

CONSUMER DEMAND FOR THE FAIR TRADE LABEL: EVIDENCE FROM A MULTI-STORE FIELD EXPERIMENT

Jens Hainmueller – Stanford University
Michael J. Hiscox – Harvard University
Sandra Sequeira – London School of Economics

February 2014

JEL Classification Numbers: C93, D12, D64, M14.

Word Count: 14,490

Jens Hainmueller, Stanford University, Department of Political Science and Graduate School of Business, 616 Serra Street Encina Hall West, Stanford, CA 94305-6044. E-mail: jhain@stanford.edu. Michael J. Hiscox, Harvard University, Department of Government, 1737 Cambridge Street, Cambridge, MA 02138. E-mail: hiscox@fas.harvard.edu. Sandra Sequeira, London School of Economics, Houghton Street London WC2A 2AE, United Kingdom. E-mail: S.Sequeira@lse.ac.uk: corresponding author.

ABSTRACT:

We provide new evidence on consumer demand for ethical products from experiments conducted in a U.S. grocery store chain. We find that sales of the two most popular coffees rose by almost 10% when they carried a Fair Trade label as compared to a generic placebo label. Demand for the higher priced coffee remained steady when its price was raised by 8%, but demand for the lower priced coffee was elastic: a 9% price increase led to a 30% decline in sales. While consumers attach value to ethical sourcing, there is significant heterogeneity in willingness to pay for it.

I. INTRODUCTION

Ethical product labels and marketing messages are an increasingly common sight in retail settings, calling attention to particular aspects of the way goods have been made (e.g. labor practices, environmental standards, the treatment of animals), and to particular causes that stand to benefit when the goods are purchased (e.g. research on HIV/AIDs or the provision of clean drinking water). The Fair Trade label, which aims to guarantee a “better deal” for poor farmers in developing countries, is perhaps the most well-known ethical label. Fair Trade coffee, tea, and chocolate are now marketed not just on college campuses and in fashionable cafes, but also in many major supermarket chains across the United States and in Europe (including Walmart, Target, Safeway, Giant, Tesco and Sainsbury’s, among others), and global sales of Fair Trade products have risen by around 30% annually over the past decade (FLO 2012).

This is a new form of politicized consumption in which citizen-consumers vote with their shopping dollar to influence firm behavior and bring about political and social change. Its potential long-term impact, in terms of the size of the market and the associated effects on firm behavior, is difficult to assess. Skeptics dismiss Fair Trade and other ethically labeled products as cheap public relations ploys by companies, and highlight the fact that such products currently account for a tiny share of retail sales. Supporters argue that if it continues to grow at the current rate, politicized consumption could have a large impact on firm behavior. Much attention has been devoted to survey evidence showing that a majority of consumers say they would prefer, and would be willing to pay extra for, any products they could identify as being made in ethical ways (Elliott and Freeman 2003). As yet, however, there is no clear evidence that consumers will actually behave this way when they are shopping, thus giving firms strong incentives to change their behavior and invest in ethical labeling (Devinney, Auger and Eckhart 2010).

This paper reports new evidence on the impact of ethical labels on consumers' willingness to pay from a field experiment conducted among actual consumers in 26 stores of a major U.S. grocery store chain. The tests reveal that the Fair Trade label has a substantial positive effect on sales. Sales of the two most popular bulk coffees sold in the stores rose by almost 10% when the coffees carried a Fair Trade label as compared to a generic placebo label. Yet consumers also reveal different levels of price sensitivity when informed of the ethical product attribute. Demand for the higher priced coffee was less elastic: sales of the Fair Trade labeled coffee remained fairly steady when its price was raised by 8%. Demand for the lower priced coffee was more elastic: a 9% increase in its price led to a 30% decline in sales as buyers switched to low-priced unlabelled alternatives. Overall, the findings suggest that consumers value ethical labeling as an important product attribute in the absence of any price differential relative to similar unlabelled products. However, in the presence of a price premium, we observe significant heterogeneity in the weight different consumers place on ethical sourcing when making their purchasing decisions. Such behavioral responses to ethical labels might be driven by several factors including differences in social preferences or in levels of information about the importance of ethical sourcing.

This study makes several contributions. First, our results have implications for an extensive literature in industrial organization and applied microeconomics that attempts to understand consumer behavior, how firms respond to consumer preferences, and how this interaction affects firm profits, market structure, and consumer welfare (Spence 1976; Carlton 1978; Hausman, Leonard, and Zona 1994; Berry, Levinsohn, and Pakes 1995; Nevo 2010). The proliferation of ethical branding is based on the assumption that this is an effective means of product differentiation given altruistic consumers. Our results provide evidence of consumer heterogeneity in the valuation of the Fair Trade label, suggesting that firm-level marketing strategies can be designed to optimally account for market segmentation based upon the complex interaction between price, ethical labels, and other

product attributes. Second, to the best of our knowledge, this is the first paper to report results from a field experiment in which the researchers simultaneously manipulate product attributes like prices and labels to estimate demand effects across multiple retail stores.¹ Previous empirical research in the industrial organization literature has relied almost exclusively upon estimating models of demand using observational data with a variety of techniques (and restrictions) applied to account for the endogeneity of pricing and marketing. Our tests highlight the advantages from the field experimental approach applied to a multi-store setting. Third, our findings add new empirical evidence to complement a growing theoretical literature on the extent and implications of social preferences (Fehr and Schmidt 1999; Andreoni 2006; Benabou and Tirole 2006).

II. FAIR TRADE AND CONSUMER DEMAND FOR ETHICALLY CERTIFIED PRODUCTS

The Fair Trade certification and labeling program was developed by a group of humanitarian organizations aiming to alleviate poverty and promote sustainable development in developing countries by establishing more direct relationships between producers in those countries and sympathetic consumers in developed economies. Fair Trade certified farmers receive a guaranteed minimum price for their crops and a price premium above the minimum or the current market price for the commodity, whichever is higher.² In addition, Fair Trade certified importers must agree to long-term (minimum of one year) contracts with farmers and make available pre-harvest credit (up to 60% of the contract value). Fair Trade certification prohibits forced and child labor on farms along with ethnic and other forms

¹While Hilger et al (2011) conducts an experiment that manipulates the labels of different types of wines in a retail setting, there is no price variation associated with the label.

²For example, the minimum price for coffee (Arabica, unwashed) is currently \$1.35 per pound and the premium over the current market price is 20 cents per pound.

of discrimination, and restricts the use of potentially hazardous chemicals. Certification is generally restricted to small, family-owned farms and requires that farmers organize into cooperatives that decide democratically how to distribute or invest the fair trade premium paid on each contract.³

As with other types of third-party certification and labeling, the Fair Trade program can be seen as a way to remove a market inefficiency that exists due to incomplete information on the part of consumers about the manner in which goods are produced. Removing this information asymmetry can facilitate product differentiation that increases consumer welfare by introducing additional product variety (Elliott and Freeman 2003; Becchetti and Solferino 2005) and enabling the fulfilment of social preferences (Camerer 2002; Sobel 2002). In the simplest type of models, lack of information about the ethical quality of goods available to consumers can lead to welfare losses, as consumers who prefer goods with high

³The program is administered by a collection of non-profit Fairtrade Labelling Organizations (FLO) that oversees certification and licenses the use of the Fair Trade trademark in each national market (in the United States, certification and licensing is organized by Fair Trade USA, formerly known as TransfairUSA and a member of FLO until 2011 when it began operating independently). FLO has developed standards for production and trade for a range of agricultural products, including coffee, tea, cocoa, bananas, sugar, rice, and cotton (see http://www.fairtrade.net/generic_standards.html). FLO conducts inspections of producers in developing countries, examines contracts, and monitors the chain of custody by which the certified goods are supplied to traders and retailers who are licensed to use the Fair Trade label and logo only when all the standards have been met. In 2012 the program included over 1.3 million farmers in 70 nations in Africa, Asia, and Latin America, with annual global sales of certified products exceeding \$6.6 billion in 2011. FLO estimates that approximately \$103 million in premium payments was distributed to communities in 2012 for use in community development (FLO 2012).

ethical quality cannot identify (and thus adequately reward) high-quality producers, and the latter are driven from the market by low-quality producers who face lower costs (Bonroy and Constantatos 2003, 2008).

To a large degree, the success of the Fair Trade model hinges on the depth and strength of support for ethically labeled goods among consumers.⁴ At present there is a great deal of uncertainty about whether the Fair Trade market can become large enough to have a substantial impact across a range of producers in the developing world. Total sales of Fair Trade goods in the United States in 2011 amounted to roughly \$1.4 billion (FLO 2012). This represents only about one fortieth of the U.S. market for certified organic products and less than \$5 per person annually. But the average annual rate of growth in U.S. sales of Fair Trade certified goods was close to 40% between 1999 and 2008. By way of comparison, U.S. sales of certified organic products grew by around 20% annually between 1990, when certification began, and 2002 (Dimitri and Green 2002). Fair Trade coffee, the largest selling certified product, accounts for over 3% of the total retail market for coffee and for close to 20% of the market for specialty coffees, the fastest growing segment of the US coffee market (TransFair USA 2009a; 2009b).⁵

⁴A second necessary condition for the sustainability of this model is that producers in the developing world actually benefit from participating in the fair trade system. Research to date has provided only crude assessments of the impact of Fair Trade certification among developing-country producers in the form of case studies of certified farmers that do not provide general measures of impact (Ronchi 2002; Murray et al. 2005) and surveys of certified and non-certified producers that do not account for the non-random selection of farmers into certification (Arnould, Plastina, and Ball 2006; Becchetti and Constantino 2008; Bacon et al. 2008).

