Journal of Health Economics 45 (2016) 47–54

Contents lists available at ScienceDirect

Journal of Health Economics

journal homepage: www.elsevier.com/locate/econbase

Habit formation in children: Evidence from incentives for healthy eating

George Loewenstein[a], Joseph Price[b,*], Kevin Volpp[c,d,e,f]

[a] Carnegie Mellon University, Social and Decision Sciences, USA
[b] Brigham Young University, USA
[c] Philadelphia VA Medical Center, United States, USA
[d] Center for Health Incentives and Behavioral Economics, Leonard Davis Institute, USA
[e] Penn Medicine Center for Innovation, USA
[f] Perelman School of Medicine and the Wharton School, University of Pennsylvania, 1120 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021, USA

ARTICLE INFO

Article history:
Received 9 September 2014
Received in revised form 26 October 2015
Accepted 18 November 2015
Available online 3 December 2015

JEL classification:
J13
I18
I28

Keywords:
Habit formation
Incentives
School lunch
Field experiments

ABSTRACT

We present findings from a field experiment conducted at 40 elementary schools involving 8000 children and 400,000 child-day observations, which tested whether providing short-run incentives can create habit formation in children. Over a 3- or 5-week period, students received an incentive for eating a serving of fruits or vegetables during lunch. Relative to an average baseline rate of 39%, providing small incentives doubled the fraction of children eating at least one serving of fruits or vegetables. Two months after the end of the intervention, the consumption rate at schools remained 21% above baseline for the 3-week treatment and 44% above baseline for the 5-week treatment. These findings indicate that short-run incentives can produce changes in behavior that persist after incentives are removed.

Currently, there is vigorous debate about when it is either effective or appropriate to incentivize positive behaviors in children. Opponents of the use of incentives argue that extrinsic rewards crowd out intrinsic motivation and results in outcomes being worse after the end of the incentive period than prior to the introduction of rewards (Deci et al., 1999), and there is, indeed, evidence of such effects in studies conducted by economists (see Frey and Jegen, 2001 for a review). However, arguments against the use of incentives sometimes overlook the role that habit formation can play in promoting long run behavioral change. Dictionary.com defines a habit as "an acquired behavior pattern regularly followed until it has become almost involuntary." If this habit formation process occurs while individuals are incentivized to engage in a behavior, then short-term efforts that encourage children to engage in a particular activity can, if sufficient to overcome any crowding out of intrinsic motivation, result in positive behavior change even after the incentives are removed.

In this paper, we examine the role of incentives in promoting healthy eating behaviors in children. We focus on fruits and vegetables since inadequate consumption of fruits and vegetables is widely seen as an important contributor to suboptimal health worldwide, and increases the risk for cardiovascular diseases, stomach cancer and colorectal cancer. Achieving high rates of fruit and vegetable consumption among children has proved a considerable challenge and has been the focus of a number of recent school-based interventions.

We implemented an incentive program at 40 elementary schools in Utah in which children could receive a special token each day as a reward for consuming at least one serving of fruits or vegetables. The tokens were worth $0.25 and could be spent at the school store, school carnival, or book fair. Schools were randomly assigned to implement the incentives for a period of either 3 or 5 weeks. We observed detailed fruit and vegetable consumption

* Corresponding author at: 162 FOB, Brigham Young University, Provo, UT 84602, USA. Tel.: +1 801 422 5296.
E-mail address: joseph_price@byu.edu (J. Price).

http://dx.doi.org/10.1016/j.jhealeco.2015.11.004
0167-6296/

data at these schools before, during, and for 2 months after the intervention ended. This experimental design allows us to examine whether the increase in fruit and vegetable consumption that we observe during the incentive period persists once the incentives are removed.

1. Background

The results of this paper complement other recent studies that examine the impact of incentives on children's in-school food choices. Just and Price (2013) provided incentives for 5 days over a 2–3 week period and found lingering effects during the first 2 weeks after the intervention, but these did not persist 4 weeks after the intervention. Belot et al. (2013) provided students with stickers and little gifts for choosing healthy lunch items for a period of 4 weeks and find that the rewards increased fruit and vegetable consumption during the incentive period (though these effects vary by how the rewards are provided and the age and gender of the child). They find little evidence that the changed behavior persists 6 months after the end of the rewards period. List and Samek (2015) provided low income school students with a small prize as a reward for choosing a healthier snack (dried fruit) over a less healthy snack (a cookie). They observed a large impact of incentives on the children's choices that persisted even after the incentives were removed, especially when incentives were combined with a health message.

Studies of habit formation in domains other than school children's food choices have yielded mixed effects. Charness and Gneezy (2009) randomly assigned college students to one of three conditions: no incentive for gym attendance, $25 to attend the gym one time, or $25 to attend the gym one time plus $100 to attend the gym another 8 times. Their key finding was that, consistent with habit formation, subjects in the high incentive treatment group had higher gym attendance (about 0.6 more visits per week) during the post-incentive period than those in the low incentive and no incentive groups.

In a replication and extension of this study, Acland and Levy (2015) observed a smaller post-incentive effect (0.26 visits per week), and found that the effect decayed over the course of the winter vacation and was highly concentrated in the upper tail of the post-treatment attendance distribution. Royer et al. (2015) also tested a similar intervention using adult workers at a Fortune 500 company and additionally tested the impact of giving workers access to a self-funded commitment contract. They found a weak persistence of gym use after the incentive was withdrawn among those provided with an incentive alone (16% of the increase in attendance during the incentive period), but substantially greater persistence (47%) among those who were provided access to the commitment contract. .

