost (we refer to this case as 'full salvage value'); the buyer can only realize a salvage value less than the full purchase cost. If low enough, this salvage value can influence the buyer's sourcing decision (we refer to this case as 'low salvage value'). The buyer's objective is to maximize expected profit by deciding how many products to procure from each supplier.

3.3.2 Normative Theory

As GLS demonstrate in their research, three factors determine the buyer's sourcing strategy: r, which influences the relative profitability of sourcing from the suppliers (i.e., $C_R - C_{NR} + r$); θ, which determines the scale of how much revenue the buyer can obtain from r; and α, which determines the scale of the consequences of sourcing from the nonresponsible supplier. These three factors permit four possible sourcing policies (as Figure B1 in Appendix B depicts): low cost (LC), i.e., sourcing from the nonresponsible supplier and selling the same product to both customer segments at price V; dual sourcing (DS), i.e., sourcing from both suppliers to sell different products to different customer segments at prices V and $V + r$; responsible mass (RM), i.e., sourcing from the responsible supplier and selling the same product to both customer segments at price V; and responsible niche (RN), i.e., sourcing from the responsible supplier and selling only to socially conscious customers at price $V + r$. Table 3.1 summarizes the buyer's profit under each of these four possible sourcing policies.[2] Based on the analysis by GLS, the buyer decides on her sourcing policy by choosing the one that yields the highest profit. An improvement in re-

[2]In implementing the experiment, we use 10,000 for the number of customers, instead of normalizing the customer volume to 1 as in GLS.

sponsible sourcing is realized when the market reactions provide incentives for the buyer to prefer a more responsible sourcing policy over a less responsible one. For example, if the buyer prefers DS over LC (respectively, RM), an improvement (respectively, a deterioration) in responsible sourcing happens because this preference results in less (respectively, more) sourcing from the nonresponsible supplier. It is noted here that since the buyer's preference between RM and RN does not influence nonresponsible sourcing, the main focus is on the buyer's preference between LC, DS, and RM (nonetheless, all insights and results of our study apply to the buyer's preference between RM and RN as well). Using the results from GLS, we structure the thresholds that lead to these preferences, both for the full salvage value case (Table 3.1)[3] and for the case of a low salvage value (Table 3.2)[4]:

Table 3.1: The buyer's profit with four possible sourcing policies.

Sourcing policy	Buyer's profit	Optimal responsible sourcing	Optimal nonresponsible sourcing
Low cost (LC)	$(10{,}000 - \phi\alpha)(V - C_{NR})$	0	10,000
Dual sourcing (DS)	$(10{,}000 - \theta)(V - C_{NR}) + (\theta - \phi\alpha)(V + r - C_R)$	θ	$10{,}000 - \theta$
Responsible mass (RM)	$10{,}000(V - C_R)$	10,000	0
Responsible niche (RN)	$\theta(V + r - C_R)$	θ	0

LEMMA 1. *LC policy dominates RM policy when $(C_R - C_{NR}) > \phi\dfrac{\alpha}{10,000}(V - C_{NR})$ (respectively, $(C_R - C_{NR}) > \phi\dfrac{\alpha}{10,000}V$ in the case of low salvage value), and RM dominates LC otherwise.*[5]

Lemma 1 indicates that the benefit of pursuing LC instead of RM lies in savings in the procurement cost by $10,000(C_R - C_{NR})$. The downside of LC compared with RM is the possible market loss by $\phi\alpha(V - C_{NR})$ (respectively, $\phi\alpha V$ in the case of low salvage value).

[3]Because it adds no special insight into the problem, we ignore the fixed cost that the buyer incurs in GLS.

[4]For simplicity, Table 3.2 considers a zero salvage value as the low salvage value in GLS. All results and conclusions hold with all positive salvage values less than or equal to the salvage value in GLS that creates a preference for LC over DS.

[5]Note that the condition outlined in Lemma 1 depends on the total number of customers, since α represents the number of customers with boycotting behavior. In GLS, α represents the percentage of such customers.

Table 3.2: The buyer's profit with four possible sourcing policies in the case of salvage value = 0.

Sourcing policy	Buyer's profit	Optimal responsible sourcing	Optimal nonresponsible sourcing
Low cost (LC)	$(10{,}000 - \phi\alpha)V - 10{,}000C_{NR}$	0	10,000
Dual sourcing (DS)	$(10{,}000 - \theta)V + (\theta - \phi\alpha)(V + r) - (10{,}000 - \theta)C_{NR} - \theta C_R$	θ	$10{,}000 - \theta$
Responsible mass (RM)	$10{,}000V - 10{,}000C_R$	10,000	0
Responsible niche (RN)	$\theta(V + r) - \theta C_R$	θ	0

Thus, the relative attractiveness of LC compared with RM is determined by the strength of these two conflicting factors. Building on Lemma 1, we now derive the thresholds of the preferences between LC/RM and DS.

LEMMA 2. *The buyer prefers LC over DS when* $(\theta - \phi\alpha)(C_R - C_{NR} - r) > 0$ *(respectively,* $\theta(C_R - C_{NR}) - (\theta - \phi\alpha)r > 0$ *in the case of low salvage value).*

The intuition behind Lemma 2 is straightforward. Because the buyer is already sourcing from the nonresponsible supplier—the possibility of losing a portion of the market is already accepted—a decision on dual sourcing depends only on the profitability of market segmentation, which is determined by the sign of $(\theta - \phi\alpha)(C_R - C_{NR} - r)$ (respectively, $\theta(C_R - C_{NR}) - (\theta - \phi\alpha)r$ in the case of low salvage value).

LEMMA 2. *The buyer prefers RM over DS when* $\phi\alpha(V - C_{NR}) - 10,000(C_R - C_{NR}) + (\theta - \phi\alpha)(C_R - C_{NR} - r) > 0$ *(respectively,* $\phi\alpha V - 10,000(C_R - C_{NR}) + \theta(C_R - C_{NR}) - (\theta - \phi\alpha)r > 0$ *in the case of low salvage value).*

The first two terms in Lemma 3's expression represent the necessary condition for the dominance of RM to LC, and we know from Lemma 1 that these two terms should be positive so that the buyer prefers RM. Hence, this expression implies that preferring RM over DS is the same as first preferring RM over LC and then keeping this preference over DS as well. If the scale of extra benefit in market segmentation, i.e., $(\theta - \phi\alpha)(r - (C_R - C_{NR}))$ (respectively, $(\theta - \phi\alpha)r - \theta(C_R - C_{NR})$ in the case of low salvage value), is not large enough to cover the cost of possible market loss, the buyer will not prefer market segmentation and thus DS over RM.

Based on Lemmas 1-2, as GLS show, although large values of r and θ (the parameters that capture the encouraging reaction) in general favor responsible sourcing, such consumer reactions may also backfire and decrease the buyer's responsible sourcing when the encouraging reaction is strong enough to make DS more profitable than RM for the buyer. On the other hand, a high α (the parameter representing the discouraging consumer reaction) reduces nonresponsible sourcing more generally and thus improves responsible sourcing. However, for situations in which the buyer faces a low salvage value for unsold products (e.g., zero refund of the procurement cost as opposed to the full refund), while the effect of the encouraging reaction remains the same as without such costs, the discouraging reaction may now also backfire. As Lemma 2 shows, the counter effect of the discouraging reaction on responsible sourcing happens when the salvage value is low so that an increase in the discouraging reaction makes LC more profitable than DS for the buyer (since LC avoids high procurement costs in the excess inventory of unsold products).

The insights above do not allow us to come up with a straightforward answer to our main research question. We summarize these insights in three points as follows. (1) For low levels of the encouraging consumer reaction, this socially conscious reaction does not affect responsible sourcing in the supply chain because it cannot lead to a preference for DS over LC (or a preference for RM over DS). (2) For high levels of the encouraging reaction, this reaction can increase the tendency for responsible sourcing in the presence of a low discouraging reaction in the market by leading to a preference for DS over LC. However, the same levels of this reaction can lead to less responsible sourcing in the presence of a medium-to-high discouraging reaction because it will lead to a preference for DS over RM. (3) In the case of a low salvage value, increasing the discouraging reaction leads to less responsible sourcing because decision makers will prefer LC over DS.

3.3.3 Behavioral Theory

Lemmas 1-3 establish which sourcing strategy a rational decision maker prefers. To examine this decision from a behavioral perspective, we reformulate the GLS model by incorporating behavioral biases into the buyer's sourcing decision. Given the abundant evidence on dual sourcing in the context of responsible sourcing, the core of our behavioral model is to account for the buyer's tendency for dual sourcing due to the potential market segmentation that the presence of the encouraging consumer reaction induces. In other words, when facing two distinct customer segments in the market (i.e., socially conscious customers and regular customers), the buyer will be more likely to go with different suppliers for these different segments to create diversification and variety in its supplies and products.

Dual-sourcing bias has been previously studied in firms' sourcing decisions in supply chains. For example, Gurnani et al. (2014) found a tendency for dual sourcing in a buyer's sourcing decision that faced one reliable supplier and one unreliable supplier, even though single sourcing was the optimal policy. They explained this behavior using the diversification bias and variety-seeking behavior (Simonson 1990, Read and Loewenstein 1995, Benartzi and Thaler 2001). Variety seeking behavior occurs when people tend to choose more variety in combined choices of quantities of goods (here, procuring multiple products) than in sequential choices (here, procuring one product each time). Diversification bias extends this behavior to contexts other than consumption, for example, asset-allocation between investment options with uncertain returns, and further proposes that people diversify their choices evenly across possible options. Gurnani et al. (2014) see the root of the dual-sourcing bias they describe in these behavioral biases.

In our context, where different consumer segments can also be a source of diversification, we further relate the dual-sourcing bias to partition-dependence decision making (Fox and Clemen 2005, Fox et al. 2005). With partition-dependence decision making, various groupings of available options lead investment decisions towards diversification. Such diversifications exist even when hedging against uncertainty in an investment portfolio is irrelevant (Fox et al. 2005). We postulate that the market segmentation potential, which originates in an encouraging consumer reaction, leads to a diversification and dual-sourcing bias in our context. Without this possible market segmentation, the buyer will not exhibit a diversification bias as there is no supply uncertainty to hedge against. To formulate and incorporate this behavioral bias into the buyer's decision making, we characterize the amount of product sourced from either supplier in the presence of the encouraging reaction in the market as follows:

$$QN = (10,000 - \theta)\beta[1_{R^+}(A)]\theta - \beta[1_{R^+}(B)](10,000 - \theta) \tag{1a}$$

$$QR = \theta - \beta[1_{R^+}(A)]\theta + \beta[1_{R^+}(B)](10,000 - \theta) \tag{1b}$$

QN and QR represent sourcing quantities from the nonresponsible and responsible suppliers. The first terms on the right-hand side of Equations (1a)-(1b) are sourcing quantities related to the market segmentation, i.e., sourcing from the nonresponsible supplier to serve regular customers and sourcing from the responsible supplier to serve the socially conscious segment. The $1_{R^+}(x)$ is an indicator function (i.e., $1_{R^+}(x) = 1$ if $x \in R^+$, and $1_{R^+}(x) = 0$ otherwise), and A and B are the expressions outlined in Lemma 2 and Lemma 3, respectively. Hence, the indicator functions with A and B allow deviating from the dual-sourcing strategy toward less responsible sourcing (if $1_{R^+}(A) = 1$ and $1_{R^+}(B) = 0$) or toward more responsible sourcing (if $1_{R^+}(B) = 1$ and $1_{R^+}(A) = 0$) by replacing one supplier with another. The behavioral parameter $\beta < 1$ prevents these deviations from fully realizing, and thus represents the tendency of the buyer to focus on dual sourcing. With $\beta = 1$, there is no dual-sourcing bias and Equations (1a)-(1b) simplify to the outcome described by the classical model. With $\beta < 1$, the buyer will have a bias toward dual sourcing and will discount value that favors deviating from it. For instance, a $\beta = 0.2$ means that a dual-sourcing policy is five times as important for the buyer as any rational deviation from it.

In the absence of an encouraging reaction, where there is no market need for dual sourcing, Equations (1a)-(1b) will take the form of

$$QN = 10,000 + \beta[1_{R^+}(-A)]0 - \beta[1_{R^+}(-B)]10,000 \tag{2a}$$

$$QR = 0 - \beta[1_{R^+}(-A)]0 + \beta[1_{R^+}(-B)]10,000 \tag{2b}$$

when LC dominates RM (based on Lemma 1, and they will the form of

$$QN = 0 + \beta[1_{R^+}(-A)]10,000 - \beta[1_{R^+}(-B)]0 \tag{3a}$$

$$QR = 10,000 - \beta[1_{R^+}(-A)]10,000 + \beta[1_{R^+}(-B)]0 \tag{3b}$$

otherwise. The expressions A and B are the same as in Equations (1a)-(1b) originating from Lemma 2 and 3, respectively. Hence, the indicator functions with $-A$ and $-B$ allow deviating from single sourcing toward dual sourcing. A $\beta < 1$ here has a similar meaning to that in Equations (1a)-(1b) in the sense that the buyer is reluctant to deviate from LC (or RM). These allow us to keep a consistent structure for our reformulation. However, we expect a stronger bias (i.e., a lower β) in the presence of an encouraging reaction, because of the market segmentation potential created by such a consumer reaction.

In addition to the dual-sourcing bias, and in order to empirically compare our behavioral model with the classical model in the experimental study in Section 3.4, we also consider an error term for human decision making. This error term allows for deviations from the optimal order quantities due to random error in decision making (Ho et al. 2010). We capture this error through the indicator function. We assume that the indicator function $1_{R^+}(x)$ will take the value of $0+\Delta$ when $1_{R^+}(x) = 0$, and $1-\Delta$ when $1_{R^+}(x) = 1$, where $0 < \Delta < 1$ is the random error term. To systematically capture this error, we relate it to how the presence and absence of the encouraging reaction changes the dynamics of the effect of α. In the absence of the encouraging reaction (i.e., when $r = 0$ and $\theta = \alpha$), the buyer loses α customers out of the total 10,000 customers if the nonresponsible supplier is the source and a violation occurs. On the other hand, in the presence of the encouraging reaction (i.e., when there are customers with $r > 0$ and θ represents this customer segment, among whom α customers are willing to boycott as well), the buyer will lose α customers out of the θ customers. Hence, although the magnitude of the possible market loss remains the same, the presence of the encouraging reaction makes it seem proportionally bigger.

Therefore, we define the error term Δ based on $\frac{\alpha}{10,000}$ and $\frac{\alpha}{\theta}$ as follows. The $1_{R^+}(A) = 0$ takes the value of $0 + (\frac{\alpha}{10,000})^{\varepsilon}$ in the absence of the encouraging reaction, and $0 + (\frac{\alpha}{\theta})^{\varepsilon}$ in the presence of the encouraging reaction, with $0 < \varepsilon < 1$.[6] In accordance with this error, the $1_{R^+}(B) = 1$ takes the values of $1 - (\frac{\alpha}{10,000})^{\varepsilon}$ and $1 - (\frac{\alpha}{\theta})^{\varepsilon}$, respectively. Similarly, the $1_{R^+}(-A) = 1$ takes the values of $1 - (\frac{10,000 - \alpha}{10,000})^{1/\varepsilon}$ and $1 - (\frac{\theta - \alpha}{\theta})^{1/\varepsilon}$ in the absence and presence of the encouraging reaction, respectively. In accordance with this error, the $1_{R^+}(-B) = 0$ takes the values of $0 + (\frac{10,000 - \alpha}{10,000})^{1/\varepsilon}$ and $0 + (\frac{\theta - \alpha}{\theta})^{1/\varepsilon}$, respectively. These error terms keep $0 < \Delta < 1$.

Without behavioral parameters, i.e., when $\beta = 1$ and $\Delta = 0$, the behavioral model is equivalent to the classical model. In this case, there is no dual-sourcing bias in the buyer's sourcing decision and the buyer's sourcing policy is rational with no error in the optimal sourcing quantities. Hence, we cannot expect any behavioral influences from the encouraging consumer reaction. With behavioral parameters, the behavioral model changes the classical model's predictions as we outline in the following two propositions. The key difference between the behavioral and the classical model, originating from the behavioral parameters, is to make the buyer prefer dual sourcing in the presence of the encouraging reaction.

PROPOSITION 1. *The presence of an encouraging reaction by consumers, regardless of magnitude, is always (never) helpful with responsible sourcing when the discouraging reaction is weak (strong).*

The intuition behind Proposition 1 lies in the fact that the presence of the encouraging reaction (through even small amounts of extra willingness-to-pay) in the market induces a clear market segmentation for the buyer and leads its sourcing policy toward dual sourcing. When the discouraging reaction in the market is weak, this dual-sourcing bias is helpful with responsible sourcing as it avoids the buyer's low-cost sourcing (i.e., sourcing solely from the nonresponsible supplier). On the other hand, when the discouraging reaction is

[6] In the interest of simplicity in the parameter estimation procedure, we do not consider different ε's for each participant. Thus, the error term here reflects a uniform error in human decision making.

strong, the dual-sourcing bias does not help with responsible sourcing as it impedes the buyer's responsible-mass sourcing (i.e., sourcing solely from the responsible supplier).[7] Our behavioral model, therefore, modifies the classical model's prediction by relating the effect of the encouraging reaction on responsible sourcing to its presence independent of its magnitude. When the encouraging reaction in the market is high enough so that the classical model predicts dual sourcing, the behavioral model and the classical model are in agreement on the effect of the encouraging reaction on responsible sourcing, though through different dynamics. Therefore, Proposition 1 modifies the first two points summarized earlier, showing that the encouraging reaction is always (never) helpful with responsible sourcing when the discouraging reaction in the market is weak (strong).

PROPOSITION 2. *The discouraging reaction is always helpful with responsible sourcing irrespective of the product's salvage value.*

The intuition behind Proposition 2 is similar to that of Proposition 1. The presence of the encouraging reaction in the market creates a dual-sourcing bias in the buyer's sourcing decision that prevents a detrimental preference for LC over DS—a preference which the classical model predicts when the product's salvage value is low. Hence, the behavioral model modifies the prediction of the classical model and establishes that the discouraging reaction cannot be detrimental to responsible sourcing in the case of a low salvage value (even with salvage value=0). For all other cases, i.e., medium-to-full salvage values, the behavioral model is in agreement with the classical model, and both favor the positive effect of the discouraging reaction on responsible sourcing. Therefore, Proposition 2 modifies the third point summarized earlier by showing that the discouraging reaction can never be detrimental to responsible sourcing, even if the product salvage value is low.

Propositions 1 and 2 together help us reach to a straightforward answer to our main research question: An encouraging reaction (by even just a small extra willingness-to-pay) leads to more responsible sourcing for hard-to-substitute products (i.e., for products where

[7] A weak (strong) discouraging reaction refers to all levels of this reaction lower (higher) than or equal to the 'low' ('high') value in GLS that creates a preference for LC over RM (and vice versa).

discouraging reactions are most likely low). However, an encouraging reaction will lead to less responsible sourcing for products that are easily substitutable (i.e., for products where discouraging reactions are potentially high). Regardless of product type, creating a discouraging reaction among customers always increases responsible sourcing, although it may be easier to accomplish such a consumer response for products with numerous substitutes. Therefore, in supply chains of products with no (many) substitutes, the encouraging (discouraging) reaction is the most effective consumer reaction to support responsible sourcing in the supply chain.

3.4 Behavioral Experiment

In this section, we use data from a behavioral experiment to validate our proposed behavioral model by comparing its predictions with those of the classical model. This comparison is centered on the three insights from the GLS model outlined earlier, where predictions from our behavioral model differ from GLS leading to Propositions 1 and 2. Hence, by comparing our behavioral model to the rational decision-making model, we also validate those propositions: Is there a transition point at which the presence of the encouraging reaction becomes helpful with (detrimental to) responsible sourcing when the discouraging reaction is low (high)? And can the discouraging reaction be detrimental to responsible sourcing when the salvage value is low?

Table 3.3 outlines our experimental design. To have a solid comparison, we first create control groups in which the behavioral and classical models predict the same sourcing policy. Then we create treatment groups by manipulating the model parameters so that they result in conditions in which the classical and behavioral models lead to different predictions. In the first two scenarios, we test the behavioral effects of the presence of the encouraging reaction in accordance with Propositions 1. The control groups in these scenarios simply remove the encouraging reaction altogether so that the behavioral model and the classical model have the same predictions. The treatment groups add the encouraging reaction, but keep this parameter at values much below the level with which GLS will not predict dual sourcing, so that the different predictions from the behavioral and classical models are due to the mere presence of the encouraging reaction. In the third scenario (that tests the detrimental effect of increasing the discouraging reaction in the case of low salvage value), the control group keeps the discouraging reaction low enough for the behavioral and classical models to predict the same. The treatment group shifts the discouraging reaction up to a level much higher than what GLS need to predict LC in the case of low salvage value, which leads to different predictions from the two models in accordance with Proposition 2. In all scenarios, all other parameters remain the same for both treatment and control groups.

Table 3.3: Experimental groups.

Scenario	Control group	Treatment group
Preference for LC over DS	No encouraging reaction	With encouraging reaction
GLS[1] prediction	LC	LC
BM[2] prediction	LC	DS
Preference for RM over DS	No encouraging reaction	With encouraging reaction
GLS prediction	RM	RM
BM prediction	RM	DS
Preference for LC over DS (the case of low salvage value)	With encouraging reaction Low discouraging reaction	With encouraging reaction High discouraging reaction
GLS prediction	DS	LC
BM prediction	DS	DS

[1]Classical Model of GLS, [2]Behavioral Model

Because participants in each group will repeat the same task for several rounds, we draw randomly chosen parameter values from ranges of values for each parameter so that the participants repeat the task with different parameter values in each round. Therefore, in each scenario, the parameters have ranges, and these ranges change between the control and treatment groups according to the logic outlined above (Table B1 in Appendix B shows the ranges of all parameters used in the experiment).