⁵Fair Trade coffee is available in major coffee and food retailers, such as Starbucks Coffee, Peet's Coffee and Tea, Seattle's Best Coffee, Einstein Bros Bagels, Dunkin' Donuts,

Survey data suggest that a majority of consumers prefer, and are willing to pay substantially more for, products they can identify as being made in an ethical way.⁶ Several of these studies have focused specifically on Fair Trade coffee and report that consumers are willing to pay a sizeable premium for Fair Trade certification (De Pelsmacker et al. 2005). In a recent study, Hertel et al. (2009) found that over 75% of surveyed coffee buyers in the U.S. in 2006 said they would be willing to pay at least 50 cents more per pound for Fair Trade coffee versus non-certified coffee (a premium of roughly 16% over the average price of coffee at the time) and more than half said they would pay a premium of a dollar or more. But survey findings most likely reflect some degree of social desirability bias. What is required is direct evidence on how consumers actually behave when they encounter Fair Trade labels while shopping and deciding how to spend their own money.

A small set of empirical studies have examined relationships between observed sales and/or prices of goods and their ethical characteristics. For instance, Teisl, Roe, and Hicks (2002) examined scanner data on U.S. retail sales of canned tuna and found that market share (relative to other canned seafood and meat) rose substantially after the introduction and McDonald's, as well as in many large supermarket chains, including Walmart, Target, Safeway, Giant, Costco, Trader Joe's and Whole Foods Market.

⁶For example, a survey administered in 1999 by the Program on International Policy Attitudes found that 76% of respondents indicated they were willing to pay \$25 for a \$20 garment that was certified as not being made in a sweatshop (PIPA 2000). A poll conducted in the same year by the National Bureau of Economic Research found that roughly 80% of surveyed individuals said they were willing to pay more for an item if assured it was made under good working conditions (Elliott and Freeman 2003). A growing number of survey studies have provided additional evidence of consumers' stated willingness to pay for ethical qualities of products and for the ethical behavior of firms (Auger et al. 2003, 2008; Dickson 2001; Mohr and Webb 2005).

of the “dolphin-safe” label in April 1990. Elfenbein and McManus (2010) found a price premium for items sold in eBay’s “Giving Works” program (in which sellers direct a portion of the sale price to charity) compared with prices for similar items sold on eBay, and the premium was increasing in the amount donated to charity. On the Fair Trade label and coffee, specifically, Galarraga and Markandya (2004) gathered data on retail prices of coffee sold in major supermarkets in Britain and estimated that an average premium of around 11% was charged for coffee with a “green” label (they combined Fair Trade, organic, and shade-grown labels in this category). While such studies are suggestive of consumer support for ethically labeled products, because the observed outcomes reflect pricing and distribution decisions by sellers as well as consumer behavior, it is difficult for this type of approach to provide clear inferences about the real impact of ethical labels on consumer choices.

To date, very limited evidence is available from field experiments indicating whether and how consumers might alter their spending behavior when given the opportunity to distinguish Fair Trade or other ethically labeled products from alternatives. Kimeldorf et al. (2004) placed two identical groups of athletic socks in a department store, labeled one group as being made under “Good Working Conditions”, and altered the price of the labeled socks over several months. Hiscox and Smyth (2006) introduced a “Fair and Square” label describing ethical labor standards in facilities that manufactured brands of towels and candles sold in a retail store in New York City, then compared sales of labeled brands with sales of alternative brands. Arnot, Boxall, and Cash (2006) conducted tests with a university campus coffee vendor, adjusting the prices of two fresh-brewed coffees, a Fair Trade certified coffee from Nicaragua and a similar quality Colombian coffee, over the course of several days. In each of these field experiments, weaknesses in design made it impossible for the researchers to isolate the effects of the ethical labels from potential time-varying or product specific confounding factors, or to compare the effects of ethical

product labels with the effects of alternative types of marketing labels.⁷ The experiments we report below were designed specifically to overcome these problems and to gather new, direct evidence on how shoppers behave when encountering Fair Trade labels, and making real spending decisions in a multi-store retail setting.

III. RESEARCH DESIGN

A. MODEL OF CONSUMER BEHAVIOR

We employ a standard model of consumer behavior in which individuals may derive utility from a variety of characteristics of goods (Lancaster 1971; Gorman 1980). We assume consumers maximize their utility when choosing from a set of alternative products (e.g. types of coffee) available in a particular market. Each consumer’s utility from buying a particular good depends on the observed product characteristics, which may include Fair Trade certification, as well as price.

Consumers may differ in how they evaluate the different product characteristics. Our tests are designed to measure average responses among consumers when certain key product characteristics - Fair Trade certification and price - are manipulated experimentally for specific products. We allow consumers to place different values on Fair Trade certification, and to be more, or less, sensitive to prices charged for Fair Trade goods than they are to prices of unlabelled goods. We do not make specific assumptions about the motives of these consumers. The simplest type of assumption is that these consumers derive a “warm

⁷Comparing sales for products with and without the Fair trade label cannot rule out the possibility of a pure label effect (i.e., one that is irrespective of the informational content of the label). The ideal design would compare the product with the Fair Trade label versus a non-informative placebo label.

glow” satisfaction from supporting a program that is helping poor coffee farmers - this type of assumption is adopted in existing models of markets for ethically labeled goods (Richardson and Stahler 2007; Baron 2009). There are other motives that could generate a preference for purchasing ethically labeled products, some of them much less altruistic than others, however, our tests are not designed to assess the relative importance of alternative motivations among consumers favoring ethically labeled goods.

In general, the standards under which a good is made can be classified as “credence” attributes and are distinct from other types of product characteristics in that they cannot be directly assessed by the consumer examining or using the item. Other product characteristics, such as price, size, and color, can be evaluated by consumers before they purchase the good - these are sometimes called “search” attributes. Characteristics such as quality, durability, and taste, can be assessed by consumers after they have purchased the good and are known as “experience” attributes.⁸ Although these experience attributes are often not known to consumers at the point of purchase, firms can use a variety of methods to send credible signals about them, including guarantees, warranties, advertising and investments in brand reputations. The information asymmetry problem for experience attributes is also partly alleviated by the fact that consumers can punish firms for poor quality by making no further purchases of their products (Akerlof 1970; Shapiro 1983; Palfrey and Romer 1983). In the case of credence attributes, however, which are never directly observed by consumers

⁸For more detailed discussions of these different types of attributes, see Nelson (1970, 1974), Darby and Karni (1973), and Roe and Sheldon (2007). Besides Fair Trade standards for farmers, other familiar examples of credence attributes include organic standards for production of food and fiber, exclusion of genetically modified organisms from foods, dolphin safe methods for catching tuna, humane treatment of animals on farms, and various forms of environmental management standards adopted by firms to help to sustain forests and fisheries.

before or after purchasing the product, firms find it much harder to make credible assurances. Firms that have incurred higher costs to produce goods with these characteristics can make claims about them to consumers, but competing firms can incur no additional costs and make similar claims. Certification and labeling of specific credence attributes of goods (e.g., Fair Trade standards) by an independent third party (e.g., FLO), can mitigate this problem, effectively transforming the credence attributes into search attributes (Caswell and Mojdzuska 1996).⁹

B. THE SETTING

We investigated consumer demand for the Fair Trade label by conducting two experiments in 26 stores of a major US grocery store chain, located in the states of Connecticut, Massachusetts, Maine, and Rhode Island. The experiments took place in 2008 and 2009. The first test, the Label experiment, examined the impact of the Fair Trade label on sales of goods at existing prices. The second test, the Price experiment, investigated the price elasticity of demand for Fair Trade labeled goods. The experiments focused on the two biggest selling Fair Trade coffees sold in the stores, the French Roast (FR) Regular and a Coffee Blend (CB). Consumers could purchase coffee in the stores either from self-service bulk bins containing roasted coffee beans or in a separate section of shelves of packaged (whole and ground) coffee beans and instant coffees. All bulk coffee was supplied by the same company and during our experiments our test coffees – FR Regular and CB – were the only Fair Trade bulk coffees available. Sales of bulk coffee beans were about twice as

⁹The value of the Fair Trade label to firms and consumers will depend in part on the degree to which consumers regard the particular third party certifier as trustworthy. It is worth noting that our tests were not designed to assess the importance of third-party certification per se or the trustworthiness of FLO in the eyes of consumers (relative to the trustworthiness of the grocery store partner).

large as sales of packaged coffee beans in the stores and they accounted for over half the total coffee market in the stores (including sales of instant coffees).

C. THE LABEL EXPERIMENT

In the Label experiment the intervention consisted of attaching a 2x2 inch Fair Trade label to the bulk coffee bins containing the FR Regular and CB coffees in all stores assigned to the treatment condition. In stores assigned to the control condition, we attached a 2x2 inch generic placebo label to the bins containing these same coffees. The generic label was designed to be identical to the Fair Trade label in all the relevant dimensions from a marketing point of view such as the size of the label and its color. The only difference was in the meaning of the label: the treatment label indicated the Fair Trade sourcing of the product while the control label carried no specific information about Fair Trade and simply highlighted the name of the brand. We used the generic label for the control condition to allow for a generic label effect, unrelated to the specific informational content associated with Fair Trade, as past research has suggested that even seemingly meaningless forms of differentiation in marketing messages can affect consumer choices (Carpenter, Glazer, and Nakamoto 1994).¹⁰ Figure 1 shows the treatment and control labels that were displayed on the coffee bins. In the control label, we replaced the word Coffee with the coffee supplier's brand name. The store sourced exclusively from this supplier and this information was already available to consumers (the brand name was included on the standard display card on each bin that gave the price and the detailed description of each coffee type). Each coffee bin displayed the experimental label (treatment or control), the standard display card with the price and the description of the coffee type, and a sticker that indicated the

¹⁰Mullainathan, Schwartzstein, and Shleifer (2008) suggest that such effects may in part be due to advertisers encouraging consumers to transfer positive assessments of seemingly irrelevant attributes by analogy or association.

time of the last roasting.