Schofield et al. (2015) examined the impact of individually oriented, purely altruistic, and a hybrid of competitive and cooperative monetary reward incentives on older adults' completion of cognitive exercises and cognitive function. All three incentive structures approximately double the number of exercises completed during the 6-week active experimental period relative to a no incentive control condition. More relevant to habit formation, cognitive exercise use did persist to some degree beyond the official end of the study in all conditions including the control, and persistence was greater in the altruistic and cooperative/competitive incentives than in the atomistic and control conditions.

Persistence of behavior change may be easier to achieve in some contexts than in others. Volpp et al. (2009) randomized smokers into a treatment group which offered a $750 incentive ($100 for completion of a program, $250 for short-term cessation, and $400 for long-term cessation). This incentive resulted in a quit rate of 14.7% in the intervention group compared to 5.0% in the control group at 12 months. Six months after the long-term incentives were discontinued, the quit rates for the two groups were 9.4% and 3.6%, suggesting that if incentives are effective in helping an individual to stay smoke-free for 12 months, there is a reasonable chance they will develop habits that increase their likelihood of remaining smoke-free when incentives are withdrawn. In contrast, weight loss interventions have typically shown less evidence of habit formation. In two studies testing the use of lottery incentives and deposit contracts for weight loss (Volpp et al., 2008; John et al., 2011), incentives were highly effective in motivating weight loss during the incentive period, but participants regained most of the weight they had lost once the incentives ended.

One possibility for why smoking cessation is more persistent than weight loss is that weight loss involves a complex interplay between myriad decisions around food consumption and physical activity that happen at all points of the day with differing stimuli and constraints. With smoking, in contrast, quitting can be a simple decision to totally desist; one has to eat to live, but one does not have to smoke. Food choice in school cafeterias is in a few important ways a simpler behavior to change than either smoking or weight loss. Whether to take and consume fruits or vegetables in a school lunch is a relatively simple decision, and there are no immediate dire consequences to making either choice. A daily routine around a specific task such as getting a tray each day at the same time and changing one component of what is on the tray is far simpler than trying to change a whole host of elements required for more complex behavioral challenges like losing weight. There are no physiologic withdrawal symptoms for not consuming alternatives, as there is in smoking, and there is just one choice environment without myriad different stimuli and constraints, as with obesity more generally. As such, we would predict that habits can potentially form more easily when it comes to fruit and vegetable consumption during school lunches than for weight loss itself.

Although the studies just mentioned have examined habit formation in the sense of persistence of desired behaviors once incentives are removed, most or all of these studies are ambiguous about the exact mechanism that produces the effect. 'Classic' habit formation refers to, to requite the Dictionary.com definition, "an acquired behavior pattern regularly followed until it has become almost involuntary." A behavior becomes a habit, according to this definition, much as a particular path through the woods becomes easier to follow, and more difficult to depart from, as it is cleared by repeated usage. However, there is an alternative possible account of many experimental results purporting to show habit formation. It is possible that subjects acquired information—e.g., in the exercise studies, about where the locker room was and how to sign in, or about their own (latent) love of exercise. Although such learning would produce persistence once incentives were removed, it is unclear whether such persistence should be labeled habit formation.

Whether persistence varies as a function of habit formation provides a clue about which mechanism, if either, is operative. If persistence is the result of learning, one would expect the behavior to persist even after a brief intervention which would be, presumably, sufficient for learning to occur. If persistence is the result of more classic habit formation, in contrast, we would expect duration to make an important difference, because repeating a pattern of behavior more should cause it to become more ingrained. The duration required for a habit to form is likely to depend on the nature of the task, how difficult it is to learn, how much effort it takes, and whether it provides ongoing positive or negative feedback.

2. Methods

We conducted a field experiment at 40 elementary schools in Utah involving 8000 students in grades 1–6. The data were collected over a period of 18 months, from January 2012 to June 2013, with seven schools participating during the winter 2012 semester, 10 schools the following fall, and 25 schools in 2013. New lunch guidelines were put in place by the USDA at the start of the 2012–2013 school year which required that every child take a serving of fruits or vegetables with each school lunch. Eight schools participate in our program before the change in the lunch guidelines and 34 after the new guidelines. Each school participated in the experiment either completely before or completely after the change in guidelines. All of our analysis includes school fixed effects, which should account for any differences across schools in how the guidelines were implemented or other differences in their lunch program.

The schools in our study provide students two or three choices of a main entrée and allow students to choose as many additional items as they want from a selection of fruits, vegetables, and other side dishes. It was important that we could accurately measure the number of servings of fruits and vegetables that each student actually consumed, so most of the fruits and vegetables came in special cups while others, such as bananas or oranges, were quantified by the leftover peel or core. Research assistants stood by the trash cans in each cafeteria and recorded the number of fruits and vegetables both taken and consumed by each child by observing each child's tray as they exited the lunch room. In cafeterias where there were multiple trash locations, at least one assistant was stationed at each location. Prior to the rewards period, baseline data were collected at each school for 2 weeks.