3.4.1 Experimental Procedure

We recruited 219 participants via Amazon Mechanical Turk (AMT)[8] to perform the role of buyers. Participants received a flat \$0.01 participation payment and a performance-based payment based on the total profit each participant made in the experiment. The experiment was coded on the SoPHIE platform (Hendriks 2012). The computer played the roles of suppliers and customers. We used payment adjustments so that the performance-based payments fell into comparable ranges for all experimental groups (see Table B1 in Appendix B). Average payment was \$2.50 (STD=0.739), with a \$0.02 minimum and a \$3.91 maximum payment. Of the participants, 50.2% were male, 49.3% were female, and one participant preferred not to indicate their gender. The average age was 37.6 (STD=11.04). The average income was between \$25,000 and \$50,000. Most participants had some college education: 37.4% had some college credits, and 41.1% reported having earned a bachelor's degree.

Upon accepting to participate in the experiment, participants were randomly assigned to one of the six experimental groups (as in Table 3.3). Before starting the experimental task, participants were asked to watch an instructional video and were told that doing so was necessary to perform well and obtain profits in the task. They were also told that they would repeat the experimental task for 20 rounds and would be paid based on the total profit they made in all rounds. The instructional video narrated the descriptions for the participants and lasted 100-120sec, depending on the experimental group (full descriptions of the experimental tasks are available in Appendix B). We closely monitored whether the participants were going through the instructional video. In the recruitment process, 21 participants dropped out at the instructional video, and 15 participants went to the experimental task without finishing the instructional video. These participants dropped out after

[8] Amazon Mechanical Turk (AMT) is a popular avenue for subject recruitment, and several studies have endorsed AMT as a reliable source of experimental data (see Lee et al. 2018 and references therein).

a couple of rounds, and thus were not included in the final pool. Table 3.4 shows the final number of participants in each experimental group. We note that 97.3% of the participants agreed/strongly-agreed that the description and the video instruction were clear. In addition, at the end of the experiment, participants were asked to indicate if any part of the description was unclear or confusing to them. 95% of the participants found none of the parts confusing. Inclusion/exclusion of the other participants (2.7% and 5%) in the analyses did not make any difference in the results. Participants observed the same description of the task in each round with different parameter values randomly drawn from the parameter value intervals discussed earlier.

Table 3.4: Number of participants.

Scenario	Control group	Treatment group
Preference for LC over DS	38	35
Preference for RM over DS	36	36
Preference for LC over DS (the case of low salvage value)	36	38

In each round, participants determined their sourcing decisions, i.e., sourcing quantities from the responsible and nonresponsible suppliers. They were shown the profit they made with that sourcing decision (according to the profit formulations in Tables 3.1 and 3.2) along with the total profit they had made so far. They then continued with the next round. Upon completion of all 20 rounds, participants were shown the dollar value of their earnings,[9] and they finished the experiment by answering questions on the clarity of the task (as mentioned above) and supplying the demographic information. All results were independent of the

[9]Note that we compensate participants based on the expected profits resulting from their decisions; we do not implement a random draw that determines whether a violation occurs or not in the experiment.

participants' demographics, such as gender and age.

3.4.2 Preliminary Results

Table 3.5 shows preliminary results for all experimental groups. Because our problem setting did not incorporate supply uncertainty, we did not expect to see any oversourcing (Gurnani et al. 2014, Kalkanci 2017). In line with this expectation, neither the mean nor the median of total order (QN+QR) differs statistically from the total demand (i.e., 10,000) in any of the groups. We calculate the average order quantity for each participant and use t-tests and Wilcoxon rank-sum tests, respectively, for equality of observed means and medians with predicted values. As one can see in Table 3.5, the classical model of GLS fails to predict the observed sourcing policy in terms of order quantities. The only exception is the case when DS is the optimal policy, and thus the classical model and the behavioral model largely coincide in their prediction.

In addition to the order quantities, we also study the ratio $\frac{\text{QR}}{\text{QN+QR}}$ as a measure of diversification (Gurnani et al. 2014). As Gurnani et al. (2014) discuss, with optimal single-sourcing, the theoretical value of this measure is either zero or one, and hence, an observed value close to 0.5 indicates a diversification bias. As seen in Table 3.5, the presence of the encouraging reaction in our treatment groups pushes this measure closer to 0.5. Without the encouraging reaction, this measure is closer to 1 and 0; this was expected because there was no supply uncertainty in our setting to cause the kind of diversification bias reported by Gurnani et al. (2014).

Another important observation lies in the overall tendency of change in the amount of nonresponsible sourcing (QN) between the control and treatment groups in each scenario. As seen in Figure 3.1, these changes are far from the classical model's predictions and are in line with the behavioral model's predictions.

Table 3.5: Observed order quantities and predictions of the classical model.

			Scenario 1		Scenario 2		Scenario 3	
			Control	Treatment	Control	Treatment	Control	Treatment
Total order	Observed	Mean	9,928	9,301	10,088	9,921	9,972	10,053
		Median	10,000	10,000	10,000	10,000	10,000	10,000
	Predicted (GLS)	Mean	10,000	10,000	10,000	10,000	10,000	10,000
		Median	10,000	10,000	10,000	10,000	10,000	10,000
QN	Observed	Mean	6,634***	3,739***	2,880***	3,252***	4,334	3,481***
		Median	8,068***	3,970***	1,602***	3,585***	4,265	3,728***
	Predicted (GLS)	Mean	10,000	10,000	0	0	4,200	10,000
		Median	10,000	10,000	0	0	4,200	10,000
QR	Observed	Mean	3,294***	5,562***	7,207***	6,669***	5,638	6,572***
		Median	1,932***	5,700***	8,510***	6,278***	5,742	5,868***
	Predicted (GLS)	Mean	0	0	10,000	10,000	5,799	0
		Median	0	0	10,000	10,000	5,800	0
$\frac{QR}{QR+QN}$	Observed	Mean	0.33***	0.59***	0.71***	0.70***	0.56	0.66***
		Median	0.19***	0.58***	0.82***	0.64***	0.57	0.58***
	Predicted (GLS)	Mean	0	0	1	1	0.58	0
		Median	0	0	1	1	0.58	0

Note. *** *indicates* $p_{value} \leq 0.001$ *in t-test and Wilcoxon rank-sum test, respectively, for equality of observed means and medians with predicted values.*

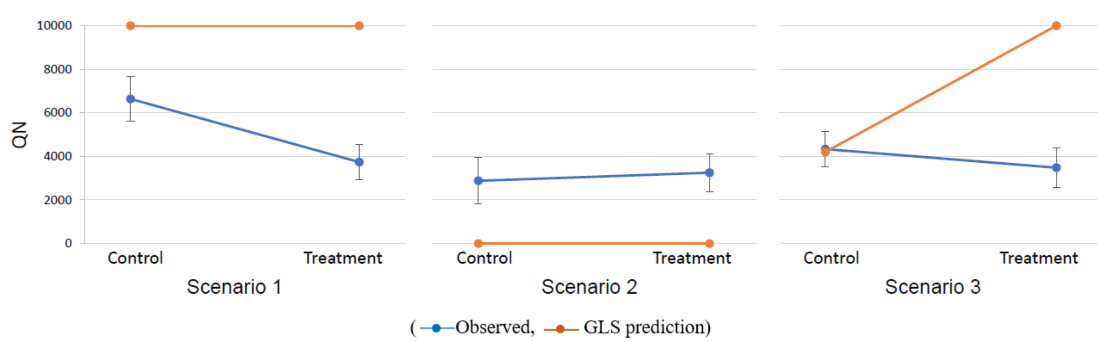

Figure 3.1: Nonresponsibly sourced products—observed versus predicted by the classical model. *Note: Error bars represent 95% confidence intervals.*

3.4.3 Behavioral Predictions

Having established that the classical model of GLS does not fit the experimental data well, we now examine whether the behavioral model can fit with the observed results in Table 5 and Figure 1. We do so by using statistical methods to estimate the behavioral parameters of the model. We follow the approach of Ho et al. (2010) and use maximum likelihood estimation assuming each participant's order quantities follow a normal distribution with the behavioral model's prediction as the structural mean. Thus, the objective is to estimate the unknown parameters of the behavioral model through estimating the structural mean. Similar to the approach of Ho et al. (2010), the parameters are specified to be common across all experimental groups while the mean structure changes for each group according to Equations (1a)-(3b). That is, we consider uniform values for the behavioral parameters across all groups. This keeps the focus on examining whether the extended behavioral model is more powerful than the existing rational model in terms of predicting larger ranges of data rather than on finding the best fit and parameter values for each group (Rabin 2013a,b).

Because our setting contains two decision variables from participants, i.e., QN and QR, we use a bivariate normal distribution to simultaneously estimate the means of both variables. This ensures that our estimation procedure takes into account variations in both extracted variables. Based on our discussion in Section 3.3.3, we define the full model here as one with different β's depending on whether the encouraging consumer reaction is present or not. Each of these β's remains the same across all scenarios and groups. Also, we consider one ε for all scenarios and groups to retain consistency in error terms across all experimental groups.

Table 3.6 shows the results of the estimation. For the estimated β's (β_{enc} and $\beta_{no\,enc}$ are the estimated values in the presence and the absence of the encouraging reaction, respectively), we use a t-test to test if they differ statistically from 1. The standard errors are clustered at the subject level to account for within-subject correlation in ordering decisions (Ho et al. 2010). We cannot perform a t-test for ε because there is no null-value for this parameter (the classical model removes the whole error term Δ rather than setting its ε to a null-value). We thus use a likelihood ratio test to statistically test its effect in the model.[10] The second column of Table 3.6 shows the estimated parameters, β and ε. As seen, β_{enc} is significantly less than $\beta_{no\,enc}$, which means the dual-sourcing bias in the presence of the encouraging reaction is significantly stronger than that in the absence of such a consumer reaction. The third and fourth columns test the effect of the two behavioral parameters (β and ε) in the model by using a likelihood ratio test. Both parameters improve the model fit at the 0.001 significance level (note that when the error term is removed from the model, the first β loses its effect in Equations (2a)-(3b), and its estimated value remains at an initial value). The last column repeats the same analysis for the fully reduced form of the behavioral model, which is in fact the classical model.

The estimation results here show that the behavioral model statistically outperforms the classical model and that both behavioral parameters are essential in obtaining better predictions. Next, we check if the behavioral model, with the estimated uniform values for the behavioral parameters, can consistently explain the observed data (in 95% confidence intervals) in all scenarios and groups. Table 3.7 shows the predictions of the behavioral model. The behavioral model provides consistent predictions (in 95% confidence intervals) for all scenarios and groups. Furthermore, as Figure 3.2 depicts, the behavioral model fits the observed nonresponsible sourcing quantities across different scenarios and groups very well.

3.4.4 Robustness Checks

We conducted robustness checks to examine whether (a) a uniform dual-sourcing bias and (b) a naïve diversification (Benartzi and Thaler 2001) can also explain the observed data. We also further examine whether the presence of an encouraging consumer reaction in fact

[10] Likewise, we cannot perform a Wald test for ε because of the lack of a null-value for the parameter. To be consistent, we use a likelihood ratio test for β as well. However, we check a Wald test for β to ensure that the dependence of observations from the same participant in the data does not change the concluded results, and the outcome is in line with the likelihood ratio test (same as in Ho et al. 2010 and Kalkanci 2017).

Table 3.6: Parameter estimation of the behavioral model.

Behavioral parameters	Full BM	BM without β	BM without ε	BM without β and ε (equivalent of GLS)
$\beta_{no\ enc}$	0.76*** (0.025)	—	0.50 (4.320)	—
β_{enc}	0.31*** (0.098)	—	0.01*** (0.013)	—
ε	0.43 (0.148)	0.48 (0.051)	—	—
-LL	86,593	86,959	87,128	88,467
Likelihood ratio test and p-value (against the full BM)	—	$\chi^2(2) > 100$ p<0.001	$\chi^2(1) > 100$ p<0.001	$\chi^2(3) > 100$ p <0.001

Note. Numbers in parentheses are standard errors clustered at the subject level.
*Significance levels are based on t-tests against 1 (***=0.001, **=0.01, and *=0.05). No t-test is performed for ε.*

Table 3.7: Observed order quantities and predictions of the behavioral model.

			Scenario 1		Scenario 2		Scenario 3	
			Control	Treatment	Control	Treatment	Control	Treatment
Total order	Observed	Mean	9,928	9,301	10,088	9,921	9,972	10,053
		Median	10,000	10,000	10,000	10,000	10,000	10,000
	Predicted (BM)	Mean	10,000	10,000	10,000	10,000	10,000	10,000
		Median	10,000	10,000	10,000	10,000	10,000	10,000
QN	Observed	Mean	6,634	3,739	2,880	3,252	4,334	3,481
		Median	8,068	3,970	1,602	3,585	4,265	3,728
	Predicted (BM)	Mean	6,643	4,332	2,816	3,330	4,454	3,186
		Median	6,652	4,359	2,815	3,324	4,457	3,191
QR	Observed	Mean	3,294	5,562	7,207	6,669	5,638	6,572
		Median	1,932	5,700	8,510	6,278	5,742	5,868
	Predicted (BM)	Mean	3,357	5,668	7,183	6,670	5,545	6,814
		Median	3,348	5,641	7,185	6,676	5,543	6,809
$\frac{QR}{QR+QN}$	Observed	Mean	0.33	0.59	0.71	0.70	0.56	0.66
		Median	0.19	0.58	0.82	0.64	0.57	0.58
	Predicted (BM)	Mean	0.34	0.57	0.72	0.67	0.55	0.68
		Median	0.33	0.56	0.72	0.67	0.55	0.68

Note. In all cells, p_{value}>0.1 in t-test and Wilcoxon rank-sum test, respectively, for equality of observed means and medians with predicted values.

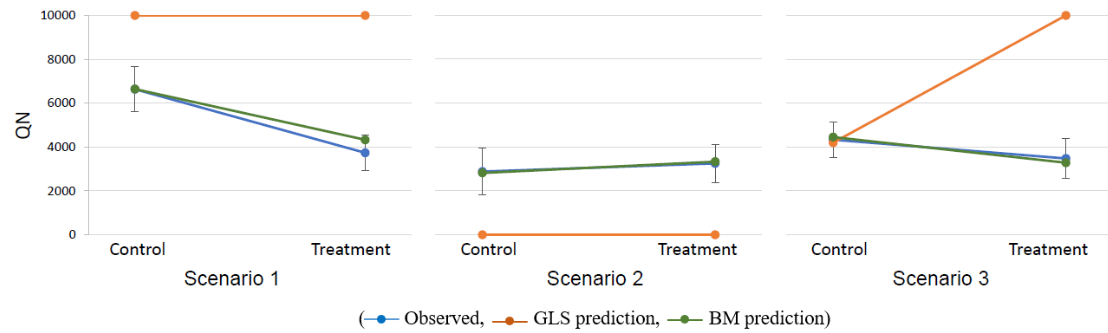

Figure 3.2: Nonresponsibly sourced products—observed versus predicted by the classical and behavioral models. *Note: Error bars represent 95% confidence intervals.*

induces a dual-sourcing bias.

For the first robustness check, we repeat our parameter estimation considering a uniform β instead of the split β_{enc} and $\beta_{no\,enc}$. For the second robustness check, we repeat the estimation assuming that the buyer pursues a half-half dual-sourcing policy due to having two options (i.e., suppliers) to source the product from. This naïve diversification ignores the role of the market segmentation induced by an encouraging consumer reaction and only looks at the diversification from the biases on the supply side, similar to the diversification bias discussed by Gurnani et al. (2014) for dual sourcing in their problem. Finally, for a third robustness check, we repeat our estimation without assuming that the presence of an encouraging consumer reaction induces a dual-sourcing bias. That is, we assume that the buyer's sourcing decision follows Equations (1a)-(1b) only when dual sourcing is the classical optimal, and otherwise it is Equations (2a)-(2b) and (3a)-(3b) that define the buyer's sourcing decision.

Table 3.8 presents the results of our robustness checks. The third column shows that the model with a uniform β does not work as well as the full model with different β's in covering the salience of the dual-sourcing bias in the presence of the encouraging reaction (and the market segmentation induced by that). The fourth column shows the result of the estimation using the naïve diversification model explained above. Comparing the outcome of this estimation with that of the full behavioral model (presented in the second column), we find that the naïve diversification is not able to explain the observed data better than the model that relates the dual-sourcing bias to the market segmentation. We use the Akaike Information Criterion (AIC) for this comparison because the two models contain the same parameters while changing the structure of the model. The relative likelihood of the naïve diversification model against the full model, calculated by $e^{(AIC_0 - AIC_{model})/2}$ (Burnham and Anderson 2004), shows that this model is 9.85×10^{-34} as probable as the full model to minimize the information loss in the maximum likelihood estimation.

Table 3.8: Robustness checks on the nonuniformity of the dual-sourcing bias and the market induced dual-sourcing.

Behavioral parameters	Full BM	BM with equal $\beta's$	BM with naïve divers.	BM without market induced dual sourcing
$\beta_{no\ enc}$	0.76*** (0.025)	0.56*** (0.001)	0.99 (0.078)	0.81*** (0.001)
β_{enc}	0.31*** (0.098)	0.56*** (0.001)	0.40*** (0.190)	0.26*** (0.007)
ε	0.43 (0.148)	0.44 (0.001)	0.42 (0.097)	0.46 (0.001)
-LL	86,593	86,675	86,669	86,815
Likelihood ratio test and p-value (against the full BM)	—	$\chi^2(1) > 100$ p<0.001	—	—
Relative likelihood of the model (against the Full BM)	—	—	9.85×10^{-34}	3.86×10^{-97}

Note. Numbers in parentheses are standard errors clustered at the subject level.
*Significance levels are based on t-tests against 1 (***=0.001, **=0.01, and *=0.05). No t-test is performed for ε.*
Relative likelihood is calculated by $e^{(AIC_0 - AIC_{model})/2}$, *in which AIC is the Akaike Information Criterion.*

The last column shows the result of the estimation that ignores any dual-sourcing induced by the presence of the encouraging reaction unless it is the classical optimal policy (note that here β_{enc} is for when there is an encouraging reaction and dual sourcing is the classical optimal policy, and $\beta_{no\,enc}$ is for the otherwise). We use a similar comparison (using the AIC) here and find that this model is not as powerful as our behavioral model in explaining the observed data (the model is 3.86×10^{-97} as probable as the full behavioral model to minimize the information loss in the maximum likelihood estimation). It is also worth noting that, in contrast to what we obtain by our full behavioral model in Table

3.7 (i.e., nonsignificant differences between the predicted and observed values), none of the alternative models tried here or presented in Table 3.6 could provide such consistent predictions (in 95% confidence intervals) for all scenarios and groups.

The result of the estimation for the last robustness check above uncovers a meaningful insight. When the dual-sourcing is the (classical) optimal policy, it seems to be stronger than the dual-sourcing that is based on the presence of the encouraging reaction regardless of it being the classical optimal policy or not (0.26 compared with 0.31). Therefore, as another robustness check, we further look at more classifications for the parameter β. In our model, we have tied its value to the presence and absence of the encouraging reaction (and the market segmentation induced by that). Here, we tie its value to how likely the induced market segmentation is to occur. In other words, the magnitude of the extra willingness-to-pay r can determine how valuable the market segmentation by θ and $(10,000-\theta)$ is for the buyer, and thus can implicitly determine the weight of the dual-sourcing bias for the buyer.

The best way to consider the magnitude of r is through its magnitude relative to $C_R - C_{NR}$ because it determines the value of r. In our experimental setting, in the interest of keeping the classical optimal policies unchanged between the control and treatment groups, we have $r < C_R - C_{NR}$ for when LC is the classical optimal, r close to $C_R - C_{NR}$ for when RM is the classical optimal, and $r > C_R - C_{NR}$ for when DS is the classical optimal. This can expand our previous classification for β as follows: β in the absence of the encouraging reaction, β in the presence of the encouraging reaction and when LC is the classical optimal, β in the presence of the encouraging reaction and when RM is the classical optimal, and β in the presence of the encouraging reaction and when DS is the classical optimal. Hence, we repeat our parameter estimation using four β's reflecting these four cases. Table 3.9 shows the result of this robustness check. As seen, the values of β's correspond to the relative magnitudes of r. Nonetheless, the likelihood ratio test is only barely significant. A Wald test also shows nonsignificant differences compared to the full model. The result of this robustness check shows that no further classification of the parameter β is necessary for the behavioral model to explain the observed data.

Table 3.9: Robustness check on the classification of the parameter β.

Behavioral parameters	BM with four $\beta's$ (new full BM)	BM with two $\beta's$ (previously established full BM)
$\beta_{no\ enc}$	0.76^{***} (0.030)	0.76^{***} (0.025)
β_{enc,r_L}	0.47^{***} (0.111)	0.31^{***} (0.098)
β_{enc,r_M}	0.34^{***} (0.031)	0.31^{***} (0.098)
β_{enc,r_H}	0.24^{***} (0.005)	0.31^{***} (0.098)
ε	0.43 (0.016)	0.43 (0.148)
-LL	86,590	86,593
Likelihood ratio test and p-value (against the new full BM)	—	$\chi^2(2) = 6$ p>0.049
Wald test and p-value (against the new full BM)	—	$\chi^2(2) = 3.72$ p>0.155

Note. Numbers in parentheses are standard errors clustered at the subject level.
*Significance levels are based on t-tests against 1 (***=0.001, **=0.01, and *=0.05). No t-test is performed for ε.*

3.5 Conclusion

In this paper, we developed a behavioral model of a firm's responsible sourcing to investigate the effect of a market's socially responsible reactions to firms' sourcing decisions. Building on the stylized model and framework of GLS, our problem setting included a "buyer" (i.e., the manufacturer), a "responsible supplier," and a "nonresponsible supplier," and focused on two socially responsible reactions: customers' extra willingness-to-pay for responsibly sourced products (as an encouraging reaction) and their willingness-to-boycott the buyer in the event of a social/environmental violation on its suppliers' side (as a discouraging reaction). The core of our behavioral model, which was validated via an experimental study, was to incorporate an important behavioral consideration: In the presence of an encouraging reaction by consumers, the buyer has a tendency for dual-sourcing bias due to the perceived potential for market segmentation. Our behavioral reformulation, incorporating this behavioral bias into the classical formulation, modified some of the predictions of the classical model and lead to very different predictions of the buyer's sourcing decisions.