For the duration of the Label experiment, we removed all other references to Fair Trade from the product descriptions of both Fair Trade bulk coffees so that the 2x2 bin label was the only reference to Fair Trade in the bulk coffee sections of the stores. Prior to our experiment, the FR Regular coffee bins had displayed a small (half inch square), black-and-white Fair Trade logo beneath the coffee description on the standard display card, which was almost indiscernible to the casual shopper. These small logos were removed in all stores several weeks before the start of the first test. Packages of the FR Regular coffee, sold in the separate section of the store, carried a small Fair Trade logo on their reverse side before and during the experiment. It is not impossible that some perceptive repeat consumers did not react to the bin labels because they identified the FR Regular coffee as Fair Trade certified in the control condition of our label test if they recalled the previous label being on the bulk bins in the past or if they closely examined packages of the FR Regular coffee in the packaged coffee section of the store before shopping in the bulk section. As a consequence, the results we report below can be interpreted as a lower bound of the true effect of the Fair Trade label

D. THE PRICE EXPERIMENT

In the Price experiment the intervention consisted of raising the prices for the Fair Trade labeled FR Regular and CB coffees. In the treatment condition, prices were raised by \$1 for both types of coffee.¹¹ Given the base price of \$11.99 per pound for the FR Regular, and

¹¹The magnitude of the price increase was decided jointly with the grocery store partner. Our goal was to identify a salient price increase (one dollar per pound) while ensuring that the final price of the FR Regular and CB test coffees would not be significantly outside of the normal price range for bulk coffee in the stores. This satisfied our grocery store partner's participation constraint.

\$10.99 for the CB coffee, this represents price increases of about 8% and 9% respectively.¹² In the control condition, prices remained at their usual levels. Notice that changing the prices changed the price ranking of the test coffees among the set of bulk coffees available at the stores. Almost all the other bulk coffees were sold at \$11.99 per pound, so in the treatment condition FR Regular moved from being an average-priced coffee to being among one of the most expensive coffees, at \$12.99. Only two other specialty bulk coffees were sold at this higher price and these accounted for a lower sales volume than the FR Regular (see the summary statistics in Section IV for details of the consumers' choice set). The CB coffee was one of only two bulk coffees usually sold at \$10.99 (the other was Colombian Supremo). So during the treatment period, this coffee moved from being one of the cheapest bulk coffees on offer to being an average-priced coffee, at \$11.99. As we discuss below, this had potentially important implications in terms of substitution effects.

In addition to the price increase, the stores in the treatment condition displayed a prominent 3x3 inch Fair Trade label on the bulk bins containing the FR Regular and CB coffees that carried a message aimed at inducing consumers to connect the higher price specifically with Fair Trade certification. The label read: "A Fair Price to Support Fair Trade." Stores in the control condition, where prices were not altered, displayed a Fair Trade label with the message: "Support Fair Trade." The two labels are shown in Figure 2. By explicitly directing consumers to associate the price premium with Fair Trade certification in the treatment condition, the test provides an assessment of their willingness to pay extra for this specific ethical product attribute. In the absence of such a message, it is possible that some customers would associate higher prices with some other type of unobserved product characteristic - an experience attribute, such as quality or flavor - thus

¹²Prices of the packaged versions (ground and whole bean) of the FR Regular and CB coffees were raised by the same amount, so that particularly savvy customers could not avoid paying the higher price by switching to the pre-packaged versions.

making it more difficult to interpret the results (Bagwell and Riordan 1991).¹³

To provide a benchmark for examining consumer responses to price changes in the absence of any message prompting them to associate these changes to ethical sourcing, we examined historical data on sales of all the bulk coffees in the stores at different prices during a two-year period prior to the tests. The historical sales data allow us to estimate the price elasticities of demand for all bulk coffees in the stores in the absence of the test labels. This helps us benchmark the results from the Price experiment.

E. CROSSOVER DESIGN

In both experiments we relied on a two-group, two-phase crossover design (Jones and Kenward 2003) whereby stores were randomly assigned to a sequence of treatment-control or control-treatment. In each store, the treatment or control condition was in place for an initial phase of four weeks, after which stores switched to the opposite condition for another four weeks. Thus, both experiments lasted eight weeks in total.¹⁴

The crossover design provides higher efficiency than a simple parallel group design, because we can exploit within-store variation for each store (assuming no carry-over). The no carry-over assumption may be violated if perceptive repeat customers remember that the test coffees are Fair Trade certified and therefore disregard the label changes during the experimental period (in particular in the stores in which the treatment labels are assigned

¹³There is evidence that price serves as a signal for unobserved quality even when consumers are actually able to assess quality directly via consumption of the good (Plassman et al. 2008).

¹⁴In the Price experiment, we extended the second phase to six weeks to accommodate the fact that in a small number of stores, the label switch was delayed for several days. Extending the second phase gives us the opportunity to discard the two weeks immediately following the label switch.

in the first phase and replaced by the control labels in the second phase). Presumably this should result in an attenuation bias for the label effect, since customers who value Fair Trade certification would simply continue to purchase the test coffees even under the control condition. In the discussion section we report various robustness checks which support the no carry-over assumption. In particular, we find that the effects are similar when (1) we only consider the first phase of the experiment (where no carry-over is present) and (2) when we replicate the crossover analysis while restricting the sample to sales during the last two weeks of each experimental phase when carry over-effects are less likely to occur. For the randomization, all 26 stores in our sample were initially matched into pairs on important covariates such as their history of average coffee sales, total sales, sales growth, and location characteristics. Within each pair, one store was then randomly assigned to the treatment-control and the other to the control-treatment condition, leading to a fully balanced design.

F. DATA AND MONITORING

To conduct the initial matching of stores, we combined store-level information on sales with socioeconomic data for the 5-digit zip code areas for each store drawn from the 2000 U.S. Census. To analyze the results of the experiments we relied on weekly register data on coffee sales in each store.

All stores received detailed instructions on how to attach the labels and change prices during the experiments. To ensure compliance with the experimental protocol of each experiment, we had our own monitors visit each of the participating stores during the first two days following the beginning of each treatment and control phase, and once a week after that. Observers checked the label displays, prices, and whether there were any product stock-outs that might affect sales. At no time during the experiments were the FR

Regular and CB coffees included in any promotional events or sales at the stores.¹⁵ Store managers and coffee department personnel at the stores were extensively briefed on the experiments. Overall, compliance was high: in only a few cases, the labels were switched a few days behind schedule.

G. RANDOMIZATION CHECKS

To verify whether the randomization successfully orthogonalized the treatment with respect to confounding factors, Tables 1 and 2 display the covariate balance for a range of pre-treatment characteristics. We report the mean covariate values in the treatment and in the control group as well as p-values from a two sample t-test (with unequal variances) and a bootstrapped two-sample Kolmogorov-Smirnov test (Abadie 2002). The pre-treatment characteristics include total store sales and total sales growth, as well as average dollar sales not only for each of the test coffees, but also for all bulk, all packaged coffee beans and all instant coffee available in the stores. The averages are provided both for a four week and for a 52 week period prior to the tests. The balance tables also include a range of socioeconomic characteristics for the 5 digit zip code areas in which the stores are located. For both experiments, we obtain very good balance on observed characteristics as variable means are close and none of the p-values indicate significant differences at conventional levels.

IV. ANALYSIS

¹⁵For the duration of the experiment, there was one type of bulk coffee that was placed on sale both in treatment and control stores. These sales promotions were routinely administered in all stores simultaneously nationwide and should therefore not lead us to reject our unconfoundedness assumption since it affected both treatment and control stores equally.

A. STATISTICAL MODEL

For the estimation we follow a standard framework in the discrete-choice literature (Akerberg et al. 2007; Nevo 2010). Let there be $i = 1, \dots, \infty$ consumers who maximize their utility by choosing one of $j = 0, 1, \dots, J$ goods (i.e. various bulk coffees and an outside good) in $t = 1, \dots, T$ markets. Markets are defined as store-weeks and for both experiments $n = 1, \dots, 26$ stores are observed over $w = 1, \dots, 8$ weeks; each store is observed for four weeks under the treatment and the control condition respectively. Consumer i 's utility from buying the j -th good in market t is given by

$$U_{ijt} = U(x_{jt}, \xi_{jt}, \nu_{it}; \theta) \tag{1}$$

where x_{jt} is a vector of observed product characteristics (which may include the price p_{jt}), ξ_{jt} indicates product characteristics that are unobserved by the researchers (these can also be thought of as demand shocks), ν_{it} are unobserved differences in consumer tastes, and θ is a vector of model parameters (to be estimated) that includes how sensitive consumers are to each of the observed product characteristics.¹⁶

For identification we normalize the utility of the outside good, $j = 0$, to zero and proceed with a simple logit specification where

$$U_{ijt} = \delta_{jt} + \varepsilon_{ijt} \text{ with mean utility levels } \delta_{jt} = x_{jt}\beta + \xi_{jt} \tag{2}$$

and the error term for idiosyncratic tastes is assumed to be $\varepsilon_{ijt} \stackrel{\text{iid}}{\sim}$ extreme value type II. Aggregate market shares are thus given by $s_{jt}(x, \beta, \xi) = \frac{\exp(x_{jt}\beta + \xi_{jt})}{\sum_{j=1}^J \exp(x_{jt}\beta + \xi_{jt})}$ and following

¹⁶The set of consumers choosing good j depends on the unobservable ν and is given by $S_j(\theta) = \{\nu | U_{ij} > U_{ik} \forall k\}$. Market shares can be recovered by specifying a distribution $f(\nu)$ and integrating over the values that meet the conditions in a given market $s_j(x|\theta) = \int_{\nu \in S_j(\theta)} f(\nu) d\nu$.

Berry (1994) we can solve for the mean utility as a function of observed market shares using $\delta_{jt} = \log(s_{jt}) - \log(s_{0t})$, and estimate the model by regression.