During the rewards period, children receiving a school lunch who ate at least one serving of fruits or vegetables received a special coin which had a picture of an apple and a carrot on it. These coins had a value to the students of 25 cents and could be redeemed at a school store, school carnival, or book fair. We used redeemable tokens instead of cash in response to a concern expressed by some school principals that children might use the money to purchase candy or other junk food after school. At the end of the program, each school received a check for the value of tokens that were redeemed, providing the double benefit to schools of encouraging healthy eating and providing additional funds for the school and PTA.

Just prior to the start of the rewards period, an announcement about the program was included in a newsletter sent home to parents, and information about the program was provided in the school's morning announcement. Reinforcing these announcements, the research assistants handing out the tokens were instructed to explain to students why they were distributing the tokens and also reminded children who had not eaten a full serving of fruits or vegetables that if they went back and finished their fruit or vegetable they could receive a token. Thus the change in behavior during the incentive period may result from both the direct effect of the incentives as well as any effects operating through the presence and interaction of students with the data collectors.

Some schools expressed concerns that during the token period some children might be cheating by hiding their vegetables in their milk cartons or throwing food on the floor in order to receive a token. Schools provided an announcement to students about the importance of honesty and warned that cheating would result in the end of the program for the school. In addition, since our data collectors were in the cafeteria during the entire lunch period, they were able to check for food being thrown on the floor or to ask the students personally if they did indeed eat an item before handing them a token. It is important to note, however, that, while any cheating will bias upwards the estimates of consumption during the incentive period, it should have no effect on our estimates of habit formation, since those estimates are based on measures recorded after the end of the token period (when there would be no incentive to cheat).

During most of the week, the tokens were only available to students who purchased or received a school lunch since our data collection approach is not well suited to measure the fruit and vegetable consumption of students with a sack lunch. As an accommodation to these children, we made the tokens available to all students who consumed a fruit or vegetable on Fridays, including those who consumed one from a sack lunch. The children with a sack lunch were not included in any of our data collection. It is possible that students may have switched from getting a sack lunch to getting a school lunch on token days. Whether this will bias the effects during the incentive period depends on whether the switchers are more or less likely to consume fruits and vegetables during the incentive than the kids who always get a school lunch. It is likely that students more likely to eat fruits and vegetables will switch in response to the incentives.[1] These switchers may bias upwards our estimates of the effects during the incentive period but similar to the issue of cheating, they should not have any effect on our estimates of behavior change after the incentive period has ended.

All 40 schools that participated in our experiment expressed willingness to implement an incentive program at their school for up to 5 weeks. Schools that elected to participate were randomly assigned to have the rewards in place for either 3 weeks or 5 weeks. We chose 3 weeks as our smallest length since we knew from Just and Price (2013) that 5 days would be insufficient to create habit formation. We chose the 5-week treatment to match the conventional wisdom that it takes 21 days to form a habit. We had originally planned to have an 8-week treatment but found it difficult to find schools willing to agree to this length of time for an intervention. To ensure that we would have similar schools in each of our treatment groups, we stratified the randomization on two school characteristics: the baseline consumption rate at the school, and the fraction of children at the school who are on free and reduced price lunch (FRPL rate).

Unlike previous studies (e.g., Just and Price, 2013; List and Samek, 2015) this study did not include a control group. We made this decision so as to allocate scarce research dollars to the two experimental groups, and because it would have been difficult to recruit schools while having to fully disclose that some of them would incur the costs of hosting monitors, but would not get the intervention. To be a true control school, the school would have had to be willing to agree to 2 weeks of baseline data collection, 3 or 5 weeks to match the treatment period, and then 2 months of follow-up data without receiving any direct benefits of the program. In addition, schools which found themselves being measured but not treated may have been more likely to drop out, creating selective attrition from our experiment.

One purpose of including a control group is to rule out any changes in behavior resulting from environmental and/or policy changes occurring at the same time as the treatment. We address this issue by staggering the start times of the treatment across the schools in our sample. Another purpose of a control group would be to net out the effect that being observed has on behavior, as well as any secular changes in diet that tend to occur.

[1] Just and Price (2013) find that their rewards program increased the fraction of children eating school lunch that day by 4%. They note that if all of the children who switched over on the incentive days were already eating a serving of fruits or vegetables with their sack lunch that the effect of the incentive would need to be scaled down by one eighth. The bias could be negative if the switchers are less likely to consume fruits and vegetables than those students who have been getting a school in the past. This difference in consumption could result from sack lunches, on average, being much less likely to include fruits and vegetables (Rainville, 2001; Sweitzer et al., 2009).

Table 1
Comparison of baseline characteristics across treatment groups.

	3 weeks	5 weeks	p-value
Number of students per school	663 [190]	629 [150]	0.55
Gender ratio			
Male	50.6 [2.6]	50.6 [1.8]	0.97
Grade	3.51 [0.37]	3.35 [0.27]	0.13
Ethnicity			
White	79.4 [14.7]	75.4 [20.3]	0.47
Hispanic	14.5 [13.1]	18.5 [17.5]	0.41
Other	6.1 [3.4]	6.0 [4.5]	0.98
FRPL rate	40.5 [16.2]	45.3 [22.2]	0.43
Fraction of children eating at least one serving of fruits or vegetables	39.9 [11.3]	37.6 [12.0]	0.58
Amount of servings consumed per child	0.577 [0.171]	0.602 [0.174]	0.83
Number of schools	22	18	

Notes: Values listed above are averaged over all schools in each group. FRPL rate is the fraction of students who receive a free or reduced price lunch. Standard deviations are included in brackets. The p-value is based on a t-test for the difference in characteristics between the 3- and 5-week schools.