With these modifications, our analyses show that: (1) the encouraging reaction, regardless of its magnitude, is always (never) helpful with responsible sourcing when the discouraging reaction is (not) sufficiently prevalent in the market, (2) the discouraging reaction is always helpful with responsible sourcing regardless of the product's salvage value. We obtained these findings by investigating scenarios in which the classical model yields inconsistent predictions regarding the effect of the two socially responsible reactions. These findings enable us to reach to a more consistent and straightforward answer to the main question in this context (i.e., which type of consumer reaction is more effective in improving responsible sourcing by firms in different supply chains and industries?): (1) in supply chains of products that do not have perfect substitutes, the creation and expansion of an encouraging reaction among customers is always sound policy, (2) in supply chains of products with many substitutes in which an encouraging reaction is not helpful with responsible sourcing, promoting and enlarging a discouraging reaction among customers is preferable, and (3) enhancing a discouraging reaction in the market is always good practice, irrespective of the supply chain type, although doing so may be easier in supply chains of products with many substitutes.

These recommendations are meaningful for practice and are in line with what we can expect from customers if we consider the nature of the products. For products that have several substitutes in the market, it is unrealistic to expect higher willingness-to-pays from customers for a product they can buy from other manufacturers. Therefore, promoting a discouraging reaction among customers becomes the priority. On the other hand, for products without perfect substitutes, it is realistically possible to sell them at higher prices by promoting their responsibly sourced attributes, a situation that gives an encouraging reaction powerful leverage. Meanwhile, although a discouraging reaction can also be helpful here, it may not be easily achievable because customers lack other options. Thus, our behavioral model yields recommendations that can be implemented by NGOs trying to promote proper socially responsible behavior among customers in different supply chains to improve firms' responsible sourcing.

Finally, given some level of similarity between the problem of our context and that of firms' sourcing from an unreliable (i.e., disruption prone) and a reliable supplier, our results can also have implications for manufacturers' management of their supply uncertainty. A discouraging reaction, when realized, has an indirect impact on the supply side by determining how many products a manufacturer can sell in the market. This can be similar to a supply uncertainty that also determines manufacturers' potential sales. Our findings show that responsible sourcing can be a viable way to manage indirect supply uncertainty, in line with other supply-risk mitigation efforts, such as dual sourcing and process improvement (Wang et al. 2010). However, our analyses highlight when and how this supply-risk mitigation can be pursued by structuring supply chains for socially responsible behavior. A discouraging reaction, although recommended for any supply chain, is very likely to take effect only in supply chains of products with many substitutes. Therefore, manufacturers may perceive responsible sourcing as supply-risk mitigation only in supply chains of this product category.

CHAPTER 4

CHARGING FOR RECYCLING AND CONSUMER ADOPTION OF GREEN ELECTRONICS

4.1 Introduction

Waste Electrical and Electronic Equipment (WEEE), also called e-waste, is a complex and fast-growing waste stream. Worldwide, e-waste is growing at an annual rate of 4-5% (Balde et al. 2015). Of the various options for managing e-waste, recycling has proven to yield the most environmental benefits (Demeester et al. 2013). The inherent complexity of e-waste recycling, however, has been a burden to recycling operations and a discouraging factor to practitioners: "There is almost nothing as hard to recycle as electronics" (Securis 2017). Meanwhile, regulators and legislators in a majority of countries have mandated that manufacturers collect WEEEs from their customers free of charge. While such policies have led to noticeable growth in recycling rates in some industries (e.g., steel and ferrous recycling), they have not done the same in others (e.g., plastic and glass recycling). In these industries, where recycling costs are relatively higher, recycling rates remain low (Demeester et al. 2013, US EPA 2016). The high costs of recycling core materials have kept the recycling of e-waste far below an ideal point (Balde et al. 2015). This raises a debate as to whether policies and environmental laws that insist on free recycling are effective in increasing e-waste recycling rates over policies that would allow charging for recycling to assist recycling systems with operational costs.

The first state in the U.S. to pass an e-waste recycling law was California in 2003. Since then, 25 states have passed e-waste recycling legislation.[1] All states except California and Utah use the Extended Producer Responsibility (EPR) approach, whereby man-

[1] http://www.electronicstakeback.com/promote-good-laws/state-legislation/

ufacturers take financial responsibility for recycling their old products. Utah's law does not require collection and only focuses on recycling education. California's Electronic Waste Recycling Act of 2003 charges customers a small recycling fee, similar to a tax, on each purchase.[2] Recently, Best Buy implemented a recycling program that charges $25 for accepting TVs/monitors for recycling. The new recycling program is implemented in all states except Illinois and Pennsylvania, where state laws bar retailers from charging for recycling (Best Buy has shut down its recycling program in these states). Each of these policies—free recycling, recycling tax,[3] and charging for recycling—has its own advantages and drawbacks. From the point of assisting with operational costs, charging for recycling can directly help manufacturers and recyclers manage their recycling costs based on their needs and can thus improve recycling rates of collected e-waste. Meanwhile, free recycling (and a recycling tax policy, for that matter) hopes to keep the volume of product return at higher levels because charging can deter customers from returning their products for recycling. Although many states in the U.S. have enacted bans on landfilling and/or incinerating some types of e-waste (such as TVs, monitors, desktops, laptops, and tablets),[4] which limits households' options in disposing of their used electronics, the main rationale in a free recycling policy is to keep the product return a convenient option for customers. From this point of view, Best Buy's new charging-for-recycling program is a counterintuitive policy that does not keep the same priority and focus on increasing the volume of e-waste return. While the above debate is an obvious one, how charging for recycling can potentially impact the efficiency of e-waste recycling in a bigger picture is worth uncovering.

In this paper, we look at the potential benefits of charging for recycling beyond helping with the operational costs. Specifically, we explore whether charging for recycling, com-

[2] https://www.calrecycle.ca.gov/electronics/retailer

[3] Although with California's Electronic Waste Recycling Act retailers are required to remit the collected fee to the California Department of Tax and Fee Administration, the fee is posed as a recycling fee, rather than a tax, at the purchase. Nonetheless, we refer to this policy as a recycling tax policy in the interest of making a better distinction from the charging-for-recycling program.

[4] http://www.electronicstakeback.com/wp-content/uploads/Disposal_Ban_Bills.pdf

pared with free recycling and recycling tax policies, can result in higher consumer adoption of recyclable electronics. Recently, there have been serious efforts to decrease the recycling costs of e-waste through improving the recyclability of consumer electronics (for example, EPEAT certification, which emphasizes recyclability aspects) and encouraging retailers to market and sell more recyclable and greener electronics (for example, Energy Star Partner of the Year recognition for retailers by the EPA in the U.S.).[5] However, evidence has shown that direct green marketing to increase consumer adoption of green products has not been successful (see, e.g., Allcott and Taubinsky 2015, Allcott and Sweeney 2017). Hence, retailers usually do not find investing in green marketing financially attractive (Deutsch 2010a, Deutsch 2010b, Allcott and Sweeney 2017).

We postulate that Best Buy's charging-for-recycling program can be a viable alternative that could potentially increase consumer adoption of green (recyclable)[6] electronics by indirectly nudging their attention towards the fact that recycling is costly, and thus making recyclability a valuable attribute, and by improving their perceived self-efficacy to take this eco-friendly action. Indirect resonation of "recyclability is valuable" by "recycling is costly" at the purchase point lies in the concept of green nudges (Sunstein 2009), which aims to enhance people's intentions toward environmentally friendly actions. With the long standing practice of free recycling, "recycling is costly" may never effectively resonate with consumers at the purchase point, despite the growing concerns about high recycling costs among recyclers (Demeester et al. 2013, Securis 2017). This sort of recycling charge, in addition, can provide customers a proxy for the consequences of not buying a recyclable product and thus improve their self-efficacy in taking an environmental action on their side, lack of which has been proven to be a hindrance to consumers' green purchases (see, e.g., Ellen et al. 1991, Berger and Corbin 1992, Rice et al. 1996).

To gain insights into this aspect of Best Buy's new program, through controlled experi-

[5] https://www.epeat.net/resources/criteria-2/ and https://www.energystar.gov/about/awards

[6] In general, the criteria of "green" products focuses on three main aspects: products being free of toxic compounds, being energy efficient, and being made of recyclable and renewable materials (Speer 2012). Here, our focus is on the recyclability aspect, and we thus refer to the recyclable version as the green version.

mental studies, we explore whether such a charging-for-recycling scenario, compared with free recycling and recycling tax scenarios, can lead to a relatively higher level of adoption of recyclable versions among customers (see, Shogren 2002, on using experimental methods for guiding environmental and climate policy). We use a subject-level measure for the adoption of the green (recyclable) version over the standard (non-recyclable) version in terms of the subject-level percentage of extra willingness to pay (WTP) for the Green version. In line with previous studies on consumer durables replacements (see Guiltinan 2010 for a review) and emergency purchases (Samson and Voyer 2014), we study regular (or new) purchase situations and emergency (or replacement) purchase situations separately as they have been proven to have different decision-making dynamics. We find that, compared with free recycling and recycling tax scenarios, charging for recycling increases regular buyers' adoption of green electronics by around 5%. For emergency buyers, the adoption level remains the same as with the current recycling policies, in line with previous findings that emergency buyers' decision-making situation hinders them from paying for environmental attributes (Samson and Voyer 2014).

Our findings have clear implications for green marketing, operational aspects of recycling, and environmental policy. The most promising finding in the adoption of energy-efficient products has yielded an effect size of around 5%, while requiring both incentives to salespeople and rebates to customers (Allcott and Sweeney 2017). Therefore, the observed effect here is not negligible, especially since it comes as a byproduct of a recycling program. From an operational perspective, our findings extend previous research that ties recycling operations to green marketing and contradict the conventional wisdom that favors free/promotional recycling as the beneficial bridge to coordinate the two (see, e.g., Westley and Vredenburg 1991, Shrum et al. 1994). Precisely, our study findings show that charging for recycling not only directly assists retailers with the operational costs of recycling, but can also help enhance the recyclability of products entering the economy in the first place, a factor that has been emphasized as vital to creating an efficient circular economy (Geyer et al. 2015, Zink and Geyer 2017).

While one cannot be definite about the outcomes on an environmental policy, the scope of our experimental studies is to sharpen the best guesses on the potential of a new environmental policy (Shogren 2002, Chetty 2015). A limitation on charging for recycling as an environmental policy can be pointed out as potentially having a deterring effect on consumers' product returns, compared with free recycling policy, and thus potentially reducing the volume of products returned for recycling. However, the current state of TV and monitor recycling can result in retailers closing their recycling programs for these products and thus setting the volume of recycling at zero. For example, Best Buy has the largest e-waste recycling program in the U.S., but in the states that do not allow charging for recycling, it still had to shut down its recycling channel for TVs and monitors due to their high recycling costs. Meanwhile, previous evidence showing strong positive correlations between green purchasing behavior and recycling behavior (Biswas et al. 2000) keeps us hopeful that a policy of charging for recycling may not significantly deter consumers' recycling as it enhances their green purchasing behavior. Hence, by changing consumers' purchase behavior, such a policy may result in greater environmental benefits, with no or little drawbacks on recycling volume, over merely focusing on increasing the recycling volume alone.

Our findings also suggest potential benefits in implementing these sorts of new programs in industries other than consumer electronics to increase consumer adoption of socially/environmentally friendly products. For example, in the apparel industry, utilizing a similar charging-for-return policy by Goodwill to accept used clothes can potentially increase consumers' adoption of organic clothing, which will, in turn, motivate manufacturers to increase using organic materials in their productions.

The remainder of the paper is organized as follows: Section 4.2 reviews the related literature and builds our hypotheses; Section 4.3 describes the experimental design; Section 4.4 presents results and analyses; and finally, Section 4.5 discusses the implications of our findings for green marketing and environmental policy.

4.2 Hypotheses

4.2.1 Charging for Recycling and Adoption of Green Versions

Our main discussion here is that the presence of a recycling charge at the purchase point would influence customer purchasing behavior and lead to a higher adoption of green versions of products. Based on Best Buy's practice of charging for recycling, i.e., announcing its recycling charge and its sale of green electronics at the same time, we find two behavioral explanations most supportive of our postulation: nudging and the theory of planned behavior. Nudging, i.e., piquing attention through indirect interventions (Thaler and Sunstein 2008, Leonard 2008, Sunstein 2009), has been used in many different contexts to lead people's actions in desired directions. The sort of nudge that aims to promote environmentally friendly actions, for example, decreasing the household energy consumption (see, e.g., Allcott and Rogers 2014), is known as the green nudge (Sunstein 2009). The theory of planned behavior, first proposed by Ajzan (1991), has been widely used in studying people's green purchasing behavior (see, e.g., Ellen et al. 1991, Berger and Corbin 1992, Rice et al. 1996, Kalafatis et al. 1999, Chan and Lau 2002).

To discuss nudging in this context, we draw a parallel with previous work on the provision of direct information to increase consumer adoption of green products. This stream of work uses energy efficiency in green products to motivate consumers via savings in their electricity bills as a win-win benefit in purchasing such products (see, e.g., Allcott and Greenstone 2012). However, the work has found direct provision of information to be an unsuccessful marketing strategy for increasing the sale of green products. Allcott and Taubinsky (2015) found that a significant portion of the market still preferred incandescent lightbulbs even after being fully informed about benefits of energy-saving lightbulbs. Kallbekken et al. (2013) found that a salesperson's effort had no significant effect on the demand for energy-efficient products. Similarly, Allcott and Sweeney (2017) found that incentivizing salespeople to provide consumers with information on the cost savings of energy-efficient water heaters had no effect on consumer demand.

In contrast to the evidence on the ineffectiveness of direct marketing strategies, there are reports of successful uses of nudges to reduce energy consumption among consumers. Allcott and Rogers (2014), in their studies of comparison-based households' energy bills, found the comparison-based bill to be a successful nudge that reduced households' energy consumption. Similarly, Dolan and Metcalfe (2013) found a significant reduction in energy consumption by nudging households to reduce their energy consumption to comply with social norms. In reducing households' water consumption, Ferraro and Price (2013) found messages based on social norms and social comparisons to be more effective than instructive descriptions.

In parallel with the above evidences, we postulate that indirectly nudging the value of recyclability by charging for recycling will work effectively in increasing the adoption of recyclable versions of products. How charging for recycling serves as a nudge is to indirectly draw customers' attention to the fact that "recycling is costly" and thus presents "increased recyclability" as a more valuable attribute. Consumers, if asked, might have some knowledge of the expenses/harms that recycling imposes on society and the environment. What a nudge does is to draw their attention to this fact at the purchase point, without which this knowledge might never come to the surface or effectively play its role at the decision point. To the best of our knowledge, our study is the first that explores a green nudge in the domain of consumer adoption of green products. It is worth noting that the lack of full awareness that sometimes appears in nudges and raises criticism of restricting people's autonomy to act upon their own preferences (see, e.g., Hausman and Welch 2010) does not apply to our study since we do not architect a nudge and rather investigate a byproduct of a recycling program as a nudge.

In the Theory of Planned Behavior (TPB), the main argument is that while one factor of the planned behavior, attitude, is predominantly in favor of environmentally friendly purchases, the other factor, perceived behavioral control, is not always realized and thus hinders the final behavior. According to Ajzen (1991), the nature of the perceived behavioral control is uncertainty about how impactful people think their action would be on the environment. This concern has been found to be a significant determinant of eco-friendly behavior (see, e.g., Ellen et al. 1991, Berger and Corbin 1992, Rice et al. 1996). We postulate that a recycling charge will induce a belief on consumers that not buying a recyclable product imposes a similar monetary cost to the environment. This monetary, relatable cost will make it clearer to customers what impact they are making on the environment by adopting recyclable versions and will thus enhance the perceived behavioral control. This improved self-efficacy in making an eco-friendly action will, in turn, close the gap in TPB and result in adoption of green (recyclable) products. Combining the two discussions above, we summarize our expectation in the following hypothesis:

HYPOTHESIS 1. *Compared with free recycling, charging for recycling will increase adoption of green versions of products.*

Although, compared with free recycling, both charging for recycling and a recycling tax collect recycling fees from consumers, there are fundamental differences between them. More explicitly, in contrast to charging for recycling, a recycling tax is reflected as a mandatory price inflation at the purchase point and at a relatively low magnitude. On one hand, being mandatory and tying the nudge to purchase costs rather than recycling costs disturb nudging dynamics with the recycling tax. Successful green nudges motivate people towards voluntary, rather than forcible, contributions to environmental protection actions (Schubert 2017), and the recycling fee under the recycling tax policy lacks this. Furthermore, one of the main reasons behind the success of nudging households with their neighbors' energy consumption is that the information is delivered to the households together with their energy bills, which makes the information salient and thus work as a nudge (Allcott and Rogers 2014, Schubert 2017). In contrast to a charging-for-recycling policy that charges a recycling fee because of accepting a product for recycling, a recycling tax policy charges a fee because of selling the product, which does not fit well in tying the information to the right activity and thus creating an informational nudge.

On the other hand, even if a recycling tax can, to some level, fit as green nudge, its relatively low magnitude may not allow it to make a strong impact. Due to being collected upfront at the purchase point and for all products sold, a recycling tax is sensibly lower than the recycling charge that a charging-for-recycling policy will collect at the recycling point and only for recycled products (for example, for TVs of the same size, California's Electronic Waste Recycling Act charges $7 recycling tax, which is significantly lower than the $25 recycling charge in Best Buy's program). Hence, any possible nudging effect with a recycling tax policy will be significantly lower than that with a charging-for-recycling policy.

From the TPB point of view, we can postulate that, by collecting a fee at the purchase point, the recycling tax will suggest to consumers that they are paying their contributions to the environment and will thus leave no room for further pro-environmental actions at this point. In this vein, Merritt et al. (2010) discuss how one act of pro-social behavior can suppress further actions by making people feel as if they have done "enough." With the charging-for-recycling program, on the other hand, the similar pro-environmental action, i.e., paying for recycling, while being present, is tied with recycling action and thus does not interfere with or hinder the other pro-environmental action at the purchase point, i.e., paying more for the recyclable version. This postulation, together with the above discussion on the dynamics of recycling tax, leads us to the following hypothesis:

HYPOTHESIS 2. *Compared with recycling tax, charging for recycling will increase adoption of green versions of products.*

4.2.2 Consumer Heterogeneity

Bao and Ho (2015) discuss that a nudge can have heterogeneous effects on people's decisions depending on the characteristics of the situation they are in and that making the most of a nudge requires understanding possible heterogeneities. Allcott and Rogers (2014), in their study of nudging households with comparison-based energy bills, found that households with consumption above average started to decrease their energy usage, while those with consumption below average tended to increase their usage, which limited the overall effect to 1-2% savings in electricity. Similarly, Beshears et al. (2015), in their study of retirement savings, found that peer information increased savings among non-unionized recipients but decreased it among unionized recipients, due to differences in norms between these two groups of workers. In this vein, Bronchetti et al. (2011) segmented people based on their financial limitations and found that opt-in/out defaults had no impact on the saving behavior of low-income tax filers as they had strong intentions to spend their refunds. More related to our study, Samson and Voyer (2014) discuss that emergency purchases, compared with regular purchases, are more prevention-focused, and thus emergency buyers may only be willing to spend on necessary attributes in products and not much on environmental attributes.

Emergency purchases in consumer electronics, such as TVs and monitors, in a broader picture can be considered as replacement purchases, for which previous research has found more nuances in behavioral influences compared with new purchase situations (see Guiltinan 2010, for a review). For example, Bayus (1991) and Cripps and Meyer (1994) found that purchases due to product failure are based on different decision-making processes than regular purchases. In line with most of the work on replacement purchases, our study concerns a durable good (TV/monitor). Like most durable electronic goods, purchase cycles for TVs/monitors have been significantly shortened,[7] and in today's consumer electronics markets, the main portion of the market purchases new versions of products while having a fully functioning version in use (Gordon 2009). Therefore, we define emergency/replacement buyers as the ones who are buying a new TV to replace their broken TV, and regular/new buyers as the ones who are buying a new TV while still having an-

[7] On average, around 5-7 (3-4) -year cycle is assumed for TVs (monitors). http://www.nytimes.com/2011/01/06/technology/06sets.html

other fully functioning one at home. This segmentation is in line with previous research and captures both the dynamics of regular vs. emergency purchases as well as those of new vs. replacement purchases. Based on previously established evidence (e.g., Bronchetti et al. 2011, Samson and Voyer 2014), we postulate that the cost-oriented nature of emergency purchases will overshadow the behavioral effects of the charging-for-recycling program for emergency buyers.

HYPOTHESIS 3. *The effect of charging for recycling in increasing the adoption of green versions will be stronger for regular buyers compared with emergency buyers.*

The main point of the discussion above is that a possible heterogeneous nudging effect from charging for recycling is driven by a strong reason for a change in behavior. There can be, of course, other consumer heterogeneity in the market that can segment consumers into several sub-groups. Age, gender, annual income, etc., are among common sources of consumer heterogeneity that direct green marketing usually faces. For instance, research on direct green marketing has found that supporters of environmentalism tend to be younger in age, as young people are more open to new ideas (see, e.g., Lee 2008 and references therein). Houde (2014) discusses such heterogeneity in consumer adoption of energy-efficient products via direct green marketing in more detail by relating that to consumer sophistication with respect to collecting and processing direct information.

One of the main issues in the direct green marketing of energy-efficient products has been discussed as cognitive limitations such as hyperbolic discounting. Direct green marketing strategies usually push customers to evaluate the cost-benefit of paying a higher purchase price for energy-efficient products in exchange for obtaining energy bill savings in the future (Anderson and Claxton 1982, Newell and Siikamäki 2014). Such evaluations also occur when households evaluate the benefits of taking steps to reduce their energy consumption to achieve future energy bill savings. However, green nudges in this area, e.g., nudging households with neighbors' energy consumptions, has been found to be successful by going around such cognitive limitations (Allcott and Rogers 2014, Schubert 2017). With green nudges avoiding this issue in direct green marketing, we can expect that they would also go around other issues that direct green marketing faces—in particular, consumer heterogeneity as discussed above.