Our quantities of interest are the effects of the experimentally manipulated product characteristics (i.e., the Fair Trade label and the test price) on sales of the test coffees and on sales of the main alternative coffees that may be affected by substitution. We estimate the following model:

$$\log(s_{jt}) - \log(s_{0t}) = \alpha + \mathbf{M}\beta + \xi_{jn} + \xi_w + \Delta\xi_{jt} \quad (3)$$

where \mathbf{M} is a $(J \cdot T \times J)$ matrix that contains one indicator variable for each of the inside goods $\mathbf{M} = [m_{j=1}, \dots, m_{j=J}]$. For each inside good, the indicator variable is coded as one for store-weeks in which the treatment condition was assigned to the test coffees (i.e., the Fair Trade label or the test price) and zero for store-weeks in which the control condition was assigned to the test coffees (i.e. the control label or the regular price). Accordingly, $\beta = \{\beta_1, \dots, \beta_J\}$ is a $(J \times 1)$ vector of coefficients that measures the effect of the various product characteristics on product sales. The sales effect of the experimentally manipulated product characteristics are allowed to vary across the J coffees. The ξ_{jn} provide a full set of product/store fixed effects so that the identifying variation for the treatment effects is across time based on deviations from product/store specific means. We also include a set of week fixed effects, ξ_w , to account for weekly demand shocks that are common to all stores.¹⁷

The key identifying assumption, $E[\Delta\xi_{jt}|\mathbf{M}] = 0$, is supported given that the randomization orthogonalizes our treatments (i.e. the Fair Trade label and the price) with respect

¹⁷Notice that common shocks are also directly accounted for via the balanced experimental design (i.e. at each point in time half of the stores are assigned to treatment or control); the treatment effect coefficients are therefore unaffected by the inclusion of week fixed effects.

to all other observed or unobserved product characteristics of the test coffees, and of all the competitor coffees. Unlike in almost all other studies involving demand estimation, endogeneity of product characteristics or pricing is not a concern here. For the Label experiment we include the product prices p_{jt} as a regressor, although excluding the prices does not affect the point estimates of the label effect as expected given the randomization. For the Price experiment we omit prices because our treatment indicators measure the contrast between the test price and the control price. We use the coefficient estimates to compute own and cross-price elasticities. We cluster standard errors at the store level in order to allow for potential within-store correlation across time. For each experiment, we restrict the estimation window to the weeks when the experiment was underway.¹⁸

We include among the inside products the two test coffees, the FR Regular and CB coffees, as well as the five main alternative bulk coffees that were available across all stores: French Roast (FR) Extra Dark, Breakfast Blend, Regional Blend, Colombian Supremo, and Mexican.¹⁹ We compute market shares by converting volume sales to pounds and dividing by the total potential number of pounds of coffee in a given market. The potential coffee market is assumed to be equal to one cup of coffee per customer per day in a given store-week.²⁰

¹⁸As indicated above, for the Price experiment we discard two hybrid weeks following the switch of the two conditions so that we also have 8 weeks for each store; 4 weeks under each condition.

¹⁹Notice that we discard about 6% of the cases (ie. product-store-weeks) where sales are unavailable because of occasional stock outs and/or bulk bin rotations. The missing observations mostly involve the less popular coffees such as the Mexican and the Colombian. There are almost no missing observations for the two test coffees.

²⁰The International Coffee Organization estimates that in 2007 the average coffee consumption in the United States was .40 ounces per person per day, or roughly one cup (International Coffee Organization, 2008). The total number of customers per store-week

B. SUMMARY STATISTICS

Table 3 reports summary statistics for the test coffees and the alternative bulk coffees in 2009. The bulk coffees are all regularly priced at \$11.99 per pound with the exception of the CB and the Colombian Supremo, which are cheaper at \$10.99 per pound. The FR Regular coffee is the best-selling bulk coffee with an average sales share of about 11% among the bulk coffees and average weekly sales of about 17.2 pounds (or \$199) per store. The CB coffee has a sales share among the bulk coffees of about 7% and average weekly sales of about 9 pounds (or \$99) per store. The alternative bulk coffees, none of which are Fair Trade certified, all have weekly sales of about 11 pounds (or \$115 to \$126 dollars) per store, except for the Mexican coffee, for which sales are somewhat lower. It is worth noting that prior to our Price experiment, the store did not impose a price premium for Fair Trade certified coffee: the certified FR Regular and CB coffees were priced the same as similar non-certified coffees. This appears to be common practice among US coffee retailers.²¹

is based on total stores sales divided by the average basket size. Our approach here follows previous studies that similarly approximate market potential based on population and average consumption in the relevant markets (Berry et al. 1995; Nevo 2001).

²¹Reinstein and Song (2012) report that Fair Trade coffee is typically priced the same as comparable coffees in most of the largest US retailers such as Starbucks, Peet's Coffee and Tea, and Tully's. Presumably the companies are absorbing the additional costs themselves rather than passing it on directly to consumers, or relying on cross-subsidization. FLO (2010b) points out that the additional costs of Fair Trade certification for the final product can be so small (a small percentage of the farm gate price of the raw commodity, that is itself often only a small percentage of the total cost of the retail item after shipping, processing, packaging, and marketing) that it is possible for firms to absorb them entirely.

C. RESULTS

C.1. THE LABEL EXPERIMENT

Before presenting the results from the discrete choice model described above, it is useful to examine the impact of the FT label in a simple reduced-form specification. To do so we regress the log of weekly dollar sales of the test coffees — FR Regular and Coffee Blend — in each store on a binary treatment indicator equalling 1 if the test coffee displays the Fair Trade label and 0 if it displays the generic placebo label. We also add a full set of store and week fixed effects, so that the average treatment effect of the Fair Trade label is identified based on deviations from store/week means. Standard errors are clustered by store.

Table 4 presents the results from this reduced-form analysis. In column 1 we report that the Fair Trade label increased weekly sales of both FR and CB coffees by approximately 8% ($p < .03$). Columns 2 and 3 show the effect of the Fair Trade label on each of the test coffees separately: we find that the increase in demand is estimated at 7% ($p < .10$) for the FR Regular and at 12% ($p < .03$) for the CB coffee.

In Table 5 we present the results from the discrete choice analysis, where we have imposed more structure in terms of the choice sets of the consumers and we can examine substitution effects. The findings reported in the first two columns confirm the results from the reduced-form analysis: the Fair Trade label has a positive and significant effect on sales of both the FR Regular and CB coffees. The first column examines combined sales for both labeled coffees: sales increased by about 10% with the Fair Trade label ($p < .01$). The second column considers the effect of the Fair Trade label treatment on sales for each of the test coffees and all other inside bulk coffees. We find that the application of the label increased sales of CB coffee by about 13% ($p < .03$) and increased sales of FR Regular coffee by about 8% ($p < .09$).

The additional rows in column 2 in Table 5 examine possible substitution effects. We find that sales of four out of the five alternative (unlabelled) bulk coffees decreased, although individually the effect is only significant at conventional levels for the French Roast (FR) Extra Dark, which is presumably the closest substitute to the labeled French Roast (FR) Regular coffee in terms of flavor. Sales for the less popular French Roast Extra Dark coffee decreased by about 9% ($p < .05$) as a result of placing the Fair Trade label on the FR Regular counterpart. Notice that total sales of all bulk coffees increased by about 1.6% for stores in the treatment condition, although this increase is not statistically significant at conventional levels.²²

Taken together the results provide strong evidence that consumers reacted positively to the Fair Trade label by increasing demand for labeled coffees. In the absence of customer-level data, we are unable to firmly establish the extent to which this represents new demand, as opposed to a substitution effect away from unlabelled coffees.²³ However, the relatively small elasticities of substitution suggest that part of the observed increase in demand for FR Regular and CB coffees represents new demand and not only switching by customers

²²Since the total sales increase for FR Regular is still significantly higher than the reduction in sales of FR Extra Dark, these results do not fully support the hypothesis of a “reputation stealing externality” imposed by the Fair Trade label on unlabelled competitor products (Benabou and Tirole 2006).

²³The grocery store partner did not collect customer-level data so we are unable to identify the average size of each customer’s purchases and/or distinguish between repeat and new customers.

between Fair Trade labeled and unlabelled coffees.²⁴

D. THE PRICE EXPERIMENT

Columns 3 and 4 in Table 5 present the results for the Price experiment. The third column examines combined sales for the FR Regular and CB coffees: sales decreased by around 17% as a result of the one dollar price increase applied to these two test coffees, but this aggregate result masks important heterogeneity of treatment effects. In column 4 we allow the price effect to vary across the different coffees. We find that for the FR Regular coffee, the 8% increase in price did not reduce sales: sales were actually 2% higher at the test price of \$12.99 compared to the regular control price of \$11.99. As shown in the right panel of the table, this corresponds to an own-price elasticity of .28 with a 90% confidence interval of (-1.54; 2.12) suggesting a relatively less elastic demand for this coffee when the price increase is explicitly linked to Fair Trade certification. In contrast, the 9% increase in the price of the CB bulk coffee from \$10.99 to \$11.99 resulted in sales falling by more than 30%, suggesting that demand for this less expensive bulk coffee is quite elastic despite the Fair Trade label with the message associating the higher price with the ethical certification. As shown in the right panel, the decline in sales of the bulk CB coffee corresponds to an estimated own-price elasticity of -3.32 (-4.26; -2.38).

The additional rows in column 4 examine the effect that the price increase for the FR Regular and CB coffees have on sales of the alternative bulk coffees in the stores. Most notably, the decline in sales for CB is matched by a strong substitution effect that increased sales of Colombian Supremo, the only other bulk coffee that was offered at the lowest price

²⁴As an additional check, we also tested whether the Fair Trade labels on the FR Regular and CB coffees in the bulk coffee section of the stores had any impact on sales of packaged versions of these coffees sold separately. We found no significant impact on sales of packaged coffees.

of \$10.99 per pound. Sales for the Colombian Supremo coffee increased by almost 16%, corresponding to a substantial cross-price elasticity of 1.90 (-1.38; 5.17). There seem to be no strong substitution effects for the other competitor coffees. In particular, there was no substitution towards FR Extra Dark, the coffee that is closest in type to the FR Regular coffee and that was now being sold at a lower price relative to the FR Regular test coffee.

The results suggest that different customers react to the price increases for the Fair Trade labeled coffees in different ways. For customers buying the more expensive and more popular FR Regular coffee, demand for the labeled coffee was significantly less elastic: this segment was willing to pay a sizeable premium (8%) for Fair Trade labeled coffee. Customers buying the cheaper CB coffee, on the other hand, appeared to switch to the less expensive alternative coffee in response to a price rise, indicating that they were not willing to pay a premium for Fair Trade.²⁵ In the absence of detailed consumer-level data we are unable to fully specify the preferences of these two types of consumers and examine segmentation. Nonetheless, it is worth recalling from the summary statistics that the FR Regular coffee accounts for about 11% of total bulk coffee sales, while the CB coffee accounts for only 5.7% of sales, suggesting that the group of customers for which demand was less elastic was substantially larger. Also, note that total sales of all bulk coffees increased by about 1.8% under the test prices, although this increase is not statistically significant at conventional levels.