Although we do not have direct evidence of whether either of these effects occurred in the current study, evidence from previous studies weighs against either of these effects leading to exaggerated treatment effects. For example, Just and Price (2013) included control schools in their experiment and contrary to the idea that simply observing students increase fruit and vegetable consumption, they observed a decline in the fraction of students eating a serving of fruits and vegetables over the course of the observation period. This was true across the 5 baseline days for all of the 15 schools in their sample as well as across the full 15 days at the two control schools. List and Samek (2015) also incorporated a control group, which also experienced a slight decline in the fraction of children eating the healthy item over the course of the observation period.

Our primary outcome measure is an indicator for whether or not the child ate at least one serving of fruits or vegetables which was the criterion we used to determine whether a student received a reward or not. For each student we recorded the number of servings that they took and how many they actually ate. Measuring actual consumption was key since the majority of our schools participated in the field experiment after the implementation of the new lunch guidelines that required that students place at least one serving of fruits or vegetables on their tray. Our rewards program was designed to counteract the fact that these new guidelines were leading to such a large number of fruit and vegetables being thrown away. We also report results for a secondary measure: the number of servings of fruits or vegetables actually consumed per student.

Unfortunately, we were unable to collect any data on food consumption or calories consumed outside of school lunch. In general, adopting a diet with more fruits and vegetables will tend to crowd out more energy-dense foods since fruits and vegetables are high in fiber but low in calories. It is possible that encouraging increased fruit and vegetable consumption during lunch will add calories to a child's diet, though these calories might be offset by less unhealthy snacking after school. However, the health benefits of encouraging fruit and vegetable consumption extend well beyond any effects on obesity since fruits and vegetables are rich in vitamins, antioxidants, and fiber and have been shown to protect against various common diseases (Kant 2004; Hung et al., 2004).

Our final sample includes 40 elementary schools, 22 of which had the rewards period in place for 3 weeks and 18 which had it in place for 5 weeks. For our analysis, we use the student-day as the unit of observation but cluster all of our standard errors at the school level. Our combined sample has 403,922 child-day observations including the baseline, incentive, and follow-up periods. Running our analysis at the child-level allows us to include controls for student characteristics such as gender or grade. Our data collection procedure and IRB restrictions made it impossible to collect information on each student's identity so we are unable to include any student fixed effects though we include both school fixed effects and day of the week fixed effects in our analyses.[2]

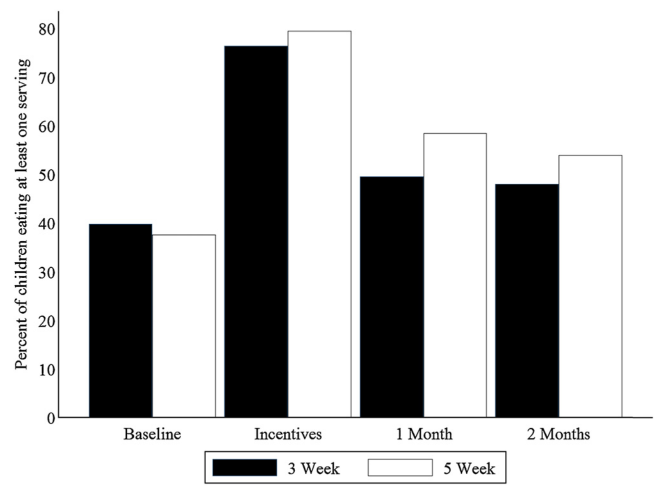

Fig. 1. Levels of consumption before, during and after the incentives period. The baseline period includes the 10 days prior to the start of the incentives. "1 month" and "2 months" refer to the first and second month after the end of the incentive period.

3. Results

In Table 1, we use data from the Common Core of Data from the National Center for Education Statistics to provide some basic characteristics of the schools in our sample. We also provide the p-value of the t-test for whether the characteristics differ between the two treatment groups. At the bottom of Table 1, we provide some measures from our baseline data collection. Among the students at the schools in our sample, about 39% were eating at least one serving of fruits or vegetables and the average student was eating about 0.59 servings of fruits or vegetables every day. We find that none of the differences between the two treatment groups at baseline are statistically significant, suggesting that our randomization procedure was successful in balancing the characteristics across the two treatment groups.

Fig. 1 provides the fraction of children eating at least one serving of fruits or vegetables during each of the four periods of the study

[2] All of the IRB documents used in this study are available from the authors. Our data provided no easy way to exclude individual students from the study once a school decided to participate. As a result, the PTA had to reach consensus about whether to participate in the study. If parents voiced concerns about the study then the study did not occur at that school.