Therefore, upon Hypothesis 3 characterizing the main source of difference in behavior, we expect that charging for recycling, as a green nudge, will have a homogenous effect across all consumers, irrespective of their demographic heterogeneity. This expectation is in line with previous findings in nudge in other contexts, as well. For example, Azmat and Iriberri (2010), in studying how students' performance was influenced by learning their relative ranks, found a homogeneous nudging effect for both genders after differentiating the effect for the tails of the distribution from middle-rank students. Referring to common sources of heterogeneity, such as age and gender, by "consumer heterogeneity," we present this expectation in the following hypothesis:

HYPOTHESIS 4. *The effect of charging for recycling in increasing the adoption of green versions will be robust to consumer heterogeneity.*

4.3 Experimental Study

4.3.1 Design

The objective of our experiment is to compare charging for recycling with two current recycling policies (i.e., free recycling and recycling tax policies) as the baseline to extract its relative different effect on the adoption of green versions of products. To keep our experimental design in line with reality, we choose $25 as the dollar amount charged for recycling. In addition, since Best Buy's program charges only for TVs/monitors, we use a TV as the electronic product in our experiment. With California's Electronic Waste Recycling Act, the fee charged at the purchase point is based on the size of the purchased electronic product, which is $7 for products larger than 35 inches. Hence, we choose $7 as the recycling tax on the purchased TV in our experiment. We use technical specifications of a real market TV (product specifications are available in Appendix C) and create two identical versions of the TV, Standard and Green, where "designed for increased recyclability"[8] is the only extra attribute in the Green version.

Similar to real market situations, where customers discover one version of a product first and then find about the other version, participants in our experiment are first offered one version of the TV (Standard or Green),[9] and after indicating their WTPs for that version, are offered both versions and asked to indicate their WTPs for both. Participants are allowed to change their starting WTPs for the Standard (or Green) version after being offered the other. We use participants' final WTPs for the Green and Standard versions in the analysis to extract their adoption of the Green version. Realistically, the starting WTP might influence the following WTPs for the Green and Standard versions and hence the gap between them by taking effect as an arbitrary anchor (Ariely et al. 2003, Levav et al. 2012). Therefore, by directly asking participants about their WTPs, our design leaves it to the participants how much they are willing to pay for the TV to start with. This further takes into account subjective factors, such as the income level, on participants' valuation of a given TV and consequently their extra willing to pay for the Green version. These are essential for real market simulation and thus could not be ignored in the experimental design.

The measure of adoption we use in our analyses, nonetheless, nullifies both a possible anchoring effect from the starting WTP and the inherent subjectivity issues, and extracts the adoption in a way that allows us to have between-group comparisons. That is, any difference unaccounted for in the subject-level extra WTP for the Green version in different groups comes purely from the manipulations of our interests, namely, recycling policy and purchase situation. To do so, we use the subject-level percentage of extra WTP for the Green version relative to the participant's own WTP for the Standard version as the measure of adoption, and compare its average between different experimental groups.

[8]This is the phrase that is being used in the marketing of Green TVs/monitors by Best Buy and is a description of EPEAT-registered electronics as well.

[9]This ordering is counterbalanced in the experiment, and we perform further analyses to ensure that the results are independent of the ordering.

We refer to this variable as (extra WTP_G)% and build any between-group comparisons around that rather than simply comparing the average of extra WTP for the Green version between different groups. This variable further allows us to investigate both the mere adoption—through (extra WTP_G)%$>$0—and the intensified adoption—through the value of (extra WTP_G)%—in our between-group compositions. It is noted that an alternative to asking participants about their WTPs could be using a choice-based design, providing participants with multiple ranges of prices for the starting TV, and then multiple ranges of price gaps between the Standard and Green versions to capture all the same information our design provides. This would, however, result in choice sets with too many options, yet not covering every possible WTP and thus not being able to go around the possible anchoring and subjectivity issues completely.

To create a clear separation of the purchase situations in the experiment, for emergency/replacement buyers, the situation involves customers replacing a broken TV with a new one, and for regular/new buyers, the situation involves purchasing a new TV for another room of the house. With the two baseline recycling policies and the charging-for-recycling program, and separating the purchase situations for regular/new buyers and emergency/replacement buyers, we thus have a 3×2 design as summarized in Table 4.1. Experimental descriptions will be discussed in detail in Section 4.3.3.

4.3.2 Incentive Compatibility, Social Desirability, and Attention Check

To capture the exact setting and time point of decision making, e.g., deciding to purchase a new TV while already owning a broken (or fully functioning) one, we use an imaginary situation in the experiment and ask participants to indicate their WTPs assuming the described purchase situation (see the next section for the experimental descriptions). With imaginary situations, although the experiment cannot be an incentive-compatible design, there is ample evidence that, if designed properly, respondents' decisions do not differ significantly from real situations (see Carmon and Ariely 2000 and references therein) and

Table 4.1: Experimental framework.

	Regular/new purchase	Emergency/replacement purchase
Free recycling	• Recycling is free of charge • Already have a TV (in good condition) at home, and wanted to buy another one for another room	• Recycling is free of charge • Current TV not working anymore, and need to buy a new one
Recycling tax	• $7 recycling fee on the purchase price • Already have a TV (in good condition) at home, and wanted to buy another one for another room	• $7 recycling fee on the purchase price • Current TV not working anymore, and need to buy a new one
Charging for recycling	• $25 charge for recycling • Already have a TV (in good condition) at home, and wanted to buy another one for another room	• $25 charge for recycling • Current TV not working anymore, and need to buy a new one

the experiment could yet yield similar insights. More similar to our context, Ungemach et al. (2018) used imaginary situation experiments with flat participation payments to study consumer preference between environmentally friendly and regular cars. As Meloy et al. (2006) discuss, these sorts of experimental designs are of particular importance in environmental studies (e.g., green purchase behavior, environmentally friendly actions, etc.) where incentive-aligned payment often hinders understanding people's intrinsically motivated behaviors.

In addition, the comparison-based approach of our study, together with the nature of the adoption measure we use, enables us to go around the need for incentive compatibility in obtaining reliable results. Precisely, we do not use the raw gap in participants' WTPs for the Green and Standard versions in our analyses, and rather utilize participants' extra WTP for the Green version relative to their own WTP for the Standard version. More importantly, we compare participants' adoption of the Green version under the charging-for-recycling scenario with that under the baseline scenarios and are only interested in a net difference that the charging-for-recycling scenario can make. Hence, any possible price inflation for the Green version, due to lack of incentive compatibility, will be cancelled out in extracting the net difference, and the results will be driven purely by the effect of charging for recycling. We keep the same comparison-based approach in analyzing both the regular/new purchase situation and the emergency/replacement situation.

Another important point to consider in the design and analyses is the possible social desirability effect in the results. The social desirability effect, or Experimenter Demand Effect (EDE), is about change in experimental subjects' decisions due to what might constitute appropriate behavior (Zizzo 2009). In other words, subjects may give priority to being "good subjects" and feel committed to making a decision that assists the experimenter. As Zizzo (2009) discusses, in most environmental/social studies, this can be an issue from two points: first, using an incentive-compatible design and inducing to participants to maximize their monetary benefits may deter them from their environmental/social utilities in honoring the induced, desired performance; second, if the environmentally/socially conscious decision is positively correlated with experimenter demand (i.e., if participants show environmental/social consciousness in their decision because of assuming that it is what the experimenter is looking for), that would lead the results in a biased direction. Our experimental study takes care of the latter by using the comparison-based design. In these comparisons, there are comparable sources of environmental/social desirability in both the treatment and the control (baseline) groups. In both the free recycling and charging-for-recycling scenarios, participants have comparable opportunities to show a socially desirable decision, i.e., paying more for the Green version, if they are influenced by the EDE. This neutralizes the effect of the EDE, further ensuring that the net outcome is driven by our hypothesized effects. When comparing the recycling tax and charging-for-recycling scenarios, the situation is the same. Both scenarios may partly satisfy participants' environmental/social consciousness (with the recycling tax scenario, participants are default with $7 recycling fee; and with the charging-for-recycling scenario, they are default with a $25 recycling charge) while leaving comparable opportunities for showing social desirability (i.e., by paying more for the Green version). Hence, there are comparable levels of possible EDE in both. Therefore, a composition-based analysis will cancel out possible EDE in the treatment scenario, and the net difference with the charging-for-recycling scenario will be due to the hypothesized effects.

To ensure the reliability of the responses, we use a strict attention check question at the end of the experiment to filter inattentive respondents. Abbey and Meloy (2017) discuss a wide range of attention check questions commonly implemented in experimental studies. Our attention check question tests whether participants understand both of the treatments in the descriptions: whether the situation was a regular/new purchase or an emergency/replacement purchase; and whether the recycling was free or whether there was a fee associated with it. With such a test that covers the focal points of the task, we do not further filter participants based on their response time; neither do we use out-of-context questions to test participants' precision.

4.3.3 Descriptions

The experimental description includes three short parts: first, a starting description with manipulation on the purchase situation (regular/new vs. emergency/replacement); second, a description for manipulation on the recycling policy; and third, a description stating the difference between the Standard and Green versions of the TV. Full experimental descriptions are available in Appendix C.

The experimental description starts with telling participants whether the situation is a regular/new or an emergency/replacement purchase. For the former, we ask participants to imagine they already have an almost-new TV at home and are thinking about buying an extra one for another room. For the latter, we ask participants to imagine their TV is not working anymore and that they need to buy a new one.

The second part of the description creates the treatment effect regarding the recycling policy. In doing so, we try to match the description with Best Buy's announcement on its new charging-for-recycling program as much as possible. With the free recycling scenario, the description mentions that the retailer also has a recycling program and accepts broken TVs for recycling free of charge, in line with Best Buy's previous free recycling program. With the charging-for-recycling scenario, on the other hand, the description mentions that the retailer also has a recycling program and, due to the cost of managing TV recycling, charges $25 to accept broken TVs for recycling. This short description captures Best Buy's announcement on its new recycling program together with the reasoning behind it as mentioned in the announcement. We avoid adding any further detailed information in the interest of having the best reflection of reality. For example, the description leaves it to the participants what to do with their broken TVs in the emergency/replacement purchase groups, either with the free recycling or the charging-for-recycling scenario.

To have as much consistency between different groups as possible and to leave the sources of difference only to the intended manipulation in this part, descriptions use the same wording when describing the recycling policy (accepts broken TVs for recycling free of charge, or charges $25 to accept broken TVs for recycling) in both regular/new and emergency/replacement purchase situations. This leaves the difference between the two purchase situations only to the first part of the experimental description and keeps the focus of this part on the recycling policy manipulation. The word "broken" is to ensure aligning the descriptions between the two parts in the emergency/replacement purchase situation in which participants face broken TVs. At the same time, it makes sense for the regular/new purchase situation as participants are meant to assume they are not concerned with returning broken TVs for recycling at the purchase point.

Similar to the charging-for-recycling scenario, the recycling policy manipulation in the recycling tax scenario is about a recycling fee, and in line with reality, the description mentions that the retailer also has a recycling program and, due to the cost of managing TV recycling, charges a $7 recycling fee in addition to the sale price for all TVs, regardless of the brand or technical specifications. We do not use the word "tax" in the description because the $7 fee associated with California's Electronic Waste Recycling Act of 2003 is not being stated as a "tax" in practice.

The experimental descriptions, and in particular the first two parts conveying the two manipulations, were refined through several pre-runs, where participants were asked to provide explanations behind their thinking. This open-ended question was to ensure the clarity of the descriptions to participants without having to list further detailed information in the descriptions. From what we obtained from this open-ended question with the final version of the descriptions, participants' thinking was around environmental concerns with recycling (e.g., not being green, etc.) rather than rational monetary calculations.

The last part of the experimental description tells participants about the technical specifications of the TV (Green or Standard version), presented alongside the picture of the TV, and asks them about their WTPs for the offered TV. For the Green version, the description clearly states that it also is a Green TV, designed for increased recyclability at its end of life. Upon indicating their WTP for the offered version (Green or Standard), participants go to the next stage, where they are first reminded about the WTP they previously indicated for the offered version and then are told to assume that they find there is also a Green [or Standard] version of that TV, wherein both versions are shown together. The description points out clearly that both versions have the exact same features except the increased recyclability in the Green version. Participants are asked, in light of this new information, about their WTPs for each of the TVs. It is noted that in the recycling tax groups, participants are asked to indicate their WTP inclusive of the $7 fee, in order to capture the final price paid for the TV as in reality.

4.3.4 Procedure

We recruited a sample of 838 participants (46.5% female and 53.1% male;[10] M_{age}=39.8, SD=12.1) through Amazon Mechanical Turk (AMT)[11] and paid a flat fee for their participation (the same as in similar studies in AMT, e.g., Ungemach et al. 2018). The payment proportionally was over the minimum wage per hour at the time the experiment was run. Participants were eligible to participate if they were at least 18 years old and were U.S. residents. Upon agreeing to participate in the study, participants were randomly assigned to one of the six experimental groups (as shown in Table 4.1) and were asked about their WTPs for the Standard and Green TVs in the order explained in the previous section. In all groups, following the main experimental task, participants also provided an (optional) explanation of their thinking behind their decision. We also asked participants about the clarity of the descriptions and experimental tasks. The average clarity score on a 1-5 scale (5 being the highest) was 4.5. The experiment ended with the attention check question and demographic questions. A majority of the participants (73.6%) indicated having earned a bachelor's degree or having a college-level credit. The average income level was between $26,000-$75,000, and most of the participants had an income level between $26,000-$50,000.

4.4 Results and Analyses

Among 838 participants, around 5-7 in each group (36 in total; 4.3%) indicated WTPs for the Green TV lower than their WTPs for the Standard TV. Based on their explanations, we call them anti-Green participants—consumers with strong doubts about green and environmentally friendly products (Ginsberg and Bloom 2004). Exclusion or inclusion of these participants in the sample does not make a difference in the main results and findings, and we keep them in the sample as we do not have a predetermined reason to exclude them.

[10] 0.4% of the participants preferred not to indicate their gender.

[11] Amazon Mechanical Turk (AMT) has become a popular platform for experimental studies, and evidence has endorsed AMT as a reliable source of experimental data (see, e.g., Lee et al. 2018 and references therein).

We filter responses based on the attention check question at the end of the experimental task, which yielded a rejection rate under 22.1%.[12] Finally, we exclude six participants who were not able to enter a valid number as to their WTPs in either/both of the questions.

The focal point of our analyses is isolating the effect of the $25 charge for recycling on the subject-level percentage of extra WTP for the Green version, which we have defined as the measure of adoption. We refer to this variable as (extra WTP_G)%. In analyzing this variable, our experiment uses a between- and within-subject mixed model design: having three recycling policies as a between-subject manipulation while measuring within-subject WTPs for the Standard and Green versions. In addition, the two purchase situations (regular vs. emergency) and the two orders (Standard first vs. Green first) add to the between-subject manipulation. Hence, in total, our experimental design is a $3\times2\times2\times2$ mixed model based on recycling policy $\times$ purchase situation $\times$ order $\times$ product version. Before going into the details of analyzing our main variable of interest, i.e., (extra WTP_G)%, through regressions, it is worth analyzing this mixed model first in order to uncover sources of variation in (extra WTP_G)%. We use mixed-model ANOVA, which is the perfect methodology for a mixed between- and within-subject experimental design (see Abbey et al. 2015 that used this method to study customer preference for new and remanufactured products).

As Table 4.2 shows, the mixed-model analysis suggests a statistically significant effect from recycling policy under some conditions. The first row of Table 4.2 shows that there is a statistically significant difference between WTPs for Green and Standard versions. The rest of the table investigates any effect from between-subject manipulations on this difference. The second row shows that without differentiating the purchase situation, recycling policy manipulation has no statistically significant effect on within-subject difference between WTPs for Green and Standard versions. However, the third row uncovers that depending on the purchase situation, recycling policy can, in fact, increase within-subject extra WTP

[12]Given its relatively small size (see Abbey and Meloy 2017 that discuss data exclusion rates in attention check failures), inclusion of this data set in the analysis did not change the final results, nor it influenced the significance levels, and only slightly changed the p-values.

for the Green version. Finally, the last row shows that the order in which participants discover the first version has no statistically significant effect on the effect from recycling policy manipulation.

Table 4.2: Mixed-model analysis.

Term	Full model		
	F	Power	Effect size
version	201.74**	0.241	1.000
version*policy	0.32	0.001	0.100
version*policy*purchase	4.15*	0.013	0.733
version*policy*purchase*order	0.92	0.003	0.210

Note: Effect size is based on partial-eta-squared. Power reflects observed power.
*Significant levels: *=0.05 and **=0.01.*

To further explore whether the observed effect from recycling policy manipulation is driven by the charging-for-recycling scenario, the recycling tax scenario, or both, Tables 4.3a-4.3c repeat the same analysis of Table 4.2 by contrasting only two scenarios of recycling policy (charging for recycling vs. other policies, recycling tax vs. other policies, and free recycling vs. other policies). In this case, the mixed-model design reduces to $2 \times 2 \times 2 \times 2$, which allows us to isolate the observed effect for different recycling policies. As Tables 4.3a-4.3c demonstrate, it is the charging-for-recycling scenario that, compared with free recycling and recycling tax scenarios, makes a statistically significant difference on participants' extra WTP for the Green version.

The mixed-model ANOVA uncovered that charging for recycling can have a statistically significant effect on consumers' extra WTP for the Green version in some purchase situations. This suggests that it is worth conducting subsequent detailed analyses to further investigate and understand the observed effect. These detailed investigations would also determine whether our hypotheses in Section 4.2 are supported. We structure our subse-

Table 4.3a: Mixed-model analysis with charging for recycling scenario vs. other recycling scenarios.

Term	Full model		
	F	Power	Effect size
version	190.10**	0.229	1.000
version*policy	0.45	0.001	0.102
version*policy*purchase	5.54*	0.009	0.652
version*policy*purchase*order	0.36	0.001	0.092

Note: Effect size is based on partial-eta-squared. Power reflects observed power.
*Significant levels: *=0.05 and **=0.01.*

Table 4.3b: Mixed-model analysis with recycling tax scenario vs. other recycling scenarios.

Term	Full model		
	F	Power	Effect size
version	158.74**	0.198	1.000
version*policy	0.42	0.001	0.099
version*policy*purchase	0.18	0.000	0.071
version*policy*purchase*order	0.78	0.001	0.142

Note: Effect size is based on partial-eta-squared. Power reflects observed power.
*Significant levels: *=0.05 and **=0.01.*

Table 4.3c: Mixed-model analysis with free recycling scenario vs. other recycling scenarios.

Term	Full model		
	F	Power	Effect size
version	191.88**	0.230	1.000
version*policy	0.00	0.000	0.050
version*policy*purchase	7.25**	0.011	0.767
version*policy*purchase*order	1.84	0.003	0.272

Note: Effect size is based on partial-eta-squared. Power reflects observed power.
*Significant levels: *=0.05 and **=0.01.*

quent analyses based on analyzing the effect of charging for recycling on the adoption of the Green version for the two customer segments, i.e., customers in regular/new and emergency/replacement purchase situations (Analysis 1), further analyzing the robustness of the effect to common consumer heterogeneities in direct green marketing (Analysis 2), and characterizing the effect in terms of green market expansion and green market intensification (Analysis 3).

4.4.1 Analysis 1: Charging for Recycling and Adoption of the Green Version in Regular/New and Emergency/Replacement Purchase Situations

The focus of our analysis here is to statistically compare the average of adoption measure, i.e., (extra WTP$_G$)%, under the charging-for-recycling scenario with that under free recycling and recycling tax scenarios as the baseline. Given our hypotheses structure and what uncovered in the mixed-model analysis, we perform this comparison separately for the regular/new purchase situation and the emergency/replacement purchase situation. With the mixed-model ANOVA, we analyzed what influenced any change in participants' WTPs for the Green and Standard versions. We now aim to compare participants' adoption of the

Green version in different groups. Hence, for reasons discussed earlier, we look at the adoption measure (extra WTP_G)% rather than raw WTPs. Figures 1 and 2 show the average of this variable for each experimental group. In the regular/new purchase situation, the recycling charge increases the average of (extra WTP_G)% by 4.9 compared with free recycling ($t = 2.687$, $p < 0.004$) and by 3.4 compared with the recycling tax ($t = 1.704$, $p < 0.046$). It makes no difference, however, for the emergency/replacement purchase situation relative to the baselines (-0.25 compared with free recycling, $t = 0.152$, $p > 0.879$; and -0.66 compared with the recycling tax, $t = 0.416$, $p > 0.678$).