²⁵To test the relationship between price elasticities and income levels among customers, we broke down our store sample into higher and lower income areas based on median household income data for each zip code area obtained from the Census. The results were inconclusive. In higher income areas, we observe a marginally lower price elasticity for the FR coffee but a higher price elasticity for the CB coffee, and confidence intervals were overlapping between higher and lower income areas.

E. BENCHMARK ELASTICITIES OF DEMAND FOR UNLABELLED COFFEE

In order to better interpret our results from the Price experiment, we investigated how shoppers responded in the past to changes in the prices of bulk coffees in the absence of labeling that associated pricing with Fair Trade certification. We computed the own-price elasticities for all inside bulk coffees based on historical sales data. The identifying variation in prices is based on price changes that resulted from routine sales promotions. These sales promotions typically involve lowering the retail price of a single bulk coffee by \$1 per pound for one week and are administered in all stores simultaneously nationwide. During the promotions, prices of all the other bulk coffees are held at their regular levels. Which bulk coffee is chosen for a promotion at any given time depends on a rotational schedule that is drawn up by the national sales team well in advance of implementation (there is usually a 3-4 month lead time). Given the way these sales promotions are scheduled and managed at the national level, we believe that pricing endogeneity is not a significant concern when estimating the elasticity of demand for each coffee type during sales periods (and particularly not beyond the product-store mean level).²⁶ While these non-experimental estimates are less ideal than a separate set of experimental results that would match the results presented above, they should still provide reliable benchmarks of own price elasticities for the same bulk coffees in the absence of messages linking prices with Fair Trade.

In order to estimate these benchmark elasticities, we utilize weekly sales and price data for all stores from 2007-2009, discarding the weeks during which our experiments took

²⁶Store managers do not have authority to implement sales promotions autonomously based on local conditions. As a result there is almost no between-store variation that can be exploited. This renders the use of “Hausman” instruments of average prices in other markets infeasible. Notice also that wholesale prices, which are sometimes used as instruments in this context, are not available. Wholesale prices do not vary between stores, or over time during the period under study.

place. We estimate elasticities using a logit specification where the normal utility level δ_{jt} is regressed on the product prices p_{jt} , a full set of store/product fixed effects and a quadratic time trend. Standard errors are again clustered by store. Elasticities can then be estimated from the price coefficient and product shares.²⁷

Figure 3 shows the estimated own-price elasticities with their 90% confidence intervals based on the historical sales data, alongside the own-price elasticities for the labeled test coffees previously estimated from the Price experiment (the coefficient estimates are reported in Appendix A). Not surprisingly, the elasticities from the historical sales data are more precisely estimated than those from the Price experiment given the longer time span in the historical sales data. We find that the own-price elasticities of the unlabelled bulk coffees all tend to cluster around -4, indicating highly elastic demand. This is true also for the two test coffees, the FR Regular and the CB coffee, outside the weeks of the experiment and without the label inducing consumers to connect the price change specifically with Fair Trade certification.²⁸ These estimates are consistent with previous findings. While aggregate demand for coffee as a commodity is widely regarded as being inelastic (Larson 2003), several studies have indicated that demand for specific types or brands of coffee is highly elastic with average elasticities of -7 (Krishnamurthi and Raj 1991; Bell et al. 1999).

Most importantly, and what stands out in Figure 3, is that the estimates of price elasticities for unlabelled coffees are markedly higher (in absolute terms) than the estimated price elasticity of demand for the FR Regular coffee during the Price experiment when we attached the label linking the price premium to Fair Trade certification. For the CB

²⁷In a given market, the elasticity of demand for product j w.r.t to a price change in product l is given by $\eta_j^l = \frac{\partial s_j/s_j}{\partial p_l/p_l} = \left(\frac{p_l}{s_j}\right) \frac{\partial}{\partial p_l} \left(\frac{\exp(\delta_j)}{1+\sum_j \exp(\delta_j)}\right) = \frac{p_l}{s_j} \alpha(-s_j s_l + 1\{l=j\}s_l)$.

²⁸As a robustness check we replicate the elasticities using a non-linear Almost Ideal Demand System (Deaton and Muellbauer 1980) and the results, reported in the Web Appendix, are very similar to those from the logit model.

coffee, the price elasticity measured when the coffee was sold with the Fair Trade label was actually very similar to the estimated price elasticity at other times when it was sold without the label. Customers who buy the lower-priced CB coffee are sensitive to price and this sensitivity is not affected by information linking price to Fair Trade certification. But for the higher-priced FR Regular coffee, customers are far less sensitive to price when the price premium is associated with Fair Trade certification than when the same coffee is sold without the Fair Trade label. The price elasticity of demand for FR Regular coffee when sold without the Fair Trade label is roughly the same as the elasticities of the other unlabelled coffees; demand for this coffee is only less sensitive to price when the price increase is associated directly with Fair Trade certification. This suggests that customers buying the FR Regular coffee responded directly to the Fair Trade label applied in the Price experiment.²⁹

V. DISCUSSION

In this paper we provide original data on the impact of an ethical product label on consumer behavior based on a field experiment conducted in partnership with a major US grocery store chain. The first key finding is that consumers value the ethical label. Holding all other product attributes constant, the Fair Trade label by itself had a positive

²⁹An important caveat to bear in mind when interpreting the results is that these benchmark consumer elasticities are estimated based on price decreases (promotions), while in the Price experiment, elasticities are calculated based on a price increase. One concern would be that the higher benchmark elasticities derived from promotions were driven by a stocking up effect, whereby consumers stocked up on the product during promotion periods. However, the high frequency of coffee promotions in our stores, the relatively high unit cost of coffee and the fairly rapid loss of coffee flavor during storage mitigate this concern (Gupta 1988; Krishnamurthi and Raj 1991).

and significant effect on sales. Sales of the two most popular items, the FR Regular and CB bulk coffees, rose by almost 10% when the coffees carried a Fair Trade label compared with a generic placebo label. Second, we find that consumers exhibit differential levels of price sensitivity when considering the Fair Trade label. Consumers buying the lower-priced CB coffee were price sensitive and were unwilling to pay a premium of 9% to support Fair Trade. Consumers buying the higher-priced FR coffee were much less price sensitive when the coffee was labeled Fair Trade. They were willing to pay a sizeable premium (8% in the experiment) when the price premium was directly associated with support for Fair Trade certification.

A potential concern with our Label experiment is that the generic placebo label might have had a negative impact on demand for the test products in the stores under the control condition. While we are unable to completely dismiss this possibility, we argue that it is highly unlikely. To test for this possibility we exploit pre-treatment sales data from the weeks prior to the start of the Label experiment (the results from this test are reported in the Web Appendix). First, we focus on the sample of stores that were assigned to the control condition in the first phase of the experiment and compare sales of the test coffees in the period before the experiment and during the first phase of the experiment when the generic placebo label was placed on the coffees for four weeks. Combined sales of the test coffees remained stable in these stores when they entered the control condition and displayed the generic placebo label indicating that the label had no negative effect on sales (the effect estimate is 0.4%, $p. < .96$). Next, we conduct the same comparison for the stores that were assigned the treatment condition (the Fair Trade label) in the first phase of the experiment. There we find that sales of the test coffees increased substantially once they displayed the Fair Trade label (the effect estimate is 15%, $p. < .03$). These results strongly suggest that the treatment effect uncovered in the full crossover experiment is driven by the Fair Trade label increasing sales as opposed to the generic placebo label lowering sales.

Note that these findings also provide a robustness check that helps address concerns about potential carry-over effects in the crossover design from switching from the treatment to the control conditions (or vice versa). Comparing the changes in sales from the pre-experimental period to the first four weeks under the Fair Trade label and the generic placebo label yields an experimentally identified difference-in-differences estimate that implies that the Fair Trade label raised sales by 15% ($p < .13$) over the generic placebo label during the first phase of the experiment. The fact that this first phase effect is similar to the effect estimated from the full crossover experiment reported above is consistent with a no carry-over assumption since the first phase is not affected by carry-over from switching from the treatment to the control condition or vice versa. As another robustness check, we also replicated the crossover analysis while restricting the sample to sales during the last two weeks of each experimental phase, when possible carry-over effects are less likely to occur because we allow for a two week “washout” period. The results are again similar to the ones for the full crossover period. The positive label effect is if anything slightly larger in magnitude (13%, $p < .001$) which is consistent with the idea that in our context, carry-over primarily acts to attenuate treatment effects since perceptive repeat customers who value Fair Trade continue to buy the test coffees even when the label is switched to the generic placebo label.

Potential concerns for our Price experiment are the possibilities that consumers of the FR Regular coffee (but not the CB coffee) might have perceived the higher price as signalling higher product quality, independent of the Fair Trade label, or that consumers of the FR Regular coffee (but not the CB coffee) might have strong taste-based preferences and hence inelastic demand. While our design prevents us from definitively ruling out either possibility, we again argue that both are highly unlikely. Our analysis of historical sales data shows that demand for all coffee types (including the FR Regular and the CB coffee) exhibited similarly high (and negative) price elasticities, indicating that customers

typically substitute between the various coffee types in response to changes in prices. This is consistent with findings from previous empirical studies of elasticities of demand for coffee types and brands, as we noted above. Moreover, our analysis suggests that there was nothing distinctive or exceptional about demand for the FR Regular coffee prior to the experiment: the price elasticity for the FR Regular coffee was similar to price elasticities of all the other coffee types, including those typically sold at the same price and those sold at slightly lower prices (such as the CB coffee). This suggests that consumers of the FR Regular were not distinctive in either the way they interpreted signals about quality or the strength of their preferences.