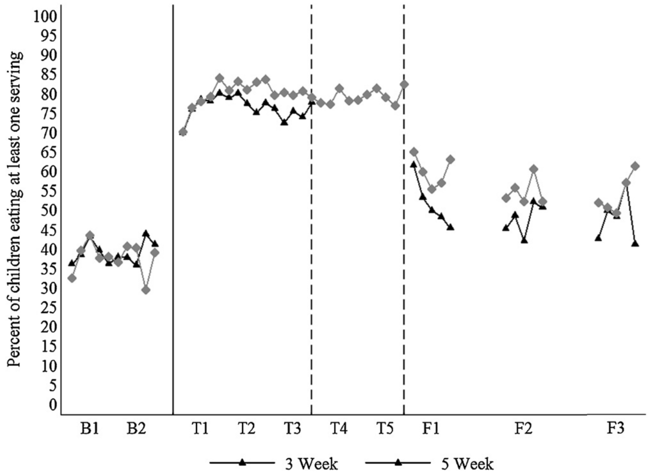

Fig. 2. Percentage of children eating at least one serving before and during the experiment. This figure provides the fraction of children eating at least one serving of fruits or vegetables each day during the three phases of the experiment. The solid vertical line indicates the start of the incentive period and the dashed vertical lines indicate the end of the 3- and 5-week incentive periods. F1, F2, and F3 refer to 1, 2, and 3 months following the end of the incentives. The post-incentive periods follow immediately after the end of the incentive period. Although this figure makes it appear there was a gap between T3 and F1 for the 3-week treatment, this is just to align the post-incentive periods for the two treatments to ease comparison.

separately for the two treatment groups. The results in this figure provide three general patterns. First, the incentives produced a very large change in the fraction of children eating fruits or vegetables during lunch (almost doubling the fraction eating at least one serving). Second, the high rates of fruit or vegetable consumption decreased after the end of the incentive period but remained at a level significantly higher than the baseline period. Third, the post-incentive fruit and vegetable consumption patterns were higher for schools assigned to the 5-week treatment group than those assigned to the 3-week treatment group.

Fig. 2 provides a more disaggregated look at the consumption rates during the baseline and incentive period. Since data were collected at each school for 10 days during the baseline period, we provide the average across all schools for these 10 days. We then provide the average consumption across the 15 days at the 3-week school and the 25 days at the 5-week schools. This graph shows that consumption rates during the baseline period were relatively stable followed by a large and immediate change once incentives were in place. There is also an incremental increase in consumption during the first few days of the incentive program suggesting either an increase in awareness or some students changing their mind about wanting to earn a reward (possibly through the influence of peers).

Table 2 provides a similar descriptive view of change in behavior in each of the treatment groups. We split the incentive period up into two periods (weeks 1–3 and weeks 4–5) to compare consumption rates during the same window of time for the two treatment groups. We provide a separate row for the last week of the incentive period (this row is not mutually exclusive of the other two rows). We also present the results for an alternative outcome measure, the number of servings of fruits and vegetables consumed per student. The results in this table provide the same insights as Fig. 1 but also show that the slightly higher consumption rates during the incentive period at the 5-week schools occurred even during the first 3 weeks (though the small difference during the incentive period is not statistically significant). The results in this table also show that the change in behavior was very similar over the course of the incentive period indicating that the effects of incentives did not fade out as they were left in place longer.

Both interventions significantly increased the fraction of students consuming at least one serving of fruits or vegetables (3-week intervention: 39.9% at baseline, average of 76.4% during intervention, p-value for difference <0.01; 5-week intervention: 37.6% at baseline, average of 79.5% during intervention p-value for difference <0.01). After the incentive period ended, the fraction of children eating at least one serving of fruits or vegetables decreased, but remained at a level about 10 percentage points above the baseline level in the 3-week intervention and 16.4 percentage points above baseline in the 5-week treatment (representing increases of 25.1% and 43.6%, respectively). These results indicate that the intervention did produce a meaningful change in post-incentive behavior for both treatment groups. This result is encouraging because it suggests that this type of incentive program did not have the type of 'crowding out of intrinsic motivation' effect that has been periodically raised as a concern when using incentives, or that such an effect, if present, was dominated by a stronger habit formation effect.

In Table 3 we provide the regression-based analog of the results in Table 2 that control for child, gender and grade and include day of week and school fixed effects. The regression-based estimates are very similar to the raw differences observed in Table 2, which is expected given the random assignment into treatment groups. For both incentive groups, we find that the increase in consumption was large and statistically significant both during the incentive period and up to 2 months after the end of the incentives.

Accounting for the persistent behavior during the 2 months after the incentive period also dramatically improves the cost effectiveness of the incentive program. Focusing on just the incentive period indicates that the intervention cost about 50 cents for each additional child induced to eat a serving of fruits or vegetables (52.1 at the 3-week schools and 47.9 at the 5-week schools). Once we include the additional consumption that occurs after the incentive period the cost per additional child eating a serving of fruits or vegetables drops to about 28 cents (29.0 cents at the 3-week schools and 28.4 at the 5-week schools). These estimates would have likely decreased even more if we had continued to measure

Table 2
Comparison of fraction of children eating a serving of fruits or vegetables by treatment group.