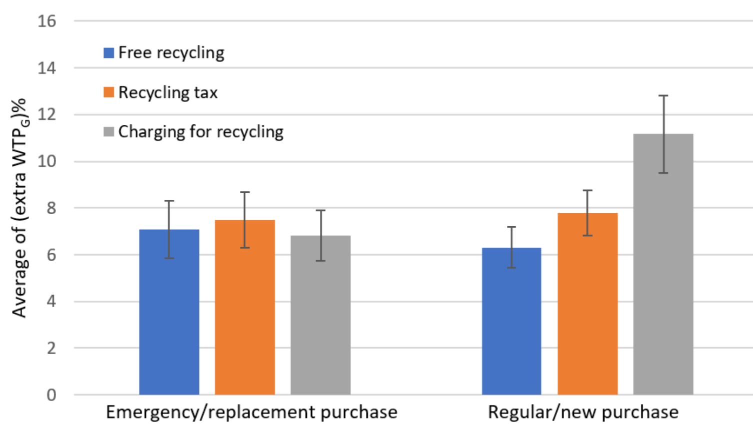

Figure 4.1: Average of (extra WTP_G)% for the two purchase situations with the three recycling policies. *Note: Error bars are standard errors of means (SEM).*

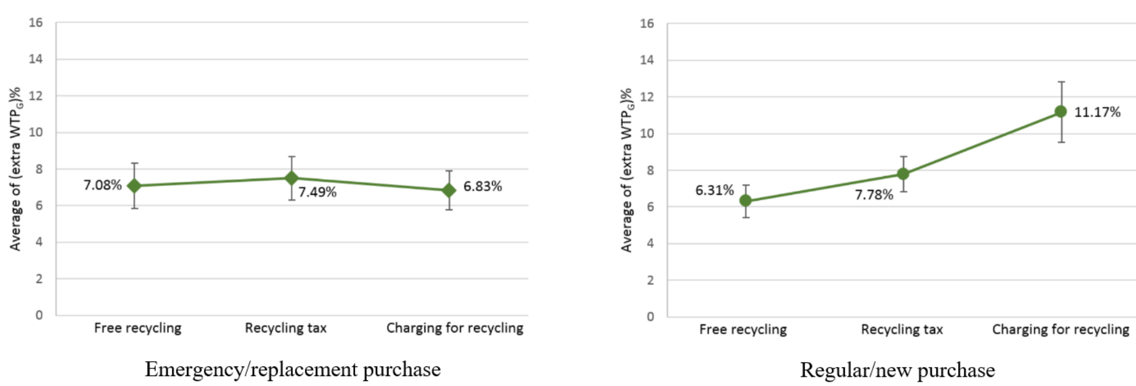

Figure 4.2: The effect of charging for recycling on the average of (extra WTP_G)% for the two purchase situations. *Note: Error bars are standard errors of means (SEM).*

Discussion: Our analysis reveals that, in line with our general expectations in Hypothesis 1 and 2, the presence of a recycling charge increases participants' adoption of the Green version through increasing their (extra WTP$_G$)%, compared with both free recycling and recycling tax policies. The results show that although the same information (i.e., "due to the cost of managing TV recycling") is being communicated to participants in both the charging-for-recycling and recycling tax groups, as predicted, the dynamics of a recycling tax are more similar to those of free recycling and thus result in an adoption level more comparable to a free recycling than a charging-for-recycling policy. Moreover, comparing the recycling tax group with the free recycling group, to which this information is not provided, our observation provides further evidence that mere information provision is not an effective policy in increasing adoption of green versions of products. Finally, the results show that as predicted in Hypothesis 3, the cost-oriented nature of the emergency/replacement purchase situation does not let the effect realize under this purchase situation; hence, the adoption remains at the same level as in the baseline. This observation also provides supporting evidence for Samson and Voyer's (2014) observation that, compared with regular purchase situations, decision making in emergency purchase situations is more prevention-focused and hinders paying for extra attributes such as environmental attributes. It is also worth noting that with the adoption level remaining unchanged in all scenarios under the emergency/replacement purchase situation, the results endorse that, as reasoned in Section 4.3.2, our experimental design does not fall under issues pertinent to the social desirability effect. This ensures the reliability of the extracted difference in the adoption level between the charging-for-recycling scenario and the baseline scenarios under the regular/new purchase situation.

4.4.2 Analysis 2: Robustness of the Effect to Common Consumer Heterogeneity beyond the Purchase Situation

Our analysis here aims to explore whether the observed effect is robust to consumer heterogeneity besides the main segmentation that Analysis 1 considered. We use participants' demographics, i.e., gender and age, to define consumer heterogeneity, which is a common approach in marketing studies. We separate the participants into young (*ageL*) and senior (*ageH*) based on the cut-off age of 40 reflecting the average age of participants (i.e., 39.8; see Section 4.3.5). We also exclude the participants (0.4% of the experimental pool) who did not indicate their gender in separating the participants into male and female groups. We start by repeating the mixed-model analysis of Table 4.2 considering gender and age extra sources of between-group manipulation. This results in a $3 \times 2 \times 2 \times 2 \times 2 \times 2$ mixed-model. Table 4.4 shows the results of this analysis. As seen, neither age nor gender has a statistically significant influence on the difference between participants' WTPs for the Green and Standard versions under any of the recycling policies or purchase situations.

Table 4.5: Mixed-model analysis including participants' demographics.

Term	Full model		
	F	Power	Effect size
version	203.58**	0.254	1.000
version*policy	0.02	0.000	0.053
*ageH	0.46	0.002	0.124
*male	0.07	0.000	0.060
*ageH*male	0.74	0.002	0.176
version*policy*purchase	4.24*	0.014	0.742
*ageH	0.31	0.001	0.099
*male	1.03	0.003	0.231
*ageH*male	1.00	0.003	0.225
version*policy*purchase*order	1.22	0.004	0.267
*ageH	0.18	0.001	0.079
*male	1.12	0.004	0.248
*ageH*male	0.92	0.003	0.210

Note: Effect size is based on partial-eta-squared. Power reflects observed power.
*Significant levels: *=0.05 and **=0.01.*

We further use regression analyses to complement the above analysis on participants' demographics. Following up on Analysis 1, we run the following regression that considers dummy variables *ageH* and male together with the three recycling policies for both the regular/new purchase situation and the emergency/replacement purchase situation:

$(extra\,WTP_G)\%_i = \alpha_0 + \alpha_1(ageH)_i + \alpha_2(male)_i + \alpha_3(ageH \times male)_i + \alpha_4(recycling\,tax)_i + \alpha_5(recycling\,charge)_i + \alpha_6(recycling\,tax \times ageH)_i + \alpha_7(recycling\,charge \times ageH)_i + \alpha_8(recycling\,tax \times male)_i + \alpha_9(recycling\,charge \times male)_i + \alpha_6(recycling\,tax \times ageH \times male)_i + \alpha_7(recycling\,charge \times ageH \times male)_i + \epsilon_i$

Table 4.5 shows the result of this regression for both purchase situations. As seen, the observed effects remains statistically independent of participants' demographics.[13]

It is worth noting that, as discussed earlier, the measure we use for the adoption of the Green version, i.e., (extra WTP_G)%, isolates the extracted effect from participants' income levels altogether. This was an important consideration because it is plausible that, for the same TV, a participant with a high income level might indicate a higher WTP than a participant with a low income level. Similarly, it might also be the case that the participants' starting WTPs, taking the role of an arbitrary anchor (Ariely et al. 2003, Levav et al. 2012), may further influence their extra WTPs for the Green version. This, however, is most likely not an issue with the measure (extra WTP_G)% as it rules out the effect of the income level. To quantitatively confirm this, we run a linear regression that considers the participant's starting WTP as a potential anchor together with the effect of different recycling policies under the two purchase situations. Since starting WTPs can potentially be higher in the Green-first order, compared to the Standard-first order, we further consider an extra interaction term for the Green-first order to capture this possibility:

$(extra\,WTP_G)\%_i = \alpha_0 + \alpha_1(starting\,WTP)_i + \alpha_2(starting\,WTP \times GreenFirst)_i +$

[13] It is noted that the low R-squares in regression models are due to the nature of the dependent variable (extra WTP_G)%, as its value cannot be directly driven by the magnitude of the recycling charge or recycling tax and is rather an indicator of the participants' adoption of the Green version. It is not uncommon to have these levels of R-squares in regressions, considering the nature of dependent or independent variables (see, for example, Allcott and Sweeney's 2017 regression results with low R-squares in the range of 0.01, 0.02, 0.04, 0.05 in a context similar to ours).

Table 4.6: The observed effect in Analysis 1 considering participants' demographics in the analysis.

	Regular/new buyers	Emergency/Replacement buyers
_intercept	6.45**	9.47**
	(1.992)	(2.309)
ageH	1.87	0.09
	(3.052)	(3.295)
male	-1.05	-5.94
	(2.577)	(3.066)
*ageH*male*	-1.51	3.28
	(4.436)	(4.545)
recycling tax	2.40	3.24
	(2.950)	(3.732)
recycling charge	8.60**	1.09
	(3.052)	(3.266)
*recycling tax*ageH*	-2.75	-3.88
	(4.432)	(5.214)
*recycling charge*ageH*	-6.17	-3.30
	(4.404)	(4.620)
*recycling tax*male*	-0.35	-0.32
	(4.037)	(4.697)
*recycling charge*male*	-7.28	1.00
	(4.029)	(4.377)
*recycling tax*ageH*male*	1.15	-1.75
	(6.567)	(6.912)
*recycling charge*ageH*male*	8.51	-1.80
	(6.292)	(6.410)
R^2	0.05	0.04
Adjusted R^2	0.02	0.01

*Note: Parenthesis include standard errors. Significant levels: *=0.05 and **=0.01.*

$\alpha_3(recycling\,tax)_i + \alpha_4(recycling\,charge)_i + \epsilon_i$

Table 4.6 shows the result of this regression. As seen, participants' starting WTP has zero effect on their adoption of the Green version irrespective of recycling policy and purchase situation. In line with the findings of Analysis 1, only the recycling charge with the charging-for-recycling scenario works as an informational anchor with a statistically significant effect on the adoption of the Green version among participants, and only in the regular/new purchase group.

Table 4.7: The effect of starting WTP on the adoption of the Green version.

	Regular/new buyers	Emergency/replacement buyers
_intercept	6.31**	4.75**
	(1.766)	(1.819)
starting WTP	0.00	0.00
	(0.003)	(0.002)
*starting WTP*GreenFirst*	0.00	0.00
	(0.003)	(0.002)
recycling tax	1.50	1.19
	(1.735)	(1.713)
recycling charge	4.71**	0.16
	(1.662)	(1.598)
R^2	0.04	0.03
Adjusted R^2	0.03	0.02

*Note: Parenthesis include standard errors. Significant levels: *=0.05 and **=0.01*

Discussion: The analysis above supports our Hypothesis 4 that, apart from the main source of difference considered in Hypothesis 3, the observed effect from charging for recycling is robust to common sources of consumer heterogeneity. The regression analysis using (extra WTP_G)%, which rules out the effect of income level and starting WTP, shows that the participants in the regular/new purchase situation, regardless of their demographics, anchor to the recycling charge ($25) to shape their subject-level valuation of the extra attribute "increased recyclability" in the Green version. In line with the findings of Analysis 1, this sort of anchoring is not statistically significant with the recycling tax, nor occurs in the emergency/replacement purchase situation. Our analysis confirms that the main and only source of heterogeneity in the effect of the recycling charge as a green nudge with the charging-for-recycling program is the purchase situation, and the effect is robust to other sources of consumer heterogeneity commonly considered in direct marketing. This result is in line with previous findings in nudge, for example, Azmat and Iriberri's (2010) observation that while a nudge can have heterogeneous effects in the presence of a strong source of change in behavior, it is robust to common sources of consumer heterogeneity.

4.4.3 Analysis 3: Characterizing the Effect in Terms of Green Market Expansion and Green Market Intensification

The objective of our analysis here is to characterize the observed effect from charging for recycling in terms of green market expansion and green market intensification. The motivation behind this analysis is some observations, for example, by Ho et al. (2016), that informational nudges to reduce negative externalities have stronger effects on people who are intrinsically inclined to be pro-social. Therefore, in this analysis, we aim to understand whether charging for recycling increases the number of participants with (extra WTP)%>0 and whether it intensifies the (extra WTP)% among participants. To do so, we first look at the portion of participants with (extra WTP)%>0 in each group. Figure 3 shows that this portion remains roughly the same across all groups (between 60-70%). Furthermore, as seen in Figures 4 and 5, we obtain a similar trend for participants with (extra WTP)%>0 to what we previously observed for all participants in Analysis 1: In the regular/new purchase situation, the recycling charge increases the average of (extra WTP$_G$)% by 6.3 compared with free recycling ($t = 2.964$, $p < 0.002$) and by 5.2 compared with the recycling tax ($t = 2.043$, $p < 0.022$); meanwhile, it makes no difference for emergency/replacement purchases (0.28 compared with free recycling, $t = 0.193$, $p > 0.847$; and 0.33 compared with the recycling tax, $t = 0.213$, $p > 0.831$). In addition, as Tables 4.7-4.10 show, we obtain similar qualitative results with mixed-model analysis and repeating Analysis 2 on consumer heterogeneity for participants with (extra WTP)%>0.

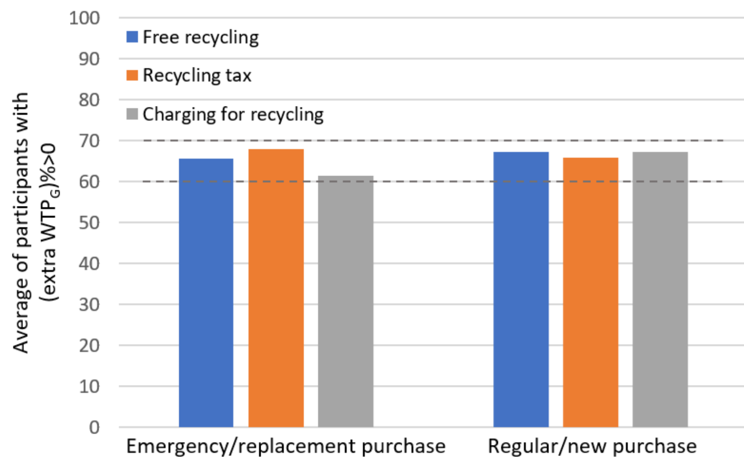

Figure 4.3: Portion of participants with (extra WTP$_G$)%¿0 for the two purchase situations with the three recycling policies.

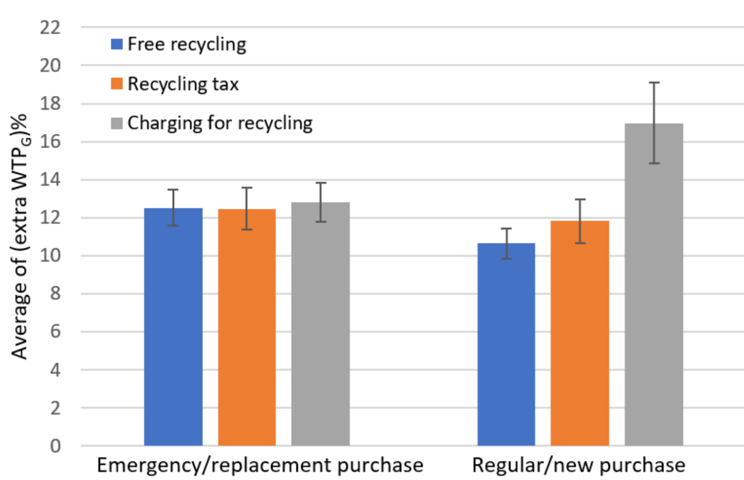

Figure 4.4: Average of (extra WTP$_G$)% for the two purchase situations with the three recycling policies for participants with (extra WTP$_G$)%¿0. *Note: Error bars are standard errors of means (SEM).*

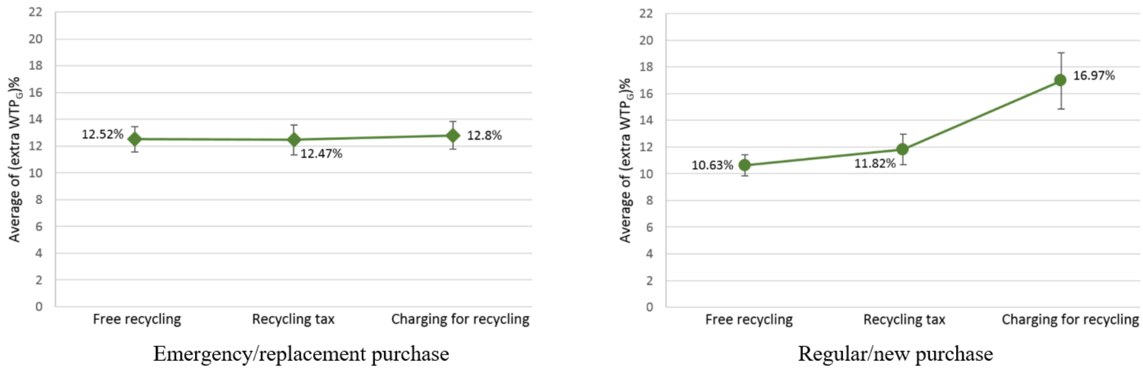

Figure 4.5: The effect of charging for recycling on the average of (extra WTP$_G$)% for the two purchase situations for participants with (extra WTP$_G$)%¿0. *Note: Error bars are standard errors of means (SEM).*

Table 4.8: Mixed-model analysis for participants with (extra WTP$_G$)%$>$0.

Term	Full model		
	F	Power	Effect size
version	298.92**	0.419	1.000
version*policy	1.06	0.005	0.235
version*policy*purchase	4.06*	0.019	0.721
version*policy*purchase*order	0.23	0.001	0.086

Note: Effect size is based on partial-eta-squared. Power reflects observed power.
*Significant levels: *=0.05 and **=0.01.*

Table 4.9: Mixed-model analysis including participants' demographics for participants with (extra WTP_G)%>0.

Term	Full model		
	F	Power	Effect size
version	279.61**	0.425	1.000
version*policy	0.42	0.002	0.117
*ageH	0.56	0.003	0.143
*male	0.59	0.003	0.149
*ageH*male	0.81	0.004	0.189
version*policy*purchase	3.29*	0.017	0.623
*ageH	0.74	0.004	0.175
*male	1.24	0.007	0.270
*ageH*male	1.16	0.006	0.254
version*policy*purchase*order	0.10	0.001	0.065
*ageH	0.62	0.003	0.154
*male	1.85	0.010	0.384
*ageH*male	0.49	0.003	0.130

Note: Effect size is based on partial-eta-squared. Power reflects observed power.
*Significant levels: *=0.05 and **=0.01.*

Discussion: In Analysis 3, we repeated our previous analyses only for participants with (extra WTP_G)%>0. With the baseline free recycling and recycling tax policies, around 60-70% of participants had (extra WTP_G)%>0. The observation made in Analysis 3 shows that the charging-for-recycling program illustrates its effect through intensifying the adoption of the Green version among this segment of participants. In other words, charging for recycling intensifies the willingness to adopt the Green version among the participants who have slight intentions to pay more for environmentally friendly and green products. This is in line with Ho's et al. (2016) observation that informational nudges in the domain of environmentalism have stronger effects on people with slight pro-social intentions. This further supports our general expectation in Hypotheses 1 and 2 that charging for recycling will improve self-efficacy in the consumer side to act upon their pre-existing environmental intentions. Nonetheless, the analysis shows that our expectation in Hypothesis 3 that the cost-oriented nature of emergency/replacement purchase situations will hinder such effects still applies. Finally, the analysis yields the same expectation as in Hypothesis 4 on

Table 4.10: The observed effect in Analysis 1 considering participants' demographics for participants with (extra WTP_G)%>0.

	Regular/new buyers	Emergency/replacement buyers
_intercept	11.27**	13.52**
	(2.212)	(1.931)
ageH	0.69	0.33
	(3.409)	(2.764)
male	-1.68	-2.90
	(5.893)	(2.644)
*ageH*male*	-0.04	1.48
	(5.202)	(4.015)
recycling tax	4.07	2.84
	(3.474)	(3.053)
recycling charge	7.96*	1.80
	(3.298)	(2.764)
*recycling tax*ageH*	-4.31	-2.94
	(5.097)	(4.388)
*recycling charge*ageH*	-5.22	-2.18
	(4.824)	(4.020)
*recycling tax*male*	-3.32	-1.69
	(4.640)	(3.964)
*recycling charge*male*	-3.77	-2.07
	(4.591)	(3.782)
*recycling tax*ageH*male*	2.627	-1.47
	(7.627)	(5.980)
*recycling charge*ageH*male*	5.59	2.45
	(7.289)	(5.767)
R^2	0.07	0.05
Adjusted R^2	0.02	0.00

*Note: Parenthesis include standard errors. Significant levels: *=0.05 and **=0.01.*

Table 4.11: The effect of starting WTP on the adoption of the Green version for participants with (extra WTP_G)%>0.

	Regular/new buyers	Emergency/replacement buyers
_intercept	12.96**	13.56**
	(2.061)	(1.652)
starting WTP	0.00	0.00
	(0.004)	(0.002)
*starting WTP*GreenFirst*	0.00	0.00
	(0.003)	(0.002)
recycling tax	0.60	-0.10
	(2.071)	(1.501)
recycling charge	5.98**	0.05
	(1.968)	(1.448)
R^2	0.07	0.02
Adjusted R^2	0.05	0.00

*Note: Parenthesis include standard errors. Significant levels: *=0.05 and **=0.01* consumer heterogeneity beyond the purchase situation.

4.5 General Discussion and Concluding Remarks

In this paper, we examined the effect of a charging-for-recycling program on consumer adoption of green versions of consumer electronics. Our experiments showed that, compared with free recycling and recycling tax policies, charging for recycling may enhance adoption of green versions by around 5%, on average. This finding is particularly rewarding from a green marketing point of view, given that costly direct green marketing strategies have been unsuccessful in increasing consumer adoption of energy-efficient products (Deutsch 2010a, Deutsch 2010b, Allcott and Sweeney 2017). Comparing it with one of the most promising findings in this area, which has found a 5% increase in adoption of energy-efficient products by incentivizing salespeople and rebating customers (Allcott and Sweeney 2017), the observed effect here has merits as an alternative green marketing strategy. More importantly, direct green marketing usually focuses on energy efficiency, highlighting the pecuniary benefits to customers from saving energy, which limits its application only to products with high energy consumption, such as air conditioners, refrigerators, washing machines, and dryers (Ward et al. 2011). Charging for recycling, on the other hand, uses the recyclability of goods and can thus target almost all consumer electronics, irrespective of their energy consumption level. In addition, charging for recycling as an indirect marketing strategy can avoid the green marketing myopia that occurs with some green advertisements (Ottman et al. 2006); can avoid cognitive limitations, such as hyperbolic discounting, that deter customers from making extra payments for energy efficiency (Anderson and Claxton 1982, Newell and Siikamäki 2014); and can be immune to common consumer heterogeneity issues in direct green marketing (Houde 2014). Moreover, it should be noted that the observed effect in this study was merely a byproduct of a recycling program for retailers that primarily aims to assist them with the operational costs of their recycling systems. Therefore, any positive impact, even a small one, is an extra benefit for retailers and the environment.