Overall our findings suggest that there is substantial consumer support for Fair Trade, although some price-sensitive shoppers, accounting for a smaller volume of sales relative to the Fair Trade supporters in our sample, will not pay a large premium for the Fair Trade label. The suggested heterogeneity in consumer willingness to pay for ethical labels highlights the importance of having a clearer understanding of how different consumers assess different product attributes. How generalizable are our findings? We conducted the experiments in partnership with a grocery retailer that is associated with, among other things, relatively high prices compared with other grocery chains and stronger support for organic farming and social and environmental causes. Shoppers in our stores may thus tend to have higher incomes and more interest in social and environmental causes than the average consumer.³⁰ It is difficult to generalize from our results to other settings and other sets of consumers, and we do not claim that our shoppers are representative of the universe of shoppers in terms of their preferences and sensitivity to prices. The overall direction of the potential bias is, however, not obvious. Individuals with higher incomes may be

³⁰Data from the 2000 U.S. Census indicate that the median household income for zip codes in which our stores were located in the Northeast region was \$60,111, compared with a median income of \$54,140 for the Northeast Census region as a whole.

more likely than others to donate money to help people in need, since they have additional resources and less anxiety about their own economic circumstances. On the other hand, evidence suggests that lower income individuals give proportionally more of their incomes to charity than do higher income counterparts (Frank 1996; Andreoni 2001). Survey studies typically find no clear connection between willingness to pay for Fair Trade and other ethically labeled products and income levels of respondents (Dickson 2001; De Pelsmacker et al. 2005). As a result, it is not readily apparent whether findings from a study of a relatively high-income sample of consumers would tend to overestimate or underestimate the strength of demand for ethically-labeled goods among the broader population.

It is also unclear whether the same tests conducted in a retail environment in which appeals to social preferences are less common would reveal a larger or smaller impact of the Fair Trade label. For one, the label may be more salient for consumers in an environment in which there is less competition in terms of cause-related marketing. On the other hand, the shoppers in our experiments are more likely to have been better informed about what the Fair Trade label represents than is the case in other retail environments.³¹ Note that, from this perspective, our findings are more likely to reflect the true preferences for ethical labels among (fully informed) consumers.

In addition, we conducted the experiments during one of the worst recessions in post-war history, raising prices of goods when retailers everywhere were cutting prices. It seems likely that consumer sensitivity to price increases in this period may have been particularly

³¹We conducted a short survey of about 450 shoppers in stores that participated in the experiments to assess the extent to which consumers were familiar with the Fair Trade label. When shown the Fair Trade label, 90% of respondents were able to identify what it represented. By contrast, a 2007 study by PSL Marketing found that 23% of U.S. consumers recognized the Fair Trade logo and a 2006 LOHAS study by the Natural Marketing Institute reported that 39% of US adults recognized the logo.

high. This should bias against finding higher willingness to pay a premium for the Fair Trade label. Ultimately, questions about generalizing the results would be best addressed by replicating the tests with different retailers, different products, and in different phases of the business cycle.

Overall, we suggest that in identifying significant support for ethical product labeling in a large-scale, multiple-store field experiment in the United States, our results could help motivate future research on consumer behavior and social preferences. An important future challenge is to provide a better understanding of the exact motivations of the consumers who respond to ethical product labeling. Intrinsic forms of motivation to purchase Fair Trade products may stem from “pure” altruism when consumers derive private satisfaction from contributing to the well-being of others or from reducing global inequality (Fehr and Schmidt 1999; Becchetti and Rosati 2007); or “impure” forms of altruism when consumers derive “warm glow” type satisfaction simply from feeling better for giving to a cause (Andreoni 1989, 1990; Baron 2009).³² Alternatively, consumers may be extrinsically motivated by the anticipated impact that purchasing Fair Trade may have on their social status (Hollaender 1990; Freeman 1997; Cialdini 2003; Goldstein et al. 2008), on their self-image (Batson 1998; Benabou and Tirole 2006) or on their reputation (Glazer and Konrad 1996; Harbaugh 1998; Fehr and Fischbacher 2003; Benabou and Tirole 2006). Finally, an additional extrinsic motivation for purchasing Fair Trade products could be the perception of higher product quality. Consumers could interpret ethical production standards, along

³²Empirical research on these specific types of motivations is limited. However, one set of findings consistent with pure altruism is from a survey experiment examining consumers’ stated willingness to pay for Fair Trade (Hicks 2007) which showed that the amount individuals were prepared to pay rose when they were provided with information about the positive impact of the program (specifically, information about the percentage of farmers participating, and their revenues from Fair Trade sales).

with support for ethical causes and corporate social responsibility initiatives more generally, as a signal that the producing firm is an honest and reliable type that will not skimp on quality (Fisman et al. 2006; Siegal and Vitaliano 2007; Elfenbein et al. 2012). Besides addressing these specific types of potential motivations among consumers, future research could also identify the market for ethically labeled products according to socio-demographic segments such as age, gender, education, and income by combining experimental designs with individual-level data on purchasing behavior and consumer characteristics.³³ This research could then potentially clarify the conditions under which firms can boost sales and increase market share by offering Fair Trade certified goods, either targeted to particular segments and priced at a premium, or marketed more generally at regular prices.

³³Existing research based mostly on survey data reveal mixed or inconclusive results as to whether support for ethically labeled products is associated with key socio-demographic characteristics (e.g., De Pelsmacker et al. 2005). More recently, Cesarini et al (2009) suggested that genetic differences can explain a significant portion of individual-level variation in preferences for giving.

ACKNOWLEDGEMENTS:

With thank the and teams at our grocery store partner for their invaluable support with this project. For excellent research assistance we thank Brooke Berman, Jill Bunting, Emily Clough, Catherine Corliss, Christina Giordano, Hamed Ibrahim, Shilpa Katkar, Janet Lewis, Emily Naphtal, Eric Portales, Elisha Rivera, Shahrzad Sabet, and Tessy Seward. For comments on earlier versions of this paper we thank Daniel Diermeier and Lakshmi Iyer and numerous participants in seminars at MIT, Princeton, the London School of Economics, and at the Harvard-Princeton Political Institutions and Economic Policy workshop. The usual disclaimer applies.

VI. BIBLIOGRAPHY

Abadie, Alberto. 2002. "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models." *Journal of the American Statistical Association*, 97 (457): 284-292.

Akerberg, Daniel, C. Lanier Benkard, Steven Berry and Ariel Pakes. 2007. "Econometric Tools for Analyzing Market Outcomes.", in *Handbook of Econometrics: Volume 6A*, ed. James Heckman and Edward E. Leamer, North Holland: chapter 63.

Akerlof, George. 1970. "The Market for "Lemons": Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics*, 84 (3): 488-500.

Andreoni, James. 1989. "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence." *Journal of Political Economy*, 97 (December): 1447-58.

Andreoni, James. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." *Economic Journal*, 100 (June): 464-77.

Andreoni, James. 2001. "The Economics of Philanthropy." in *International Encyclopedia of the Social and Behavioral Sciences*, ed Neil J. Smelser and Paul B. Baltes. Oxford, England: Elsevier: 11369-11376.

Andreoni, James. 2006. "Philanthropy." in *Handbook of the Economics of Giving, Altruism and Reciprocity*, ed Serge-Christophe Kolm and Jean Mercier Ythier, Amsterdam: Elsevier/North-Holland: 1201-1269.

Arnot, Chris, Peter C. Boxall, and Sean Cash. 2006. "Do Ethical Consumers Care About Price? A Revealed Preference Analysis of Fair Trade Coffee Purchases" *Canadian Journal of Agricultural Economics*, 54 (4): 555-565.

Arnould, Eric J., Alejandro Plastina, and Dwayne Ball. 2006. "Market Disintermediation and Producer Value Capture: The Case of Fair Trade Coffee in Nicaragua, Peru and Guatemala." mimeo. University of Wyoming.

Auger, Pat, Paul Burke, Timothy M. Devinney, and Jourdan J. Louviere. (2003).

“What Will Consumers Pay for Social Product Features?” *Journal of Business Ethics* 42 (3): 281-304.

Auger, Pat, Paul Burke, Timothy M. Devinney, and Jourdan J. Louviere. (2008). “Do Social Product Features have Value for Consumers?” *International Journal of Research in Marketing* 25: 183-191.

Bacon, Christopher, V. Ernesto Mendez, Maria Gomez, Douglas Stuart, and Sandro Flores. 2008. “The State of Smallholder Livelihoods in Nicaragua: A Participatory Evaluation Using the Millennium Development Goals to Assess Sustainable Coffee Certifications.” mimeo. University of California, Santa Cruz.

Bagwell, Kyle, and Michael Riordan. 1991. “High and Declining Prices Signal Product Quality”, *American Economic Review*, Vol. 81: 224-39.

Baron, David P. 2009. “Clubs, Credence Standards, and Social Pressure.” in *Voluntary Programs: A Club Theory Perspective*, ed Matthew Potoski and Aseem Prakash. Cambridge MA: MIT Press: 41-66.

Batson, C. Daniel, and Tecia Moran. 1999. “Empathy-Induced Altruism in a Prisoner’s Dilemma”. *European Journal of Social Psychology*, 29: 909-924.

Becchetti, Leonardo, and Nazaria Solferino. 2005. “The Dynamics of Ethical Product Differentiation and the Habit Formation of Socially Responsible Consumers.” *Universita di Roma*.

Becchetti, Leonardo, and Furio Rosati. 2007. “Globalisation and the death of distance in social preferences and inequity aversion: empirical evidence from a pilot study on fair trade consumers”, *The World Economy*, 30 (5): 807-30.

Becchetti, Leonardo, and Marco Constantino. 2008. “Fair Trade on marginalised producers: an impact analysis on Kenyan farmers”, *World Development* Vol. 36, (5): 823-842.

Bell, David R., Jeongwen Chiang and V. Padmanabhan. 1999. “The Decomposition of Promotional Response: An Empirical Generalization.” *Marketing Science* 18: 504-26.

Benabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review*, 96 (5): 1652-1678.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica*, Econometric Society, vol. 63(4): 841-90.

Berry, Steven. 1994. "Estimating Discrete-Choice Models of Product Differentiation". *Rand Journal of Economics*, 25(2): 242-262.