	Ate at least one serving		Amount of servings eaten	
	3 weeks	5 weeks	3 weeks	5 weeks
Baseline	39.9%	37.6%	0.577	0.602
Weeks 1–3 of intervention period	76.4%	79.9%	0.948	0.954
Weeks 4–5 of intervention period	–	79.0%	–	0.934
Last week of intervention period	75.4%	79.7%	0.906	0.927
1 month after	49.6%	58.5%	0.658	0.762
2 months after	48.1%	54.0%	0.648	0.716
N	191,719	212,203	191,719	212,203

Notes: The unit of observation is the student day. Both the 3- and 5-week interventions resulted in significant increases in fruit and vegetable consumption from baseline to the intervention period (p-values < 0.01). None of the differences between the 3- and 5-week treatments at any of the specific time intervals are statistically significant at the 0.10 level. The p-value for the difference for "1 month after" is 0.109 in the first two columns and 0.110 in the last two columns.

Table 3
Impact of incentives on behavior after incentives are removed.

	Ate at least one serving		Number of servings eaten	
	3 week	5 week	3 week	5 week
Incentive	0.368**	0.411**	0.370**	0.348**
	(0.025)	(0.036)	(0.027)	(0.047)
1 month post-intervention	0.117**	0.205**	0.103**	0.171**
	(0.023)	(0.038)	(0.030)	(0.045)
2 months post-intervention	0.103**	0.148**	0.090**	0.122**
	(0.023)	(0.032)	(0.029)	(0.031)
Grade	0.005	0.001	0.005	0.001
	(0.003)	(0.003)	(0.003)	(0.004)
Male	−0.060**	−0.049**	−0.085**	−0.072**
	(0.006)	(0.009)	(0.008)	(0.011)
N	191,719	212,203	191,719	212,203

Notes: The unit of analysis is the student day. The regressions include school and day of week fixed effects and controls for the child's grade and gender. Standard errors are clustered at the school level.
* Indicates statistical significance at the 5% level.
** Indicates statistical significance at the 1% level.

consumption data even longer than 2 months after the end of the incentive period.

It is possible that some of the persistence resulted from the fact that incentives were simply ceased when the incentive period ended, without an official announcement. It would not have taken students long, however, to realize that coins were no longer being handed out for fruit and vegetable consumption. When we rerun the analyses reported in Table 3 excluding the first 2 days of the post-intervention period, the results are virtually unchanged. Moreover, the fact that effects persist into the second month provides even stronger evidence that the post-intervention effect is not simply due to lack of awareness that incentives had ended.

Another possible factor that could bias the results is if students switched from eating bagged lunch to school lunch in response to treatment. This would produce a positive bias to results if those who switch already eat fruits and vegetables and would produce a negative bias if those who switch did not previously eat fruits and vegetables. This is a larger concern for differences for differences observed during the incentive period than those observed after the incentives have ended, which are the central focus of the paper. We observe some evidence that the incentives did lead some children to switch from bag lunches to school lunches since there was a 5% decline in school lunches on Fridays (from an average of 269 on weekdays to 255 on Fridays), which were the only days that students could earn incentives for eating fruits and vegetables from a sack lunch.

However, we see little evidence that this small effect led to a difference in fruit and vegetable consumption during the incentive period. In Appendix Table A1, we find that the results are very similar when we restrict our sample to just Fridays to the results when we exclude Fridays. Moreover, prior studies have found that sack lunches provide fewer fruits and vegetables than school lunches and often fail to include even one serving of fruits and vegetables (Rainville, 2001; Sweitzer et al., 2009), which suggests that those brought in to school lunches by the incentives may otherwise have consumed fewer fruits and vegetables.

In addition to examining whether providing incentives can continue to impact behavior after the incentives, we also test whether a longer incentive period leads to higher post-intervention fruit and vegetable consumption. Table 4 presents regressions that pool the two treatment groups together and estimate the interaction term between the length of the incentive and each of the periods of the study (e.g. incentive, 1-month follow-up, and 2-month follow-up). The coefficients on the interaction terms indicate the difference in the rate of persistence in the month following the intervention is statistically significant at only the 10% level though the difference between the two groups is about 8.7 percentage points. Both treatment groups though were significantly higher than their baseline rates. At 2 months post-intervention, we also find no clear evidence of greater effectiveness of the 5-week vs. the 3-week intervention in producing sustained effects (54.0% in 5-week vs. 48.1% in 3-week, *p*-value on this difference is 0.262).

Our study presents a common situation in randomized field experiments in which there is a very large sample but a much smaller number of randomization units. Various approaches have been developed to estimate appropriate standard errors that take into account the intra-class correlation between observations from the same randomizing unit. The standard errors reported in Table 4 have all been clustered at the school level. We also implemented a set of alternative approaches for calculating the standard errors used in past studies. These approaches include the cluster generalization of the wild bootstrap described by Cameron et al. (2008), the paired bootstrap method used by Prescott and Rockoff (2011), and the method of Generalized Estimating Equations (GEE) developed by Liang and Zeger (1986).

All of these alternative approaches provide very similar standard errors as the ones we report in Table 4. The standard errors using the GEE approach tend to produce the most precise estimates, and under this approach the difference between the 3- and 5-week schools 1 month after the end of the incentives would be statistically significant at the 5% level. If we did not cluster the standard errors at the school level, we would have standard errors that are about 10 times smaller. Future studies might consider ways in

Table 4
Impact of incentives on behavior after incentives are removed with interaction based on length of treatment period.