Our findings also suggest that current environmental laws that prohibit retailers from charging for recycling may be counterproductive. Free recycling that has been widely pursued (for instance, Directive 2012/19/EU, which mandates manufacturers to collect used consumer electronics from customers for recycling free of charge) primarily aims to increase the number of products being recycled in the economy as the initial step in creating a circular economy (Ellen MacArthur Foundation 2015, Lewandowski 2016). However, it may fail to achieve its purpose for two reasons: first, because retailers may be forced to leave the recycling business in order to avoid its operational costs; and second, because free recycling fails to focus consumer attention on green products at the point of purchase. Moreover, based on previous evidence of the strong positive correlation between green purchasing behavior and recycling behavior (Biswas et al. 2000), we can hope that consumer adoption of green versions will also result in higher willingness to recycle at the recycling point. As Merritt et al. (2010) discuss, a pro-social behavior of adopting a recyclable version may induce more pro-social behavior, such as a commitment to recycle at the end of a product's life, and there is abundant evidence that inducing commitment can have a significant long-term influence on recycling behavior (see Osbaldiston and Schott 2012, for a review).

In line with the above, and more related to the habit formation (see, e.g., Fuhrer 2000, Carroll et al. 2000), Best Buy's program may also change customers' purchase behavior for more consumer electronics than just TVs/monitors. In its current implementation, Best Buy's charging-for-recycling program announces the \$25 charge (for accepting TVs/monitors for recycling) on its recycling channel, where it collects many other products and where it also provides customers with the opportunity for "Buying Greener Electronics." Based on the abundant evidence on habit formations in consumer consumption/purchase behavior (see, e.g., Alessie and Lusardi 1997, Zhen et al. 2011), customers' adoption of green TVs/monitors driven by the charging-for-recycling program can also extend to adoption of green versions of other consumer electronics, as well. Even if such a habit extends only to a limited number of products (for example, computers, tablets, and printers, which customers may perceive to be in the same product family as TVs/monitors), the outcome can still be significant at the macro level.

Finally, our findings in this study also provide insights for other industries as to considering the sort of return programs and policies that can result in overall more positive impacts with the circular economy and social/environmental responsibility perspectives. One of these cases is the apparel industry and consumer adoption of organic clothing. Goodwill, for example, accepts used clothes free of charge to prepare them for reuse by those in need. Implementing a similar charging-for-return program here can potentially increase consumers' adoption of organic clothing at the purchase point. Such a consumer behavior will further encourage manufacturers towards more socially/environmentally responsible sourcing in their supply chains, which can further support and expand organic agriculture businesses.

Appendices

APPENDIX A

The original and new versions' specifications under a low innovation in the new version:

	Tablet 4.2	Tablet 5.0
Operating system	Android 4.2	Android 5.0
Processor speed (GHz)	Dual-Core 1.2 GHz	Quad-Core 1.2 GHz
RAM (GB)	1GB	1.5Gb
Memory (GB)	8GB	16GB
Battery life (hours)	9 Hours	13 Hours
Screen size (inch)	7.0"	8.0"
Resolution (px)	1024×600	1024×768
Size (inch)	4.58"×7.61"×0.38"	8.20"×5.43"×0.29"
Weight (lb)	0.68	0.69
Camera (MP)	2MP	5MP

The original and new versions' specifications under a high innovation in the new version:

	Tablet 4.2	Tablet 5.2
Operating system	Android 4.2	Android 5.2
Processor speed (GHz)	Dual-Core 1.2 GHz	Octa-Core (1.9+1.3) GHz
RAM (GB)	1GB	3Gb
Memory (GB)	8GB	32GB
Battery life (hours)	9 Hours	8 Hours
Screen size (inch)	7.0"	9.7"
Resolution (px)	1024×600	2048×1536
Size (inch)	4.58"×7.61"×0.38"	9.34"×6.65"×0.22"
Weight (lb)	0.68	0.86
Camera (MP)	2MP	8MP

Robustness check on using the same participants for both innovation levels:

The objective of our robustness check here is to ensure that no ordering effect occurs between the two questions for 'low innovation' and 'high innovation' with using the same participants for both innovation levels. The best way to check this is to repeat the analysis only with the first question that the participants answered. This would resemble an experiment that uses different participants for different innovation levels. Therefore, we collected relatively more responses in the first market setting, to have fairly enough data points after cutting them in half, and thus limit our robustness check to this market setting.

Table A1 shows the result of this robustness check. The 'LH participants' refers to the participants who answered the 'low innovation' question first and their responses is bing used for the 'low innovation' case, and the 'HL participants' are the participants who answered the 'high innovation' question first and we use their responses for the 'high innovation' case.

Table A1: Robustness check on using the same participants for both innovation level.

Low innovation

Trade-in (regression table with all participnts)

	Coeff.	Std. Err.	t	p_{value}
Sec	0.38	0.1028	3.70	0.000
$\_cons$	39.63	7.6206	5.20	0.000
$\text{Prob} > \text{F} = 0.000, \text{R}^2 = 0.069, \text{Adj.R}^2 = 0.064$				

(regression table with LH participants)

	Coeff.	Std. Err.	t	p_{value}
Sec	0.31	0.1572	1.95	0.054
$\_cons$	46.95	11.38	4.13	0.000
$\text{Prob} > \text{F} = 0.054, \text{R}^2 = 0.044, \text{Adj.R}^2 = 0.032$				

Upgrade (regression table with all participnts)

	Coeff.	Std. Err.	t	p_{value}
New	0.27	0.0111	24.47	0.000
$\text{Prob} > \text{F} = 0.000, \text{R}^2 = 0.772, \text{Adj.R}^2 = 0.771$				

(regression table with LH participants)

	Coeff.	Std. Err.	t	p_{value}
New	0.27	0.0174	15.61	0.000
$\text{Prob} > \text{F} = 0.000, \text{R}^2 = 0.753, \text{Adj.R}^2 = 0.750$				

High innovation

Trade-in (regression table with all participnts)

	Coeff.	Std. Err.	t	p_{value}
Sec	0.68	0.0954	7.18	0.000
$\_cons$	21.48	7.1440	3.01	0.003
$\text{Prob} > \text{F} = 0.000, \text{R}^2 = 0.223, \text{Adj.R}^2 = 0.219$				

(regression table with HL participants)

	Coeff.	Std. Err.	t	p_{value}
Sec	0.68	0.1267	5.35	0.000
$\_cons$	19.56	9.7044	2.02	0.047
$\text{Prob} > \text{F} = 0.000, \text{R}^2 = 0.233, \text{Adj.R}^2 = 0.225$				

Upgrade (regression table with all participnts)

	Coeff.	Std. Err.	t	p_{value}
$Curr$	0.60	0.0311	19.38	0.000
$\text{Prob} > \text{F} = 0.000, \text{R}^2 = 0.681, \text{Adj.R}^2 = 0.679$				

(regression table with HL participants)

	Coeff.	Std. Err.	t	p_{value}
$Curr$	0.66	0.0436	15.25	0.000
$\text{Prob} > \text{F} = 0.000, \text{R}^2 = 0.710, \text{Adj.R}^2 = 0.707$				

Further Analyses in Section 2.4:

In what follows, we repeat Regression (3) for when the original purchase price is the influential anchor and for when the new version's price is the influential anchor. With the the former, the $Curr$ in Regression (3) is replaced with Opp:

$$ref_j^U = \beta_o Opp_j + \beta_d dummy\_Sec_j + \beta_{od} Opp \times dummy\_Sec_j + \varepsilon_j$$

and with the latter, it is replaced with New:

$$ref_j^U = \beta_n New_j + \beta_d dummy\_Sec_j + \beta_{nd} New \times dummy\_Sec_j + \varepsilon_j$$

Table A2: Regression (3) for when the original purchase price is the influential anchor.

Variable	Coefficient
New	0.38
	(< 0.001)
dummy_Sec	-19.51
	(> 0.568)
$New \times dummy\_Sec$	-0.02
	(> 0.883)

Parentheses contain p-values. $R^2 = 0.752, Adj.R^2 = 0.751.$

Table A3: Regression (3) for when the new version's price is the influential anchor.

Variable	Coefficient
Opp	0.74
	(< 0.001)
dummy_Sec	-63.32
	(> 0.438)
$Opp \times dummy\_Sec$	0.06
	(> 0.876)

Parentheses contain p-values. $R^2 = 0.718, Adj.R^2 = 0.713.$

Proofs:

The manufacturer's optimal pricing in the anonymous case:

With $c_H = 0$ in Equation (9), Π_2^A reduces to $\Pi_2^A = [1 - \theta_H]\theta_H V_\Delta + [\theta_1 - \theta_L]\theta_L V_L$, which yields separate maximization problems for the new and the original versions in period two. Define $\theta^* = argmax(1-\theta)\theta$, and $\hat{\theta}^* = argmax(\theta_1 - \theta)\theta$ for a given θ_1. For uniform distribution, $\theta^* = \frac{1}{2}$ and $\hat{\theta}^* = \frac{\theta_1}{2}$. Using F.O.C. for θ_H and θ_L in Π_2^A, we get $\theta_H^* = \theta^*$ and $\theta_L^* = \hat{\theta}^*$. With that, we have $\Pi_2^{A*} = [1 - \theta^*]\theta^* V_\Delta + [\theta_1 - \hat{\theta}^*]\hat{\theta}^* V_L$. The manufacturer's total profit is $\Pi^A(\theta_1) = \Pi_1^A + \delta\Pi_2^{A*}$. First period's cut-off and price are obtained from $\theta_1 V_L + \delta\theta_L V_L - p_1 = 0$, where $\theta_L V_L$ is the guaranteed buyback price in period two either through the manufacturer or through the secondary market, and we have $\Pi_1^A = [1 - \theta_1][\theta_1 V_L + \delta\theta_L V_L]$ for the manufacturer's period-one profit. With backward induction, and after some mathematical simplifications, $\Pi^A(\theta_1) = \Pi_1^A + \delta\Pi_2^{A*} = [1 -$

$\theta_1]\theta_1 V_L + \delta(1-\theta^*)\theta^* V_\Delta + \delta(1-\hat{\theta}^*)\hat{\theta}^* V_L$. Using F.O.C. for θ_1, we get $\theta_1^* = \frac{1+\delta/2}{2+\delta/2}$. This gives $\theta_L^* = \frac{1+\delta/2}{4+\delta}$, and we already have $\theta_H^* = \frac{1}{2}$.

The manufacturer's optimal pricing in the semi-anonymous case, with no overlapping production in period two:

With $c_H = 0$ in Equation (7), Π_2^{Sn} reduces to $\Pi_2^{Sn} = [1-\theta_u][\theta_H V_H - \alpha p_1] + [\theta_1 - \theta_H]\theta_H V_H$. From the first expression, the optimal θ_u would be either as low as possible or equal to 1, depending on the sign of $\theta_H V_H - \alpha p_1$, thus $\theta_u^* = \theta_1$ or 1. Assume α is not very large so that $\theta_H V_H \geq \alpha p_1$ is guaranteed in optimality. Thus, $\Pi_2^{Sn} = [1-\theta_1][\theta_H V_H - \alpha p_1] + [\theta_1 - \theta_H]\theta_H V_H$. From period one, we have $p_1 = \theta_1 V_L - \delta\theta_1 V_\Delta + \delta\theta_H V_H$, by which Π_2^{Sn} can be rewritten as $\Pi_2^{Sn} = -\alpha[1-\theta_1][\theta_1 V_L - \delta\theta_1 V_\Delta + \delta\theta_H V_H] + [1-\theta_H]\theta_H V_H$. Using F.O.C. for θ_H yields $\theta_H^* = \frac{1-\alpha\delta(1-\theta_1)}{2}$. With backward induction, and given $\Pi_1^{Sn} = [1-\theta_1][\theta_1 V_L - \delta\theta_1 V_\Delta + \delta\theta_H V_H]$, after some mathematical simplifications: $\Pi^{Sn}(\theta_1) = \Pi_1^{Sn} + \delta\Pi_2^{Sn*} = (1-\alpha\delta)[1-\theta_1][\theta_1 V_L - \delta\theta_1 V_\Delta + \frac{\delta - \alpha\delta^2(1-\theta_1)}{2} V_H] + \delta[\frac{1}{4} - \frac{\alpha^2\delta^2(1-\theta_1)^2}{4}]V_H$. Using F.O.C. for θ_1 yields $\theta_1^* = \frac{(1-\alpha\delta)(V_L - \delta V_\Delta) - (1-\alpha\delta)\delta V_H/2 + (1-\alpha\delta/2)\alpha\delta^2 V_H}{2(1-\alpha\delta)(V_L-\delta V_\Delta) + (1-\alpha\delta/2)\alpha\delta^2 V_H}$, and we have $\theta_H^* = \frac{1-\alpha\delta(1-\theta_1^*)}{2}$ and $\theta_u^* = \theta_1^*$. This solution holds as long as $\theta_H^* < \theta_1^*$, since w.r.t. the market segmentation, θ_H cannot go beyond θ_1. Therefore, for when $\theta_H^* \geq \theta_1^*$, or equivalently $\frac{1-\alpha\delta(1-\theta_1^*)}{2} \geq \theta_1^*$ (or $\frac{1-\alpha\delta}{2-\alpha\delta} \geq \theta_1^*$), we would have $\theta_H^* = \theta_1^* = \frac{1-\alpha\delta}{2-\alpha\delta}$. This solution also guarantees $\theta_u^* = \theta_1^*$ since, with $\theta_H^* = \theta_1^* = \frac{1-\alpha\delta}{2-\alpha\delta}$, we always have $\theta_H V_H - \alpha p_1 \geq 0$. (It is worth noting that $p_u = \theta_H V_H - \alpha p_1 < \theta_H V_H = p_H$ for $\alpha > 0$.)

The manufacturer's optimal pricing in the semi-anonymous case, with overlapping production in period two:

With $c_H = 0$ in Equation (6), Π_2^{Sp} reduces to $\Pi_2^{Sp} = [1-\theta_u]\theta_H V_\Delta + [\theta_1 - \theta_H]\theta_H V_\Delta + [\theta_1 - \theta_L]\theta_L V_L$. From the first expression, the optimal θ_u would be as low as possible, thus $\theta_u^* = \theta_1$. With that, Π_2^{Sp} is simplified to $\Pi_2^{Sp} = [1-\theta_H]\theta_H V_\Delta + [\theta_1 - \theta_L]\theta_L V_L$, where we obtain $\theta_H^* = \theta^*$ and $\theta_L^* = \hat{\theta}^*$, using F.O.C. for θ_H and θ_L. From period one, because of the original version being offered in both periods, the cut-off θ_1 is obtained from $\theta_1 V_L + \delta\theta_1 V_L - p_1 = \delta\theta_1 V_L - \delta\theta_L V_H$; therefore, $p_1 = \theta_1 V_L + \delta\theta_L V_L$ and

$\Pi_1^{Sp} = [1-\theta_1][\theta_1 V_L + \delta\theta_L V_L]$. With backward induction, $\Pi^{Sp}(\theta_1) = \Pi_1^{Sp} + \delta\Pi_2^{Sp*} = [1-\theta_1]\theta_1 V_L + \delta[1-\theta^*]\theta^* V_\Delta + \delta[1-\hat{\theta}^*]\hat{\theta}^* V_L$, where $\theta^* = \frac{1}{2}$ and $\hat{\theta}^* = \frac{\theta_1}{2}$. Using F.O.C. for θ_1, we get $\theta_1^* = \frac{1+\delta/2}{2+\delta/2}$. This gives $\theta_L^* = \frac{1+\delta/2}{4+\delta}$, and we already have $\theta_H^* = \frac{1}{2}$. (It is worth mentioning that $p_u = \theta_H V_\Delta < \theta_H V_\Delta + \theta_L V_L = p_H$.)

Proof of Proposition 1. For $\alpha = 0$ in the results derived above, and after some mathematical simplifications, we have $\Pi^{Sn*}(\theta_1) = \frac{V_L}{4} + \delta\frac{V_\Delta}{4} + \delta\frac{V_L}{2}$, which is greater than $\Pi^{Sp}(\theta_1) = [1-\theta_1]\theta_1 V_L + \delta\frac{V_\Delta}{4} + \delta\frac{\theta_1^2}{4}V_L$ for any $\theta_1 \in [0,1]$. Assume α is such that αp_1 in the optimal solution for no overlapping production equals $\theta_L V_L$ in the optimal solution for overlapping production. Putting it simply, $[\alpha p_1^*]^{Sn} = [\theta_L^* V_L]^{Sp}$ (note that this α guarantees $\theta_H V_H \geq \alpha p_1$ in no overlapping production since $[\theta_L^*]^{Sp} \leq [\theta_H^*]^{Sn}$ and that $V_H > V_L$). With this α, overlapping production has already become more profitable than no overlapping production for two reasons: the magnitude of the reference point being the same in both while the market segmentation in the second period of the former pushes its optimal solution closer to the commitment solution. Given that Π^{Sn*} is strictly decreasing in α, there is an $\hat{\alpha}$ such that $\Pi^{Sn*} = \Pi^{Sp*}$ for $\alpha = \hat{\alpha}$, $\Pi^{Sn*} > \Pi^{Sp*}$ for $\alpha \in [0,\hat{\alpha})$, and $\Pi^{Sn*} < \Pi^{Sp*}$ for $\alpha \in (\hat{\alpha}, 1]$. This completes the proof.

Proof of Proposition 2. Based on Proposition 1, for $\alpha > \hat{\alpha}$ the manufacturer shifts to overlapping production, wherein we have $\Pi^{Sp}(\theta_1) = \Pi_1^{Sp} + \delta\Pi_2^{Sp*} = [1-\theta_1]\theta_1 V_L + \delta[1-\theta^*]\theta^* V_\Delta + \delta[1-\hat{\theta}^*]\hat{\theta}^* V_L$, which is the same as the manufacturer's profit function in the anonymous case. Using this ($\Pi^{Sp*} = \Pi^{A*}$) and the last part in the proof of Proposition 1, we have $\Pi^{S*} > \Pi^{A*}$ for $\alpha < \hat{\alpha}$, and $\Pi^{S*} = \Pi^{A*}$ for $\alpha \geq \hat{\alpha}$. This completes the proof.

APPENDIX B

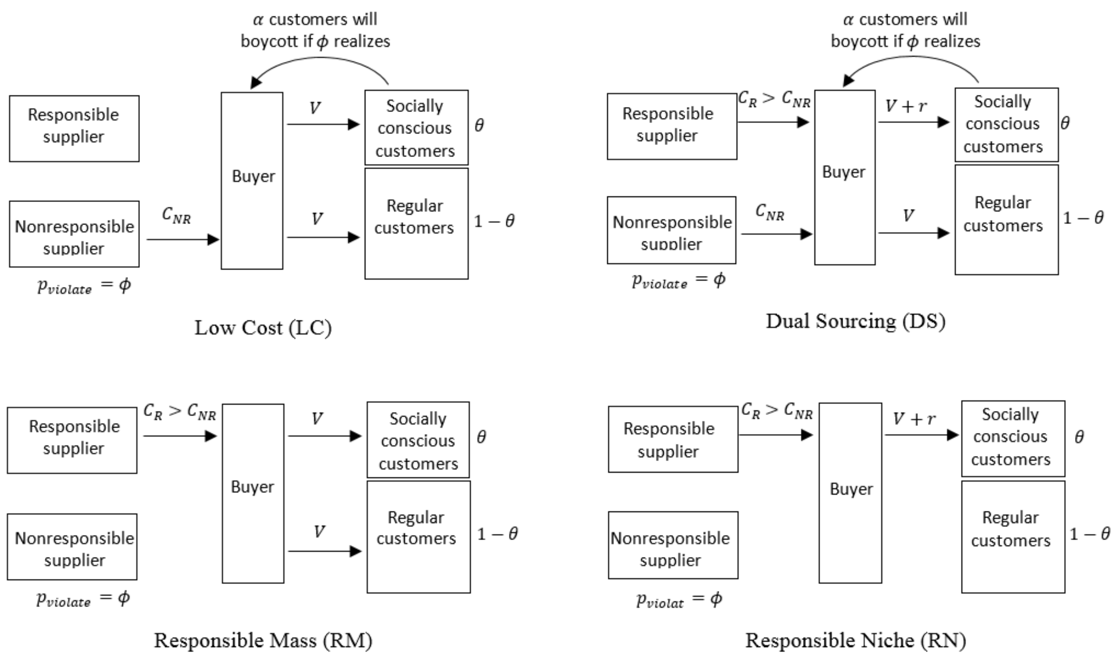

Figure B1: Four possible sourcing policies.

Figure B1 illustrates the four possible sourcing policies with the GLS model. It is noted that r and θ (representing the customers with extra willing-to-pay for responsibly sourced products) and α (representing the customers with willingness to boycott the buyer in the event of a violation on its suppliers' side) can appear independent from each other in the model, which creates two different socially responsible reactions. For example, when $r = 0$, i.e., when no customer is willing to pay more for responsibly sourced products, we can still see boycotting behavior by α customers. In this case, the only possible sourcing policies would be LC and RM, which are created due to the existence of α. On the other hand, this is the existence of $r > 0$ by θ customers, regardless of the existence or absence of α, that creates DS and RN policies (through creating a market segmentation and a price differentiation between θ and $1-\theta$ customers). Therefore, when there is no customer willing to pay more for responsibly sourced products (i.e., there is no notion of r and θ in the model), there can be no dual-sourcing bias either, and as Table 3.3 shows, both the GLS model and our behavioral model predict single sourcing (LC or RM, depending on the magnitude of α).

Proofs:

Proof of Lemma 1. LC policy dominates RM policy if $\Pi_{LC} > \Pi_{RM}$, that is, $(10,000 - \phi\alpha)(V - C_{NR}) > 10,000(V - C_R)$, which can be simplified to $C_R - C_{NR} > \phi\frac{\alpha}{10,000}(V - C_{NR})$. In the case of low salvage value (salvage value = 0), we have $(10,000 - \phi\alpha)V - 10,000C_{NR} > 10,000V - 10,000C_R$ for $\Pi_{LC} > \Pi_{RM}$, from which we get $C_R - C_{NR} > \phi\frac{\alpha}{10,000}V$.