Bonroy, Olivier and Christos Constantatos. 2003. "Consumer Uncertainty about which Firm Sells the High Quality: On the Slow Penetration of Some Credence Goods." Université Laval.

Bonroy, Olivier and Christos Constantatos. 2008. "On the Use of Labels in Credence Goods Markets." *Journal of Regulatory Economics*, 33 (3): 237-52.

Blundell, Richard and Jean-Marc Robin. 1999. "Estimation in Large and Disaggregated Demand Systems: An Estimator for Conditionally Linear Systems," *Journal of Applied Econometrics*, John Wiley & Sons, Ltd., May-June 14(3): 209-32.

Carlton, Dennis. 1978. "Market Behavior with Demand Uncertainty and Price Inflexibility." *American Economic Review*, 68 (4) : 571-587.

Carpenter, Gregory, Rashi Glazer and Kent Nakamoto. 1994. "Meaningful Brands from Meaningless Differentiation". *Journal of Marketing Research*, 31 (3): 339-350.

Caswell, Julie and Eliza Modjuszka. 1996. "Using Informational Labeling to Influence the Market for Quality in Food Products". *American Journal of Agricultural Economics*, 78 (5): 1248-53.

Cesarini, David, Christopher Dawes, Magnus Johannesson, Paul Lichtenstein and Bjorn Wallace. 2009. "Genetic Variation in Preferences for Giving and Risk Taking." *Quarterly Journal of Economics*, 124 (2): 809 -842.

Cialdini, Robert. 2003. "Crafting Normative Messages to Protect the Environment." *Current Directions in Psychological Science*, 12 (4): 105-109.

Darby, Michael and Edi Karni. 1973. "Free Competition and the Optimal Amount of Fraud." *Journal of Law and Economics* 16 (1): 67-88.

Deaton, Angus S. and John Muellbauer. 1980. "An Almost Ideal Demand System." *American Economic Review*, 70(3): 312-26.

De Pelsmacker, Patrick, Liesbeth Driesen and Glenn Rayp. 2005. "Do Consumers Care about Ethics? Willingness to Pay for Fair-Trade Coffee." *Journal of Consumer Affairs*, 39 (2): 361-383.

Devinney, Timothy, Pat Auger and Giana Eckhardt. 2010. "The Myth of the Ethical Consumer", Cambridge University Press.

Dimitri, Carolyn and Catherine Greene. 2002. "Recent Growth Patterns in the U.S. Organic Foods Market." U.S. Department of Agriculture. Agriculture Information Bulletin Number 777. Washington DC.

Elliott, Kimberly Ann and Richard Barry Freeman. 2003. "Can Labor Standards Improve Under Globalization?" Washington, DC: Institute for International Economics.

Elfenbein, Daniel and Brian McManus. 2010. "A Greater Price for a Greater Good? Evidence that Consumers Pay More for Charity-Linked Products." Forthcoming, *American Economic Journal: Economic Policy*, 2(2): 28-60.

Elfenbein, Daniel, Ray Fisman, and Brian McManus. 2012. "Reputation, Altruism, and the Benefits of Seller Charity in an Online Marketplace." *Review of Economic Studies*, 79(4): 1441-68.

Fairtrade Labelling Organizations (FLO) International. 2010a. Facts and Figures.http://www.fairtrade.net/facts_and_figures.html.

Fairtrade Labelling Organizations (FLO) International. 2010b. "Why do Some Fair-trade Products Cost More?" <http://www.fairtrade.net/faqs.html>.

Fairtrade Labelling Organizations (FLO) International. 2012. Annual Report 2012-13.<http://www.fairtrade.net/annual-reports.html>.

Fehr Ernst and Klaus Schmidt. 1999. "A Theory of Fairness, Competition and Cooperation." *Quarterly Journal of Economics*, 114 (3): 817-868.

Fehr, Ernst, and Urs Fischbacher. 2003. "The Nature of Altruism." *Nature* 425 (23 October): 785-791.

Fisman, Ray, Geoffrey Heal, and Vinay Nair. 2006. "A Model of Corporate Philanthropy". Columbia University and University of Pennsylvania.

Frank, Robert H. 1996. "Motivation, Cognition, and Charitable Giving." in *Giving*, ed Jerome Schneewind, Indiana University Press: 130-152.

Freeman, Richard B. 1997. "Working for Nothing: the Supply of Volunteer Labor". *Journal of Labor Economics*, 15(1): 140-166.

Galarraga, Ibon and Anil Markandya. 2004. "Economic Techniques to Estimate the Demand for Sustainable Products: A Case Study for Fair Trade and Organic Coffee in the United Kingdom." *Economia Agraria y Recursos Naturales*, 4(7), 109-134.

Glazer, Amihai and Kai Konrad. 1996. "A Signaling Explanation of Charity." *American Economic Review*, 86 (4): 1019-1028.

Goldstein, Noah, Robert Cialdini, and Vladas Griskevicius. 2008. "A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels." *Journal of Consumer Research*, 35.

Gorman, W. M. 1980. "A Possible Procedure for Analyzing Quality Differentials in the Egg Market." *Review of Economic Studies* 47 (October): 843-56.

Goldstein, Noah J. and Robert B. Cialdini. 2008. "Social Influences on Consumption and Conservation Behavior" in *Social Psychology of Consumer Behavior*, ed Michaela Wanke, New York: Psychology Press.

Gupta, Sunil. 1988. "Impact of Sales Promotions on When, What and How Much to Buy." *Journal of Marketing Research*, 25 November: 342-355.

Harbaugh, William. 1998. "The Prestige Motive for Making Charitable Transfers."

American Economic Review, 88 (2): 277-282.

Hausman, Jerry, Gregory Leonard and J. Douglas Zona. 1994. "Competitive Analysis with Differentiated Products." *Annales d'Economie et de Statistique*, Apr-June, Vol. 34: 159-180.

Hertel, Shareen, Lyle Scruggs, and C. Heidkamp. 2009. "Human Rights and Public Opinion." *Political Science Quarterly*, 124 (3): 443-459.

Hicks, Robert. 2007. "Performance-Based Labeling." in *New Frontiers in Environmental and Social Labeling*, eds Ulrike Grote, Arnab Basu, and Nancy Chau. New York: Physica-Verlag: 37-58.

Hilger, James, Greg Rafert and Sofia Villas-Boas. 2011. "Expert Opinion and the Demand for Experience Goods", *The Review of Economics and Statistics*, 2011, vol. 93, issue 4: 1289-1296.

Hiscox, Michael J. and Nicholas Smyth. 2006. "Is there Consumer Demand for Improved Labor Standards? Evidence from Field Experiments in Social Product Labeling." Harvard University: Cambridge MA.

Hollaender, Heinz. 1990. "A Social Exchange Approach to Voluntary Cooperation." *American Economic Review*, 80 (5): 1157-1167.

International Coffee Organization. 2008. Historical Coffee Statistics. London : ICO. Available at: http://www.ico.org/new_historical.asp?section=Statistics.

Jones, Byron and Michael G. Kenward. 2003. "Design and Analysis of Cross-Over Trials.", Second Edition (Monographs on Statistics and Applied Probability). New York: Chapman & Hall.

Kimeldorf, Howard, Rachel Meyer, Monica Prasad, and Ian Robinson. 2004. "Consumers of the World Unite: A Market-based Response to Sweatshops." *Labor Studies Journal* 29 (3): 57-80.

Krishnamurthi, Lakshman and S. P. Raj. 1991. "An Empirical Analysis of the Rela-

tionship Between Brand Loyalty and Consumer Price Elasticity.” *Marketing Science*, 10: 172-83.

Lancaster, Kelvin. 1971. *Consumer Demand: A New Approach*. New York: Columbia University Press.

Larson, Bruce. 2003. “Eco-labels for Credence Attributes: The Case of Shade Grown Coffee.” *Environment and Development Economics*, 8: 529-47.

Mohr, Lois and Deborah Webb. 2005. “The Effects of Corporate Social Responsibility and Price on Consumer Responses.” *Journal of Consumer Affairs*, 39 (1): 121-147.

Mullainathan, Sendhil, Joshua Schwartzstein, and Andrei Schleifer. 2008. “Coarse Thinking and Persuasion.” *Quarterly Journal of Economics*, May: 577-619.

Murray, Douglas, Laura Reynolds, and Peter Taylor. 2006. “The Future of Fair Trade Coffee: Dilemmas Facing Latin America’s Small-Scale Producers.” *Development in Practice*, 16 (2): 179-192.

Nelson, Philip. 1970. “Information and Consumer Behaviour.” *Journal of Political Economy*, 78(2): 311-329.

Nelson, Philip. 1974. “Advertising as Information.” *Journal of Political Economy* 82: 729-754.

Palfrey, Thomas, and Thomas Romer. 1983. “Warranties, Performance, and the Resolution of Buyer-Seller Disputes.” *Bell Journal of Economics*, 14(1): 97-117.

Nevo, Aviv. 2010. “Empirical Models of Consumer Behavior.” *Annual Review of Economics*, Volume 3: 51-75.

Program on International Policy Attitudes (PIPA). 2000. “Americans on Globalization.” University of Maryland.

Reinstein, David and Joon Song. 2012. “Efficient Consumer Altruism and Fair Trade Products” *Journal of Economics and Management Strategy*, Volume 21 Number 1, Spring 2012: 213-241.

Roe, Brian and Ian Sheldon. 2007. "Credence Good Labeling: The Efficiency and Distributional Implications of Several Policy Approaches." *American Journal of Agricultural Economics*, 89(4): 1020-1033.

Ronchi, Lorraine. 2002. "The Impact of Fair Trade on Producers and Their Organizations: A Case Study with Coocafe in Costa Rica." PRUS Working Paper No. 11. Poverty Research Unit at Sussex. University of Sussex. Brighton.

Shapiro, Carl. 1983. "Premiums for High Quality Products as Returns to Reputation." *Quarterly Journal of Economics*, 25: 659-79.

Spence, Michael. 1976. "Product Selection, Fixed Costs and Monopolistic Competition." *Review of Economic Studies*, Vol. 23, 2: 217-235.