	Ate at least one serving	Number of servings eaten
Incentive	0.368**	0.369**
	(0.025)	(0.027)
1 month post-intervention	0.117**	0.104**
	(0.023)	(0.030)
2 months post-intervention	0.103**	0.090**
	(0.023)	(0.029)
5 week × incentive	0.042	−0.021
	(0.043)	(0.053)
5 week × month 1	0.087	0.067
	(0.044)	(0.053)
5 week × month 2	0.044	0.031
	(0.039)	(0.042)
Observations	403,922	403,922

Notes: Each regression includes controls for gender, grade, day of week, and school fixed effects. Standard errors are clustered at the school level
* Indicates statistical significance at the 5% level.
** Indicates statistical significance at the 1% level.

Table 5
Heterogeneous treatment effects.

	Male	Grade	PTA collected data	After new guidelines
Incentive	0.384**	0.430**	0.440**	0.432**
	(0.0213)	(0.0211)	(0.0268)	(0.0305)
1 month post	0.165**	0.184**	0.161**	0.215**
	(0.0240)	(0.0238)	(0.0269)	(0.0329)
2 months post	0.124**	0.137**	0.0900*	0.209**
	(0.0211)	(0.0207)	(0.0344)	(0.0309)
Incentive × *X*	0.00449	−0.0130**	−0.0906*	−0.0523
	(0.0103)	(0.00291)	(0.0386)	(0.0390)
1 month post × *X*	−0.0207*	−0.00863**	−0.0115	−0.0699
	(0.00882)	(0.00316)	(0.0421)	(0.0415)
2 months post × *X*	−0.00338	−0.00428	0.0317	−0.0955*
	(0.0107)	(0.00434)	(0.0431)	(0.0371)
N	403,922	403,922	403,922	403,922

Notes: The heading of each column indicates the variable that is included as an interaction term in the regression. The unit of analysis is the student day. The regressions include school and day of week fixed effects and controls for the child's grade and gender. The regressions for "PTA collected data" and "After new guidelines" also include a main effect for these measures as a control in the regression. Standard errors are clustered at the school level.
* Indicates statistical significance at the 5% level.
** Indicates statistical significance at the 1% level.

which they can increase the number of clusters even if it requires reducing the overall sample. Some options might include having just one or two grades per school participate or possibly restricting the sample to smaller schools.

A final approach to statistical inference in this situation is to use permutation inference. Under the null hypothesis that there is no difference in follow-up period between 3-week and 5-week schools, mislabeling schools as a 3- or 5-week school would have no effect on the estimated coefficient of the interaction terms in Table 4. We randomly generated 10,000 permutations of the labeling across the 40 schools (holding constant the number of schools assigned the label of 5-week schools) and estimated the same model as Table 4. We find that only 246 of these permutations had a coefficient on the interaction between the 5-week treatment and 1-month follow-up period that was larger than the coefficient reported in Table 4. This provides a *p*-value for the two-sided test between 5-week and 3-week treatments of 0.049. In addition, the permutations that had the highest estimated coefficients were those in which the highest fraction of schools were assigned the correct label providing additional evidence that the 5-week treatment actually had a larger effect after the end of the incentive period.[3]

In Table 5, we tests for heterogeneous treatment effects by pooling the 3- and 5-week treatments together and adding terms to interacting specific variables with the treatments and follow up timing variables. We find that while boys and girls experienced almost an identical percentage point increase in fruit and vegetable consumption during the incentive period, boys had slightly lower levels of habit formation. We also find that younger children responded more to the incentives and experienced slightly higher levels of habit formation after the incentives ended, although this difference was only statistically significant 1 month after the end of the incentives.

At some of the schools the data was collected by college research assistants and at others the data was collected by PTA parents. Since PTA parents were more likely to have been familiar to students, this difference might have an impact on social distance, which could have in turn affected measured persistence if students felt more pressure to eat fruits and vegetables when monitored by someone they knew. To test for this we included a specification reported in the column marked "PTA" adding interaction terms for whether the data collection was performed by members of the PTA. We find that the change in behavior during the incentive period was larger at the schools that had RAs instead of PTA parents (contrary to what one might expect based on social distance), but that the habit formation coefficients were very similar.

Finally, Table 5 includes a column which incorporates an interaction term to test for any differences in the effects of the treatment before and after the new lunch guidelines went into effect. None of the interactions are significant at the 0.05 level, though the interaction ($p < 0.10$) for effects 2 months after the incentives were withdrawn suggests that the longer-term carry-over effect might have been smaller after the change in guidelines.

4. Discussion

The results of this paper are based on a large field experiment at 40 elementary schools in which children received a small incentive for consuming fruits and vegetables as part of their school-provided lunches. We find that these small incentives produced a dramatic increase in fruit and vegetable consumption during the incentive period and that this change in behavior was sustained for at least 2 months after the incentives ended. We also find suggestive evidence that a longer intervention period produced a more sustained response once the rewards were removed.