Proof of Lemma 2. The buyer prefers LC over DS when LC dominates DS. For that, we should have $\Pi_{LC} > \Pi_{DS}$, that is, $(10,000 - \phi\alpha)(V - C_{NR}) > (10,000 - \theta)(V - C_{NR}) + (\theta - \phi\alpha)(V + r - C_R)$, which can be simplified to $(\theta - \phi\alpha)(C_R - C_{NR} - r) > 0$. In the case of low salvage value (salvage value = 0), we have $(10,000 - \phi\alpha)V - 10,000C_{NR} > (10,000 - \theta)V + (\theta - \phi\alpha)(V + r) - (10,000 - \theta)C_{NR} - \theta C_R$ for $\Pi_{LC} > \Pi_{DS}$, from which we get $\theta(C_R - C_{NR}) - (\theta - \phi\alpha)r > 0$.

Proof of Lemma 3. Similar to the proof of Lemma 2, the buyer prefers RM over DS when RM dominates DS. For that, we should have $\Pi_{RM} > \Pi_{DS}$, that is, $10,000(V - C_R) > (10,000 - \theta)(V - C_{NR}) + (\theta - \phi\alpha)(V + r - C_R)$, which can be simplified to $\phi\alpha(V - C_{NR}) - 10,000(C_R - C_{NR}) - (\theta - \phi\alpha)(r - (C_R - C_{NR})) > 0$. In the case of low salvage value (salvage value = 0), we have $10,000V - 10,000C_R > (10,000 - \theta)V + (\theta - \phi\alpha)(V + r) - (10,000 - \theta)C_{NR} - \theta C_R$ for $\Pi_{RM} > \Pi_{DS}$, from which we get $\phi\alpha V - 10,000(C_R - C_{NR}) + \theta(C_R - C_{NR}) - (\theta - \phi\alpha)r > 0$.

In the interest of mathematical tractability, proofs of Propositions 1 and 2 are presented considering uniform values of β's irrespective of the absence or presence of the encouraging reaction. All proofs hold, and strongly so, if we consider a lower β in the presence of the encouraging reaction, since it further pushes the inequalities in the directions of the proofs.

Proof of Proposition 1. When the discouraging reaction (i.e., α) is low, the classical model's prediction is either LC or DS, depending on the level of the encouraging reaction (i.e., r and θ). Suppose r and θ are low so that the classical model predicts LS. With the behavioral model presented in Equations (1a)-(1b) and (2a)-(2b), the presence of the encouraging reaction changes the optimal responsible sourcing from $0-\beta[1-(\frac{\alpha}{10,000})^{\varepsilon}]0+\beta[0+(\frac{\alpha}{10,000})^{\varepsilon}](10,000-0)$ to $\theta-\beta[1-(\frac{\alpha}{\theta})^{\varepsilon}]\theta+\beta[0+(\frac{\alpha}{\theta})^{\varepsilon}](10,000-\theta)$. Therefore, the encouraging reaction is helpful with responsible sourcing if $\theta-\beta[1-(\frac{\alpha}{\theta})^{\varepsilon}]\theta+\beta[0+(\frac{\alpha}{\theta})^{\varepsilon}](10,000-\theta)>0-\beta[1-(\frac{\alpha}{10,000})^{\varepsilon}]0+\beta[0+(\frac{\alpha}{10,000})^{\varepsilon}](10,000-0)$. With some mathematical simplifications, this inequality can be simplified to $\theta-\beta\theta+\beta(\frac{\alpha}{\theta})^{\varepsilon}10,000>\beta(\frac{\alpha}{10,000})^{\varepsilon}10,000$, and further to $(1-\beta)\theta>[(\frac{\alpha}{10,000})^{\varepsilon}-(\frac{\alpha}{\theta})^{\varepsilon}]10,000\beta$. Because $\theta<10,000$, we have $(\frac{\alpha}{10,000})^{\varepsilon}<(\frac{\alpha}{\theta})^{\varepsilon}$ for any $0<\varepsilon<1$, and hence the right-hand side is always negative for any $0<\beta<1$. Meanwhile, the left-hand side is always positive because $0<1-\beta<1$. Thus, the presence of the encouraging reaction improves responsible sourcing by making a preference for DS over LC. Now suppose r and θ are high so that the classical model predicts DS. In this case, the encouraging reaction again helps with responsible sourcing by making a preference for DS over LC.

Similarly, when the discouraging reaction (i.e., α) is high, the classical model's prediction is either RM or DS, depending on the level of the encouraging reaction (i.e., r and θ). Suppose r and θ are low so that the classical model predicts RM. With the behavioral model presented in Equations (1a)-(1b) and (3a)-(3b), the presence of the encouraging reaction changes the optimal responsible sourcing from $10,000-\beta[0+(\frac{10,000-\alpha}{10,000})^{1/\varepsilon}]10,000+\beta[1-(\frac{10,000-\alpha}{10,000})^{1/\varepsilon}]0$ to $\theta-\beta[0+(\frac{\theta-\alpha}{\theta})^{1/\varepsilon}]\theta+\beta[1-(\frac{\theta-\alpha}{\theta})^{1/\varepsilon}](10,000-\theta)$. Therefore, the encouraging reaction is helpful with responsible sourcing if $\theta-\beta[0+(\frac{\theta-\alpha}{\theta})^{1/\varepsilon}]\theta+\beta[1-(\frac{\theta-\alpha}{\theta})^{1/\varepsilon}](10,000-\theta)>10,000-\beta[0+(\frac{10,000-\alpha}{10,000})^{1/\varepsilon}]10,000+\beta[1-(\frac{10,000-\alpha}{10,000})^{1/\varepsilon}]0$. With some mathematical simplifications, this inequality can be simplified to $(1-\beta)\theta-(1-\beta)10,000 > [(\frac{\theta-\alpha}{\theta})^{1/\varepsilon}-(\frac{10,000-\alpha}{10,000})^{1/\varepsilon}]10,000\beta$. Both sides of this inequality are nonpositive and increasing in θ, but with different slopes. Thus, we shall look at the bounds of θ. The θ varies between α and 10,000. In the extreme case, if θ is 10,000, the inequality still does not hold (and needless to mention that when θ becomes that high, the RM is no longer the optimal policy). When θ equals α, the inequality changes to $(1-\beta)\alpha-(1-\beta)10,000 > -(\frac{10,000-\alpha}{10,000})^{1/\varepsilon}10,000\beta$. For a given α, because $1/\varepsilon > 1$, $\frac{10,000-\alpha}{10,000})^{1/\varepsilon}$ in its extreme cases becomes $(\frac{10,000-\alpha}{10,000})$ and 0. When $(\frac{10,000-\alpha}{10,000})^{1/\varepsilon} = 0$, the inequality changes to $(1-\beta)\alpha-(1-\beta)10,000 > 0$, which can never hold. When $(\frac{10,000-\alpha}{10,000})^{1/\varepsilon} = (\frac{10,000-\alpha}{10,000})$, the inequality changed to $(1-\beta)\alpha-(1-\beta)10,000 > -(10000-\alpha)\beta$, or equivalently $(1-2\beta)\alpha-(1-2\beta)10,000 > 0$, which again will never hold. Thus, the presence of the encouraging reaction is detrimental to responsible sourcing by making a preference for DS over RM. Now suppose r and θ are high so that the classical model predicts DS. In this case, the encouraging reaction is again detrimental to responsible sourcing by making a preference for DS over RM. This completes the proof.

Proof of Proposition 2. Here, in the case of a low salvage value (e.g., salvage value = 0), increasing the discouraging reaction (i.e., α) shows detrimental effect to responsible sourcing by changing the classical optimal policy from DS to LC. With the behavioral model presented in Equations (1a)-(1a) and (2a)-(2b), in the presence of the encouraging reaction, an increasing α changes the optimal nonresponsible sourcing from $(10,000-\theta)+\beta[0+(\frac{\theta-\alpha_{low}}{\theta})^{1/\varepsilon}]\theta-\beta[0+(\frac{\alpha_{low}}{\theta})^{\varepsilon}](10,000-\theta)$ to $(10,000-\theta)+\beta[1-(\frac{\alpha_{high}}{\theta})^{\varepsilon}]\theta-\beta[0+(\frac{\alpha_{high}}{\theta})^{\varepsilon}](10,000-\theta)$ rather than $10,000+\beta[1-(\frac{\alpha_{high}}{\theta})^{\varepsilon}]0-\beta[0+(\frac{\alpha_{high}}{\theta})^{\varepsilon}]10,000$. Therefore, increasing α is detrimental to responsible sourcing if $(10,000-\theta)+\beta[1-(\frac{\alpha_{high}}{\theta})^{\varepsilon}]\theta-\beta[0+(\frac{\alpha_{high}}{\theta})^{\varepsilon}](10,000-\theta) > (10,000-\theta)+\beta[0+(\frac{\theta-\alpha_{low}}{\theta})^{1/\varepsilon}]\theta-\beta[0+(\frac{\alpha_{low}}{10,000})^{\varepsilon}](10,000-\theta)$. With some mathematical simplifications, this inequality can be simplified to $\theta-(\frac{\alpha_{high}}{\theta})^{\varepsilon}10,000-(\frac{\theta-\alpha_{low}}{\theta})^{1/\varepsilon}\theta+(\frac{\alpha_{low}}{\theta})^{\varepsilon}(10,000-\theta) > 0$.

The left-hand side is strictly decreasing in α_{high} and is strictly increasing in α_{low}, for any $0 < \varepsilon < 1$. The α_{low} varies between 0 and α_{high}, wherein the α_{high} is restricted by θ. When $\alpha_{low} = 0$, we have the inequality as $-(\frac{\alpha_{high}}{\theta})^{\varepsilon}10,000 > 0$, which never holds; and when $\alpha_{low} = \alpha_{high} = \theta$, we have $0 > 0$. This completes the proof.

Experimental Descriptions:[1]

[Starting paragraphs for all Scenarios and groups]

You are selling a product whose retail price, i.e., the price that you receive from customers when you sell the product to them, is $\mathbf{\$80}$ *per unit. The total demand for the product is constant and equal to* $\mathbf{10,000}$ *units, i.e., there are* $\mathbf{10,000}$ *customers in the market, and each customer will buy only one product. To sell the product to these customers, you need to procure it from a supplier in advance. You can procure the product from two suppliers: an expensive but socially responsible supplier (supplier R) and/or a cheap but nonresponsible supplier (supplier N).*

Supplier R abides with social and environmental obligations, and there is no chance that it violates these obligations. Supplier N has a $\boldsymbol{\phi\%}$ *chance of violating these obligations. The purchasing cost, i.e., the price that you have to pay the supplier to procure the product, from supplier R is* $\boldsymbol{\$C_R}$ *per unit, and the purchasing cost from supplier N is* $\boldsymbol{\$C_{NR}}$ *per unit.*

[Control groups of Scenarios 1 and 2]

$\boldsymbol{\alpha}$ *of the* $\mathbf{10,000}$ *customers will refuse to buy the product from you if any violation occurs on your suppliers' side. Product that you have procured but not sold due to a violation will be returned to the supplier for full refund of the purchase cost.*

[Treatment groups of Scenarios 1 and 2]

$\boldsymbol{\theta}$ *of the* $\mathbf{10,000}$ *customers are socially conscious customers. These customers will pay*

[1]Parameter values in each round are randomly drawn from the intervals presented in Table B1.

$\$r$ more for the product, i.e., $\$(80 + r)$ instead of $\$80$, if the product is procured from supplier R instead of supplier N. In addition, α of these θ socially conscious customers will refuse to buy the product from you if a violation occurs on your suppliers' side. Products that you have procured but not sold due to a violation will be returned to the supplier for full refund of the purchase cost.

If you procure from the same supplier to both customer groups (socially conscious and regular), you cannot offer the product at a different price for different customer groups, and you will therefore sell the product at the same price of $\$80$ to all customers.

[Control and treatment groups of Scenario 3]

θ of the $10,000$ customers are socially conscious customers. These customers will pay $\$r$ more for the product, i.e., $\$(80 + r)$ instead of $\$80$, if the product is procured from supplier R instead of supplier N. In addition, α of these θ socially conscious customers will refuse to buy the product from you if a violation occurs on your suppliers' side. Products that you have not sold will have no salvage value.

If you procure from the same supplier to both customer groups (socially conscious and regular), you cannot offer the product at a different price for different customer groups, and you will therefore sell the product at the same price of $\$80$ to all customers.

[Closing paragraph for all Scenarios and groups]

Your objective is $to\, maximize\, your\, profit$. What are your order quantities from supplier R and supplier N?

Order quantity from supplier R: _____ *Order quantity from supplier N:* _____

Parameter ranges in the experiment:

Table B1 shows the parameter ranges from which the parameter values in each round were randomly drawn. Parameter ranges are choses in a way to keep the same ranges for the control and treatment groups and keep the source of difference at the parameter of ma-

Table B1: Parameter ranges in the experiment.

<table>
<tr><th>Scenario</th><th>Control group</th><th>Treatment group</th></tr>
<tr><td rowspan="3">Preference for LC over DS</td><td colspan="2">$D = 10{,}000$, $V = 80$, $C_R = V - \{25,30,35\}$, $C_{NR} = C_R - \{10,15\}$,
$\alpha = \{1200,1400,1600,1800\}$, $\phi = \{20,25,30\}\%$,</td></tr>
<tr><td>r=0 and $\theta = \alpha^{\dagger}$</td><td>$r = \{5, C_R - C_{NR} - 5\}$,
$\theta = \{4200,4400,\ldots,7200\}$</td></tr>
<tr><td>LC profit=$411,518</td><td>DS profit=$376,951</td></tr>
<tr><td rowspan="3">Preference for RM over DS</td><td colspan="2">$D = 10{,}000$, $V = 80$, $C_R = V - \{45,50,55\}$, $C_{NR} = C_R - \{5\}$,
$\alpha = \begin{cases} 3200,\ldots,3800 & r = C_R - C_{NR} \\ 3800 & r \neq C_R - C_{NR} \end{cases}$, $\phi = \begin{cases} 35,40 & r = C_R - C_{NR} \\ 40 & r \neq C_R - C_{NR} \end{cases}\%$,</td></tr>
<tr><td>r=0 and $\theta = \alpha^{\dagger}$</td><td>$r = \begin{cases} 5,10 & \theta < 6000 \\ 5 & \theta \geq 6000 \end{cases}$,
$\theta = \{4200,4400,\ldots,7200\}$</td></tr>
<tr><td>RM profit=$501,200</td><td>DS profit=$481,479</td></tr>
<tr><td rowspan="3">Preference for LC over DS
(the case of low salvage value)</td><td colspan="2">$D = 10{,}000$, $V = 80$, $C_R = V - \{20,25,30\}$, $C_{NR} = C_R - \{25,30\}$,
$r = \{C_R - C_{NR} + 5\}$, $\phi = \{30,35\}\%$, $\theta = \{5200,5400,\ldots,6000,6400\}$,</td></tr>
<tr><td>$\alpha = \begin{cases} 1200,1400 & \theta < 6000 \\ 1200,\ldots,1800 & \theta \geq 6000 \end{cases}$</td><td>$\alpha = \begin{cases} 4200,\ldots,4800 & \theta < 6000 \\ 4600,4800 & \theta \geq 6000 \end{cases}$</td></tr>
<tr><td>DS profit=$503,445
LC profit=$489,108</td><td>DS profit= $390,285
LC profit=$410,139</td></tr>
</table>

Note. Final payments to participants (after completing all 20 rounds) were adjusted by dividing the final profits in the control/treatment group of the first scenario and the treatment group of the third scenario by 3,000,000, and dividing the final profits in the control/treatment group of the second scenario and the control group of the third scenario by 3,500,000, where the profits with the theoretical optimal policy are in the order of $400,000 and $500,000, respectively.

†Since there is no customer willing to pay and extra $r > 0$ for responsibly sourced products, we do not mention r and θ in the control group and only refer to customers with willingness to boycott (α); hence, r=0 and $\theta = \alpha$.

nipulation. Profits presented in each scenario are the average profits over 500 simulation runs (in each simulation run, a set of parameter values was generated by randomly drawing the parameter values from the parameter ranges, and the profits were calculated using the GLS formulations). As seen, in each scenario, the buyer will earn more profit by choosing the rational optimal sourcing policy, while the dual-sourcing bias in the treatment groups will result in choosing a sourcing policy with lower profit. It is noted that we do not use an affine transformation of the profits, similar to what Ho et al. (2010) did in their newsvendor problem, to create significant gaps in the profits as it is not the focus of our study. That is, in their newsvendor problem, it is important to study how much profit the newsvendor loses due to not sticking with the theoretical optimal, and hence such significant gaps matter in studying the magnitude of the effect. In our context, however, the effect is about changes in the amount of nonresponsible sourcing resulted from not sticking with the theoretical optimal policy, which we have built our analyses on. For instance, in the first scenario in Table B1, even if DS and LC had the exact same profit, what matters for our study is how the existence of the encouraging reaction in the treatment group results in a significantly lower amount of nonresponsible sourcing in this group, compared with that in the control group. As mentioned in the experimental procedure, final payments (after completing all 20 rounds) were adjusted in order to keep reasonable and comparable monetary payments in all groups (see the footnote in Table B1).

APPENDIX C

Experimental Descriptions:

Below are the experimental descriptions for the Standard-Green order. In the Green-Standard order, Question 1 asks about the willingness to pay for the Green version, and then Question 2 asks about the willingness to pays for the Green and Standard versions.

Question 1

[*for regular/new buyers*] Imagine you already have an almost-new TV at home. You are thinking about buying an extra one for another room.

[*for emergency/replacement buyers*] Imagine your TV is not working anymore, and you need to buy a new one.

[*with free recycling*] One of the largest electronic retailers would be a good place to shop. This retailer also has a recycling program and accepts broken TVs for recycling free of charge.

[*with $7 recycling tax*] One of the largest electronic retailers would be a good place to shop. This retailer also has a recycling program and, due to the cost of managing TV recycling, charges a $7 recycling fee in addition to the sale price for all TVs, regardless of the brand or technical specifications.

[*with $25 charge for recycling*] One of the largest electronic retailers would be a good place to shop. This retailer also has a recycling program and, due to the cost of managing TV recycling, charges $25 to accept broken TVs for recycling.

[*with free recycling and $25 charge for recycling*] Assume that the technical specifications of the TV below meet your basic criteria. How much (in dollars) would you be willing to pay for this TV?

[*with $7 recycling tax*] Assume that the technical specifications of the TV below meet your basic criteria. How much (in dollars) would you be willing to pay for this TV (inclusive of the $7 fee)?

Price: ——-

Question 2

You indicated that you would pay $[*price indicated in Question 1*] for the TV you were shown.

Assume you find there is also a Green version of that TV (as shown below). It has the exact same features as the standard one, while it is also designed for increased recyclability at its end-of-life.

[with free recycling and $25 charge for recycling] In light of this new information, how much (in dollar) would you be willing to pay for each of these TVs?

[with $7 recycling tax] In light of this new information, how much (in dollar) would you be willing to pay for each of these TVs (inclusive of the $7 fee)?

Standard Version
- 48.5" LED screen
- Chromecast Built-in
- Google Home and Google Assistant
- 2160p resolution and ultra HD-level quality
- Wireless Connectivity
- Two 12W speakers, DTS Studio Sound
- 3 HDMI inputs and 1 USB input

Green Version
- 48.5" LED screen
- Chromecast Built-in
- Google Home and Google Assistant
- 2160p resolution and ultra HD-level quality
- Wireless Connectivity
- Two 12W speakers, DTS Studio Sound
- 3 HDMI inputs and 1 USB input
- Increased Recyclability

Price for the Standard version: ——-

Price for the Green version: ——-

REFERENCES

Abbey JD, Blackburn JD, Guide Jr VDR (2015) Optimal pricing for new and remanufactured products. *Journal of Operations Management* 36:130-146.

Abbey JD, Meloy MG (2017) Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management* 53:63-70.

Agrawal V, Ferguson M, Souza GC (2016) Trade-in rebates for price discrimination and product recovery. *IEEE Transactions on Engineering Management* 63(3):326-339.

Agrawal V, Lee D (2019) The effect of sourcing policies on a supplier's sustainable practices. *Production and Operations Management* 28(4):767-787.

Ajzen I (1991) The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 50(2):179-211.

Alessie R, Lusardi A (1997) Consumption, saving and habit formation. Economics letters, 55(1):103-108.

Allcott H, Greenstone M (2012) Is there an energy efficiency gap? *The Journal of Economic Perspectives* 26(1):3-28.

Allcott H, Rogers T (2014) The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *The American Economic Review* 104(10);3003-3037.

Allcott H, Sweeney RL (2017) The role of sales agents in information disclosure: evidence from a field experiment. *Management Science* 63(1):21-39.

Allcott H, Taubinsky D (2015) Evaluating behaviorally motivated policy: experimental evidence from the lightbulb market. *The American Economic Review* 105(8):2501-2538.

Al-Mahmood S, Passaeriello C, Rana P (2013) The global garment trail: From Bangladesh to a mall near you. *Wall Street Journal* (May 3), http :// www.wsj.com / articles / SB 10001424127 887324766604578460833869722240.

Anderson CD, Claxton JD (1982) Barriers to consumer choice of energy efficient products. *Journal of Consumer Research*, 9(2):163-170.

Appleinsider (2016) Apple R&D spending jumps to $2.5B in Q2, accounted for 5% of total revenue. *appleinsider.com* (April 27), http://appleinsider.com/articles/16/04/27/apple-rd-spending-jumps-to-25b-inq2-accounted-for-5-of-total-revenue/.

Ariely D, Loewenstein G, Prelec D (2003) "Coherent arbitrariness": Stable demand curves without stable preferences. *The Quarterly Journal of Economics* 118(1):73-106.

Azmat G, Iriberri N (2010) The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics* 94(7-8):435-452.

Balde CP, Wang F, Kuehr R, Huisman J (2015) The global e-waste monitor–2014, United Nations University, IAS–SCYCLE, Bonn, Germany.

Bao J, Ho B (2015) Heterogeneous effects of informational nudges on pro-social behavior. The BE Journal of Economic *Analysis & Policy* 15(4):1619-1655.

Barboza D (2010) After suicides, scrutiny of China's grim factories. *New York Times* (June 6), http://www.nytimes.com/2010/06/07/business/global/07suicide.html.

Baucells M, Weber M, Welfens F (2011) Reference-point formation and updating. *Management Science* 57(3):506-519.