Teisl, Mario, Brian Roe, and Robert Hicks. 2002. "Can Eco-Labels Tune a Market? Evidence from Dolphin-Safe Labeling." *Journal of Environmental Economics and Management*, 43: 339-359.

TransFair USA. 2009a. Almanac 2009. http://transfairusa.org/sites/default/files/Almanac_2009.pdf.

TransFair USA. 2009b. Fair Trade Certified Thriving In Tough U.S. Economy. http://www.transfairusa.org/press-room/press_release/fair-trade-certified-

VII. TABLES

Table 1: Randomization Check for Label Experiment

Covariate	Mean	Mean	t-test	KS-test
	Treated	Controls	p.-val	p.-val
Total Store Sales 2008 (\$)	26,699,212	25,603,305	0.81	0.99
Total Store Sales Growth 2008-2009 (%)	3.93	3.77	0.81	0.99
Covariates for 4 Sales Weeks Prior to Test:				
Avg. Sales FR Regular Bulk (\$)	185.34	182.05	0.96	0.99
Avg. Sales Coffee Blend Bulk (\$)	101.23	104.08	0.92	0.51
Avg. Sales All Bulk Coffees (\$)	78.23	76.86	0.94	0.99
Avg. Sales All Instant Coffees (\$)	37.08	40.65	0.47	0.84
Avg. Sales All Packaged Coffees (\$)	41.01	39.15	0.70	0.99
Covariates for 52 Sales Weeks Prior to Test:				
Avg. Sales FR Regular Bulk (\$)	188.00	184.75	0.96	0.99
Avg. Sales Coffee Blend Bulk (\$)	99.21	104.94	0.83	0.99
Avg. Sales All Bulk Coffees (\$)	77.83	75.74	0.91	1.00
Avg. Sales All Instant Coffees (\$)	32.97	36.68	0.38	0.52
Avg. Sales All Packaged Coffees (\$)	38.91	37.34	0.72	1.00
Covariates from Store Zip Code Areas:				
Total Population	26,898	21,709	0.30	0.26
African American Pop (%)	3.91	4.35	0.81	0.51
Foreign Born Pop (%)	13.53	13.86	0.92	0.52
Median HH Income (\$)	65,207	63,002	0.82	1.00
HHs with Soc. Security Income (%)	10.13	9.44	0.54	0.49
Public Assistance per capita (\$)	36.59	37.26	0.95	0.84
Family HHs (%)	60.91	57.53	0.65	0.99
HH Head Aged 15-34 (%)	21.56	24.86	0.59	0.83
HH Head Aged 65+ (%)	23.76	21.20	0.35	0.83
In High School (%)	13.05	12.71	0.89	0.99
In College (%)	11.73	13.35	0.76	0.99
High School Dropouts (%)	8.55	8.18	0.88	0.84
High School Graduates (%)	18.65	16.27	0.50	0.83
BA Degree (%)	27.48	28.95	0.61	0.85
Graduate Degree (%)	26.05	28.15	0.70	0.83

Note: N=26. Columns 2 and 3 show means of covariates in the treatment and control group of stores. Column 4 shows the p-value from a two-sample t-tests assuming unequal variances. Column 5 shows the p-value from bootstrapped Kolmogorov-Smirnov test. Census covariates refer to the zip code areas in which the stores are located (based on the 5 digit Zip code tabulations areas from the 2000 Census).

Table 2: Randomization Check for Price Experiment

Covariate	Mean	Mean	t-test	KS-test
	Treated	Controls	p.-val	p.-val
Total Store Sales 2008 (\$)	25,413,676	26,888,842	0.75	0.99
Total Store Sales Growth 2008-2009 (%)	3.74	3.95	0.75	0.99
Covariates for 4 Sales Weeks Prior to Test:				
Avg. Sales FR Regular Bulk (\$)	231.16	233.74	0.97	0.84
Avg. Sales Coffee Blend Bulk (\$)	81.92	81.25	0.97	0.92
Avg. Sales All Bulk Coffees (\$)	70.30	75.72	0.73	0.82
Avg. Sales All Instant Coffees (\$)	41.39	37.98	0.54	0.53
Avg. Sales All Packaged Coffees (\$)	41.17	45.09	0.47	0.83
Covariates for 52 Sales Weeks Prior to Test:				
Avg. Sales FR Regular Bulk (\$)	179.13	181.66	0.97	0.99
Avg. Sales Coffee Blend Bulk (\$)	94.64	97.77	0.89	0.52
Avg. Sales All Bulk Coffees (\$)	72.29	76.26	0.82	0.99
Avg. Sales All Instant Coffees (\$)	43.02	38.95	0.44	0.52
Avg. Sales All Packaged Coffees (\$)	39.10	41.40	0.63	0.83
Covariates from Store Zip Code Areas:				
Total Population	25864.00	22743.31	0.54	0.52
African American Pop (%)	4.60	3.66	0.61	0.24
Foreign Born Pop (%)	13.07	14.32	0.69	0.28
Median HH Income (\$)	58449.54	69760.08	0.24	0.23
HHs with Soc. Security Income (%)	10.12	9.45	0.55	0.52
Public Assistance per capita (\$)	42.30	31.55	0.28	0.24
Family HHs (%)	58.59	59.84	0.87	0.83
HH Head Aged 15-34 (%)	24.12	22.30	0.77	0.84
HH Head Aged 65+ (%)	23.89	21.07	0.31	0.25
In High School (%)	12.46	13.29	0.72	1.00
In College (%)	15.17	9.91	0.32	0.82
High School Dropouts (%)	10.19	6.54	0.15	0.11
High School Graduates (%)	19.03	15.88	0.37	0.53
BA Degree (%)	26.72	29.71	0.30	0.51
Graduate Degree (%)	24.43	29.78	0.32	0.53

Note: N=26. Columns 2 and 3 show means of covariates in the treatment and control group of stores. Column 4 shows the p-value from a two-sample t-tests assuming unequal variances. Column 5 shows the p-value from bootstrapped Kolmogorov-Smirnov test. Census covariates refer to the zip code areas in which the stores are located (based on the 5 digit Zip code tabulations areas from the 2000 Census).

Table 3: Summary Statistics

	Regular Retail Price \$ per lb	Avg. Weekly Store Sales lb \$		Avg. Share of Store Bulk Coffee Sales Potential Coffee Market	
Test Bulk Coffees					
FR Regular	11.99	17.2	198.8	11.0%	0.5%
Coffee Blend (CB)	10.99	9.1	98.7	6.6%	0.3%
Competitor Bulk Coffees					
Breakfast Blend	11.99	10.8	124.6	7.2%	0.3%
Colombian Supremo	10.99	10.6	114.8	7.0%	0.3%
FR Extra Dark	11.99	10.9	126.3	7.1%	0.3%
Regional Blend	11.99	10.7	126.1	7.5%	0.3%
Mexican	11.99	6.4	75.2	4.1%	0.2%

Note: Summary statistics of Coffee sales and prices in sample stores (based on store-weeks in the 2009 fiscal year).

Table 4: Effect of Fair Trade Label on Sales of Test Coffees

Model	(1)	(2)	(3)
	FR Regular &		
Coffee	Coffee Blend	FR Regular	Coffee Blend
Dependent Variable	Log Sales	Log Sales	Log Sales
Fair Trade Label	0.079 (0.033)	0.072 (0.042)	0.122 (0.052)
Constant	5.281 (0.046)	4.853 (0.076)	3.861 (0.117)
Store FE	Yes	Yes	Yes
Week FE	Yes	Yes	Yes
Observations	208	208	207

Note: Models 1-3 display regression coefficients with robust clustered standard errors in parenthesis. The unit of analysis is a store week. The dependent variable in the regressions is the logged weekly dollar sales of the test coffees, FR Regular and Coffee Blend. Model 1 refers to the combined sales of both test coffees, Model 2 and Model 3 refer to sales of the FR Regular and the Coffee Blend respectively. The independent variable is a treatment indicator coded as 1 for the four weeks in which the Fair Trade label was placed on the test coffees and 0 for the four weeks when the generic label was placed on the test coffees. The design is a fully-balanced crossover experiment so each of the 26 stores is observed for 4 weeks under each condition (8 weeks in total). All models include a full set of store and week fixed effects.

Table 5: Main Results: Label and Price Experiment

Model	Label Experiment			Price Experiment			
	(1)	(2)	(3)	(4)	Own Price Elasticity		
Dependent Variable	δ_{jt}	δ_{jt}	δ_{jt}	δ_{jt}	PE	LB	UB
FR Regular & Coffee Blend	0.103 (0.035)		-0.168 (0.059)				
FR Regular		0.077 (0.043)		0.024 (0.091)	0.28	-1.55	2.12
Coffee Blend		0.129 (0.055)		-0.360 (0.075)	-3.32	-4.26	-2.38
Breakfast Blend		-0.043 (0.074)		-0.061 (0.093)			
Colombian Supremo		-0.060 (0.064)		0.159 (0.155)			
FR Extra Dark		-0.094 (0.046)		-0.009 (0.097)			
Regional Blend		0.044 (0.099)		0.041 (0.072)			
Mexican		-0.050 (0.078)		-0.006 (0.192)			
Product/Store FE	Yes	Yes	Yes	Yes			
Week FE	Yes	Yes	Yes	Yes			
Observations	1,368	1,368	1,399	1,399			

Note: Models 1-4 display regression coefficients with robust clustered standard errors in parenthesis. The dependent variables in the regressions are the normalized mean utility levels $\delta_{jt} = \log(s_{jt}) - \log(s_{0t})$. The independent variables include treatment indicators (for the Fair Trade label and the Test Price accordingly) for each coffee. The estimation is restricted to 8 weeks in which the experiment was underway (excluding a 2 week washout period for the Price experiment). All models include a full set of product/store fixed effects and week fixed effects. Models 1-2 also include product prices (coefficients not shown). The last three columns refer to the own price elasticities computed based on Model 4 for the test coffees, where the regular unit prices are used as base prices. The experiment raised unit prices by about 8% and 9% for the Coffee Blend and the FR Regular coffee respectively. PE: Point estimate. LB and UB: lower and upper bound of the 90% confidence intervals.