Bayus BL (1991) The consumer durable replacement buyer. *The Journal of Marketing* 55(1):42-51.

Benartzi S, Thaler RH (2001) Naïve diversification strategies in defined contribution saving plans. *American Economic Review* 91(1):79-98.

Benjamin DK, Kormendi RC (1974) The interrelationship between markets for new and used durable goods. *The Journal of Law and Economics* 17(2):381-401.

Berger IE, Corbin RM (1992) Perceived consumer effectiveness and faith in others as moderators of environmentally responsible behaviors. *Journal of Public Policy & Marketing* 11(2):79-89.

Beshears J, Choi JJ, Laibson D, Madrian BC, Milkman KL (2015) The effect of providing peer information on retirement savings decisions. *The Journal of Finance* 70(3):1161-1201.

Biswas A, Licata JW, McKee D, Pullig C, Daughtridge C (2000) The recycling cycle: An empirical examination of consumer waste recycling and recycling shopping behaviors. *Journal of Public Policy & Marketing* 19(1):93-105.

Bondareva M, Pinker E (2019) Dynamic relational contracts for quality enforcement in supply chains. *Management Science* 65(3):1305-1321.

Bronchetti ET, Dee TS, Huffman DB, Magenheim E (2011) When a nudge isn't enough: defaults and saving among low-income tax filers. *National Bureau of Economic Research,* Paper No. 16887, March 2011.

Brown H (2007) Gap admits possible child labor problem. *ABC News* (October 28), https://abcnews.go.com/WN/story?id=3787304&page=1.

Burnham KP, Anderson DR (2004) Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research* 33(2):261-304.

Carmon Z, Ariely D (2000) Focusing on the forgone: How value can appear so different to buyers and sellers. *Journal of Consumer Research* 27(3):360-370.

Carroll CD, Overland J, Weil DN (2000) Saving and growth with habit formation. American Economic Review 90(3):341-355.

Chan RY, Lau LB (2002) Explaining green purchasing behavior: A cross-cultural study on American and Chinese consumers. *Journal of International Consumer Marketing* 14(2-3):9-40.

Chen C (2001) Design for the environment: A quality-based model for green product development. *Management Science* 47(2):250-263.

Chen L, Lee HL (2017) Sourcing under supplier responsibility risk: The effects of certification, audit, and contingency payment. *Management Science* 63(9):2795-2812.

Cheng PY, Chiou WB (2008) Framing effects in group investment decision making: Role of group polarization. *Psychological reports* 102(1):283-292.

Chetty R (2015) Behavioral economics and public policy: A pragmatic perspective. *American Economic Review* 105(5):1-33.

Chod J, Trichakis N, Tsoukalas G (2019) Supplier diversification under buyer risk. Forthcoming in *Management Science*.

Clogg CC, Petkova E, Haritou A (1995) Statistical methods for comparing regression coefficients between models. *American Journal of Sociology* 100(5):1261-1293.

Cripps JD, Meyer RJ (1994) Heuristics and biases in timing the replacement of durable products. *Journal of Consumer Research* 21(2):304-318.

Csermely T, Minner S (2015) Behavioral decisions in inventory systems with dual sourcing option: An experiment. Working Paper, Vienna University of Economics and Business, Vienna, Austria.

Demeester L, Qi M, Van Wassenhove LN (2013) Plant networks for processing recyclable materials. *Manufacturing & Service Operations Management* 15(4):670-688.

Desai PS (2001) Quality segmentation in spatial markets: When does cannibalization affect product line design? *Marketing Science* 20(3):265-283.

Deutsch M (2010a) Life cycle cost disclosure, consumer behavior, and business implications. *Journal of Industrial Ecology* 14(1):103-120.

Deutsch M (2010b) The effect of life-cycle cost disclosure on consumer behavior: Evidence from a field experiment with cooling appliances. *Energy Efficiency* 3(4):303-315.

Dolan P, Metcalfe R (2013) Neighbors, knowledge, and nuggets: Two natural field experiments on the role of incentives on energy conservation. CEP Discussion Paper No. 1222, London School of Economics and Political Science, London.

Dutta P, Das D, Schultmann F, Frohling M (2016) Design and planning of a closed-loop supply chain with three way recovery and buy-back offer. *Journal of Cleaner Production* 135:604-619.

Ellen MacArthur Foundation (2015) Towards the Circular Economy: Economic and Business Rationale for an Accelerated Transition. (December 31) Available online: http://mvonederland.nl/system/files/media/towards-the-circular-economy.pdf.

Ellen PS, Wiener JL, Cobb-Walgren C (1991) The role of perceived consumer effectiveness in motivating environmentally conscious behaviors. *Journal of Public Policy & Marketing* 10(2):102-117.

Ferraro PJ, Price MK (2013) Using nonpecuniary strategies to influence behavior: evidence from a large-scale field experiment. *Review of Economics and Statistics* 95(1):64-73.

Fox CR, Clemen RT (2005) Subjective probability assessment in decision analysis: Partition dependence and bias toward the ignorance prior. *Management Science* 51(9):1417-1432.

Fox CR, Ratner RK, Lieb DS (2005) How subjective grouping of options influences choice and allocation: Diversification bias and the phenomenon of partition dependence. *Journal of Experimental Psychology:* General 134(4):538-551.

Fudenberg D, Tirole J (1998) Upgrades, tradeins, and buybacks. *RAND Journal of Economics* 29(2):235-258.

Fuhrer JC (2000) Habit formation in consumption and its implications for monetary-policy models. American Economic Review 90(3):367-390.

Gachter S, Orzen H, Renner E, Starmer C (2009) Are experimental economists prone to framing effects? A natural field experiment. *Journal of Economic Behavior & Organization* 70(3):443-446.

Gallagher KM, Updegraff JA (2012) Health message framing effects on attitudes, intentions, and behavior: a meta-analytic review. *Annals of behavioral medicine* 43(1):101-116.

Geyer R, Kuczenski B, Zink T, Henderson A (2015) Common misconceptions about recycling. *Journal of Industrial Ecology* 20(5):1010-1017.

Ginsberg JM, Bloom PN (2004) Choosing the right green-marketing strategy. *MIT Sloan Management Review* 46(1):79-84.

Goldschmidt K, Kremer M, Thomas D, Craighead C (2014) Supply base diversification in the presence of high impact, low probability supply disruptions. Working Paper, University of St. Thomas, St. Paul, MN.

Golson J (2016) More than half of iPhone users haven't upgraded to a 6 or 6s yet. *Yahoo Tech* (January 26), http://www.yahoo.com/tech/more-half-iphone-users-haven-003807335.html.

Gordon BR (2009) A dynamic model of consumer replacement cycles in the PC processor industry. *Marketing Science* 28(5):846-867.

Guide VDR, Wassenhove LN (2001) Managing product returns for remanufacturing. *Production and operations management* 10(2):142-155.

Guiltinan J (2010) Consumer durables replacement decision-making: An overview and research agenda. *Marketing Letters* 21(2):163-174.

Guo R, Lee HL, Swinney R (2016) Responsible sourcing in supply chains. *Management Science* 62(9):2722-2744.

Gurnani H, Ramachandran K, Ray S, Xia Y (2013) Ordering behavior under supply risk: An experimental investigation. *Manufacturing & Service Operations Management* 16(1):61-75.

Hainmueller J, Hiscox MJ, Sequeira S (2015) Consumer demand for the Fair Trade label: Evidence from a multi-store field experiment. *Review of Economics and Statistics* 97(2):242-256.

Hardisty DJ, Pfeffer J (2017) Intertemporal uncertainty avoidance: When the future is uncertain, people prefer the present, and when the present is uncertain, people prefer the future. *Management Science* 63(2):519-527.

Hausman DM, Welch B (2010) Debate: To nudge or not to nudge. *Journal of Political Philosophy* 18(1):123-136.

Heath TB, Chatterjee S, France KR (1995) Mental accounting and changes in price: The frame dependence of reference dependence. *Journal of Consumer Research* 22(1):90-97.

Hendel I, Lizzeri A (1999) Interfering with secondary markets. RAND Journal of Economics 30(1):1-21.

Hendriks A (2012) Sophie-software platform for human interaction experiments. University of Osnabruck, Osnabruck, Germany.

Ho B, Taber J, Poe G, Bento A (2016) The effects of moral licensing and moral cleansing in contingent valuation and laboratory experiments on the demand to reduce externalities. *Environmental and Resource Economics* 64(2):317-340.

Ho TH, Lim N, Cui TH (2010) Reference dependence in multilocation newsvendor models: A structural analysis. *Management Science* 56(11):1891-1910.

Ho TH, Zhang J (2008) Designing pricing contracts for boundedly rational customers: Does the framing of the fixed fee matter? *Management Science* 54(4):686-700.

Hossain T, List JA (2012) The behavioralist visits the factory: Increasing productivity using simple framing manipulations. *Management Science* 58(12):2151-2167.

Houde S (2014) How consumers respond to environmental certification and the value of energy information. *National Bureau of Economic Research,* Paper No. 20019, March 2014.

Hoyer WD, H A, Huber F (2002) When buyers also sell: The implications of pricing policies for customer satisfaction. *Psychology & Marketing* 19(4):329-355.

Huang L, Song JSJ, Swinney R (2017) Managing social responsibility in multitier supply chains. Working paper, Duke University, Durham, NC.

ILO (2017) Global estimates of child labour: Results and trends 2012-2016. *International Labour Organization* (September 19), http://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/documents/publication/wcms_575499.pdf.

Kalafatis SP, Pollard M, East R, Tsogas MH (1999) Green marketing and Ajzen's theory of planned behaviour: a cross-market examination. *Journal of Consumer Marketing* 16(5):441-460.

Kalkanci B (2017) Supply risk mitigation via supplier diversification and improvement: An experimental evaluation. Working Paper, Georgia Tech, Atlanta, GA.

Kallbekken S, Sælen H, Hermansen. EA (2013) Bridging the energy efficiency gap: A field experiment on lifetime energy costs and household appliances. *Journal of Consumer Policy* 36(1):1-16.

Kim J, Rao RS, Kim K, Rao AR (2011) More or less: A model and empirical evidence on preferences for under- and overpayment in trade-in transactions. *Journal of Marketing Research* 48(1):157-171.

Kippenberg J (2018) The path to responsible sourcing in the jewellery sector. *hrw.org* (March 21) https : // www.hrw.org / news / 2018 / 03 / 21 / path – responsible – sourcing – jewellery - sector/.

Kohler C (2016) Soon you'll buy Consoles the way you upgrade iPhones. *Wired* (May 26), http://www.wired.com/2016/05/future-of-game-consoles/.

Kornish LJ (2001) Pricing for a durable-goods monopolist under rapid sequential innovation. *Management Science* 47(11):1552-1561.

Kumar A, Lim SS (2008) How do decision frames influence the stock investment choices of individual investors? *Management science* 54(6):1052-1064.

Lee HL, Tang CS (2018) Socially and environmentally responsible value chain innovations: New operations management research opportunities. *Management Science* 64(3):983-996.

Lee K, 2008) Opportunities for green marketing: young consumers. *Marketing Intelligence & Planning* 26(6):573-586.

Lee YS, Seo YW, Siemsen E (2018) Running Behavioral Operations Experiments Using Amazon's Mechanical Turk. Production and Operations Management 27(5):973-989.

Lehner JM (2000) Shifts of reference points for framing of strategic decisions and changing risk-return associations. *Management Science* 46(1):63-76.

Leonard TC (2008) Richard H. Thaler, Cass R. Sunstein, Nudge: Improving decisions about health, wealth, and happiness. *Constitutional Political Economy* 19(4):356-360.

Levav J, Reinholtz N, Lin C (2012) The effect of ordering decisions by choice-set size on consumer search. *Journal of Consumer Research* 39(3):585-599.

Levinthal DA, Purohit D (1989) Durable goods and product obsolescence. Marketing Science 8(1):35-56.

Lewandowski M (2016) Designing the business models for circular economy—Towards the conceptual framework. *Sustainability* 8(1):43-71.

Li KJ, Fong DK, Xu SH (2011) Managing trade-in programs based on product characteristics and customer heterogeneity in business-to-business markets. *Manufacturing & Service Operations Management* 13(1):108-123.

Liebowitz SJ (1985) Copying and indirect appropriability: Photocopying of journals. Journal of Political Economy 93(5):945-957.

McKenzie-Mohr D (2011) Fostering sustainable behavior: An introduction to community-based social marketing (3rd edition). *New Society Publisher*, Gabriola, BC, Canada.

Merritt AC, Effron DA, Monin B (2010) Moral self-licensing: When being good frees us to be bad. *Social and Personality Psychology Compass* 4(5):344-357.

Monga A, Zhu RJ (2005) Buyers versus sellers: How they differ in their responses to framed outcomes. *Journal of Consumer Psychology* 15(4):325-333.

Moorthy KS (1984) Market segmentation, self-selection, and product line design. *Marketing Science* 3(4):288-307.

Moorthy KS, Png IP (1992) Market segmentation, cannibalization, and the timing of product introductions. *Management Science* 38(3):345-359.

Newell RG, Siikamäki J (2014) Nudging energy efficiency behavior: The role of information labels. *Journal of the Association of Environmental and Resource Economists* 1(4):555-598.

Nisen M (2013) How Nike solved its sweatshop problem. *BusinessInsider.com* (May 9), http://www.businessinsider.com/how-nike-solved-its-sweatshop-problem-2013-5.

Okada EM (2001) Trade-ins, mental accounting, and product replacement decisions. *Journal of Consumer Research* 27(4):433-446.

Okada EM (2006) Upgrades and new purchases. *Journal of Marketing* 70(4):92-102.

Oraiopoulos N, Ferguson ME, Toktay LB (2012) Relicensing as a secondary market strategy. *Management Science* 58(5):1022-1037.

Osbaldiston R, Schott JP (2012) Environmental sustainability and behavioral science: Meta-analysis of proenvironmental behavior experiments. *Environment and Behavior* 44(2):257-299.

Ottman JA, Stafford ER, Hartman CL (2006) Avoiding green marketing myopia: Ways to improve consumer appeal for environmentally preferable products. *Environment: Science and Policy for Sustainable Development* 48(5):22-36.

Plambeck EL, Taylor TA (2015) Supplier evasion of a buyer's audit: Implications for motivating supplier social and environmental responsibility. *Manufacturing & Service Operations Management* 18(2):184-197.

Porteous AH, Rammohan SV, Lee HL (2015) Carrots or sticks? Improving social and environmental compliance at suppliers through incentives and penalties. *Production and Operations Management* 24(9):1402-1413.

Purohit D (1995) Playing the role of buyer and seller: The mental accounting of trade-ins. *Marketing Letters* 6(2):101-110.

Rabin M (2013a) An approach to incorporating psychology into economics. *The American Economic Review* 103(3):617-622.

Rabin M (2013a) An approach to incorporating psychology into economics. *American Economic Review* 103(3):617-22.

Rabin M (2013b) Incorporating limited rationality into economics. *Journal of Economic Literature* 51(2):528-43.

Rao RS, Narasimhan O, John G (2009) Understanding the role of trade-ins in durable goods markets: Theory and evidence. *Marketing Science* 28(5):950-967.

Ray S, Boyaci T, Aras N (2005) Optimal prices and trade-in rebates for durable, remanufacturable products. *Manufacturing & Service Operations Management* 7(3):208-228.

Read D, Loewenstein G (1995) Diversification bias: Explaining the discrepancy in variety seeking between combined and separated choices. *Journal of Experimental Psychology: Applied* 1(1):34-49.

Rice G, Wongtada N, Leelakulthanit O (1996) An investigation of self-efficacy and environmentally concerned behavior of Thai consumers. *Journal of International Consumer Marketing* 9(2):1-19.

Roosevelt M (2011) Pressured by Greenpeace, Mattel cuts off sub-supplier APP. *Los Angeles Times* (June 10), http://articles.latimes.com/2011/jun/10/business/la-fi-mattel-greenpeace-20110610.

Rust J (1986) When is it optimal to kill off the market for used durable goods? *Econometrica* 54(1):65-86.

Samson A, Voyer BG (2014) Emergency Purchasing Situations: Implications for Consumer Decision Making. *Journal of Economic Psychology* 44(1):21-33.

Schubert C (2017) Green nudges: Do they work? Are they ethical? *Ecological Economics* 132:329-342.

Schweitzer ME, Cachon GP (2000) Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Science* 46(3):404-420.

Securis (2017) Let's Talk About America's Electronic Waste Problem. *securis.com* (March 2017), http://www.securis.com/lets-talk-about-americas-electronic-waste-problem/.

Seo S (2006) A review and comparison of methods for detecting outliers in univariate data sets. Ph.D. thesis, University of Pittsburgh.

Shogren JF (2002) Micromotives in global environmental policy. *Interfaces* 32(5):47-61.

Shrum, L.J., Lowrey, T.M., J.A. McCarty. 1994. Recycling as a marketing problem: a framework for strategy development. *Psychology & Marketing* 11(4):393-416.

Simonson I (1990) The effect of purchase quantity and timing on variety-seeking behavior. *Journal of Marketing Research* 150-162.

Simonson I, Drolet A (2004) Anchoring effects on consumers' willingness-to-pay and willingness-to-accept. *Journal of Consumer Research* 31(3):681-690.

Souza GC (2013) Closed-loop supply chains: a critical review, and future research. Decision Sciences 44(1):7-38.

Speer M (2012) What is a Green Product? *isustainableearth.com* (December 2011), http://www.isustainableearth.com/green-products/what-is-a-green-product.

Srivastava J, Chakravarti D (2011) Price presentation effects in purchases involving trade-ins. Journal of *Marketing Research* 48(5):910-919.

Statisticbrain (2016) iPhone 5 sales statistics. *statisticbrain.com* (September 26), http://www.statisticbrain.com/iphone-5-sales-statistics/.

Sunstein CR (2009) Green Nudges: An Interview with Obama Regulatory Czar Cass Sunstein. *grist.org* (April 7), http://grist.org/article/2009-green-nudges-aninterview-with-obama-re/.

Tang CS, Zhou S (2012) Research advances in environmentally and socially sustainable operations. *European Journal of Operational Research* 223(3):585-594.

Thaler R (1980) Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization* 1(1):39-60.

Thaler R (1985) Mental accounting and consumer choice. *Marketing Science* 4(3):199-214.

Thaler RH, Sunstein CR (2008) Nudge: Improving Decisions about Health. Yale University Press, New Haven, CT.

Time (2014) Apple's $649 iPhone 6 reportedly costs $200 to make. *time.com* (September 24), http://time.com/3426087/apple-iphone-6-cost/.

Time (2016) The iPhone 7 costs you 3x more than it costs Apple to make. *time.com* (September 26), http://time.com/money/4508221/iphone-7-mark-up-prot-margins/.

Tomlin B (2006) On the value of mitigation and contingency strategies for managing supply chain disruption risks. *Management Science* 52(5):639-657.

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

Tversky A, Kahneman D (1981) The framing of decisions and the psychology of choice. *Science* 211(4481):453-458.

Tversky A, Kahneman D (1986) Rational choice and the framing of decisions. *Journal of business* S251-S278.

Tversky A, Kahneman D (1991) Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics* 106(4):1039-1061.

Ungemach C, Camilleri AR, Johnson EJ, Larrick RP, Weber EU (2018) Translated Attributes as Choice Architecture: Aligning Objectives and Choices Through Decision Signposts. *Management Science*, 64(5):2445–2459.

US EPA (2016) Advancing sustainable materials management: 2014 fact sheet. *U.S. Environmental Protection Agency, Office of Resource Conservation and Recovery*, http://www.epa.gov/sites/production/files/2016-11/documents/2014_smmfactsheet_508.pdf.

Van Ackere A, Reyniers DJ (1993) A rationale for trade-ins. *Journal of Economics and Business* 45(1):1{16.

Van Ackere A, Reyniers DJ (1995) Trade-ins and introductory offers in a monopoly. *RAND Journal of Economics* 26(1):58-74.

Vega N (2017) Apple has been using teen labor to assemble the iPhone X. *New York Post* (November 21), https://nypost.com/2017/11/21/apple-has-been-using-teen-labor-to-assemble-the-iphone-x/.

Waldman M (1996) Planned obsolescence and the R&D decision. *RAND Journal of Economics* 27(3):583-595.

Waldman M (2003) Durable goods theory for real world markets. *Journal of Economic Perspectives* 17(1):131-154.

Wang Y, Gilland W, Tomlin B (2010) Mitigating supply risk: Dual sourcing or process improvement? *Manufacturing & Service Operations Management* 12(3):489-510.

Ward DO, Clark CD, Jensen KL, Yen ST, Russell CS (2011) Factors influencing willingness-to-pay for the ENERGY STAR® label. *Energy Policy* 39(3):1450-1458.

Westley F, Vredenburg H (1991) Strategic bridging: The collaboration between environmentalists and business in the marketing of green products. *The Journal of applied behavioral science* 27(1):65-90.

Yang Y, Vosgerau J, Loewenstein G (2013) Framing influences willingness to pay but not willingness to accept. *Journal of Marketing Research* 50(6):725-738.

Yardley J (2013) Report on Deadly Factory Collapse in Bangladesh Finds Widespread Blame. *New York Times* (May 22), https : // www . nytimes . com / 2013 / 05 / 23 / world / asia / report-on-bangladesh-building-collapse-finds-widespread-blame.html.

Zhao H, Jagpal S (2009) Upgrade pricing, market growth, and social welfare. *Journal of Business Research* 62(7):713-718.

Zhen C, Wohlgenant MK, Karns S, Kaufman P (2011) Habit formation and demand for sugar-sweetened beverages. American Journal of Agricultural Economics 93(1):175-193.

Zhu R, Chen X, Dasgupta S (2008) Can trade-ins hurt you? exploring the effect of a trade-in on consumers' willingness to pay for a new product. *Journal of Marketing Research* 45(2):159-170.

Zink T, Geyer R (2017) Circular economy rebound. *Journal of Industrial Ecology* 21(3):593-602.

Zizzo DJ (2010) Experimenter demand effects in economic experiments. *Experimental Economics* 13(1):75-98.