One question raised by this and other related studies is the mechanism that led to behavior persistence once incentives were removed. At least three mechanisms are possible. One, which we have labeled 'classic' habit formation, suggests that students became used to eating fruits and vegetables during lunch and this became an automatic pattern of behavior. A second, informational, mechanism is that consuming the fruits and vegetables may have led to either re-discovery of pre-existing tastes, a development of a new taste for a food to which a particular student may have had limited exposure, or a change in tastes, the latter consistent with prior research which shows that repeated exposure to specific items can influence an individual's food preferences (Birch & Marlin, 1982). The third mechanism is that making fruit and vegetable consumption more 'popular' (albeit with the help of an incentive) may have shifted social norms around fruit and vegetable consumption such that kids would be less likely to cast aspersions on other kids who ate fruit and vegetables at lunch. As noted, if the

[3] Heckman et al. (2010) use a permutation-based inference approach in their examination of the Perry Preschool Program and highlight some of the advantages of this approach.

longer incentive period produced greater persistence, this would tend to support the classic interpretation of habits, as well, perhaps, as the norm account (if norms take time to form). Given that the difference in persistence between the 3 week and 5 interventions is marginal, at best, the results are somewhat agnostic concerning the underlying mechanism leading to persistence. Regardless of which mechanism or combination of mechanisms is at work, however, we did observe a sustained post-intervention increases in fruit and vegetable consumption at a modest cost.

One artefactual mechanism that could explain behavior persistence in our study was the presence of data collectors in the lunchroom after the incentive was removed. Since they were the ones who distributed tokens to the students and encouraged them to eat their fruits and vegetables for several weeks, perceived social pressure on the part of the student may be enough to induce them to continue to eat fruits and vegetables. An important piece of evidence that points against this mechanism is the results of Just and Price (2013) who implemented a program very similar to ours but left the rewards in place for only 5 days. The study also had data collectors in the cafeteria after the end of the incentive period and found that consumption rates went back to their baseline levels after the incentive period ended, suggesting that the presence of the data collectors alone isn't sufficient to create the false appearance of habit formation.

None of the treatments showed any evidence that the incentives were crowding out intrinsic motivation, which would be demonstrated by the consumption of fruits and vegetables during the post-incentive period dropping to levels lower than pre-treatment levels. One concern, raised by psychologists Deci et al. (1999) has been that the use of extrinsic motivators to change behavior in children diminishes their intrinsic motivations for positive behavior. However, others have questioned the robustness of their results and noted their vulnerability to alternative interpretation (Cameron & Pierce, 1994). In the current experiment, we cannot rule out the possibility that such effects occurred but the fact that we observed persistence of a portion of the initial intervention effect instead of reduction in fruit and vegetables below baseline consumption, suggests, at a minimum, that any such effects were exceeded by the influence of habit formation.

Results from our study reinforce those from earlier research showing that the use of small incentives is an effective way to encourage children to eat more fruits and vegetables (Belot et al., 2013) and that these induced changes in behavior persist after the incentives are no longer being offered (List and Samek, 2015). We also find suggestive evidence that longer intervention periods lead to greater persistence of behavior change. While the habit formation process that we observe in this study may be most germane to food choices among children, there are many other positive health behaviors for which sustaining a period of active involvement can result in the behavioral change persisting even after the incentive is removed and where an approach similar to the intervention described here could be effective.

Acknowledgments

The authors are grateful to the USDA Economic Research Service for research funding (grant # 59-5000-1-0034) and for the many BYU undergraduates and PTA parents that helped as research assistants on the project. This paper has benefitted from the helpful comments by Sandy Black, Dan Hamermesh, Patrick Nolen, Devin Pope, Bob Slonim, Coady Wing, Jingsan Zhu and seminar participants at the Cornell University, Harris School, University of Illinois at Chicago, University of Sydney, UT-Austin, Iowa State, University of Utah, University of Oregon, University of Iowa, and Brigham Young University.

Appendix A.

Table A1
Impact of incentives on behavior after incentives are removed, by day of the week.

	Ate at least one serving			
	Friday		Not Friday	
	3 week	5 week	3 week	5 week
Incentive	0.392**	0.446**	0.362**	0.402**
	(0.031)	(0.040)	(0.026)	(0.038)
1 month post-intervention	0.123**	0.214**	0.116**	0.201**
	(0.023)	(0.053)	(0.025)	(0.037)
2 months post-intervention	0.102*	0.145**	0.103**	0.148**
	(0.047)	(0.041)	(0.022)	(0.034)
Grade	0.009*	0.006	0.004	0.000
	(0.004)	(0.003)	(0.003)	(0.003)
Male	−0.059**	−0.046**	−0.060**	−0.050**
	(0.009)	(0.009)	(0.006)	(0.009)
Observations	34,551	38,580	157,168	173,623

Notes: Each regression includes controls for gender, grade, day of week, and school fixed effects. Standard errors are clustered at the school level.
* Indicates statistical significance at the 5% level.
** Indicates statistical significance at the 1% level.

References

Acland, D., Levy, M., 2015. Naiveté, Projection bias, and habit formation in gym attendance. Manag. Sci. 61 (1), 146–160.
Belot, M., James, J., Nolen, P., 2013. Changing eating habits—a field experiment in primary schools. University of Edinburgh (Working paper), Edinburgh, Scotland.
Birch, L.L., Marlin, D.W., 1982. I don't like it; I never tried it: effects of exposure on two-year-old children's food preferences. Appetite 3 (4), 353–360.
Cameron, A.C., Gelbach, J.B., Miller, D.L., 2008. Bootstrap-based improvements for inference with clustered errors. Rev. Econ. Stat. 90 (3), 414–427.
Cameron, J., Pierce, W.D., 1994. Reinforcement, reward, and intrinsic motivation: a meta-analysis. Rev. Educ. Res. 64 (3), 363–423.
Charness, G., Gneezy, U., 2009. Incentives to exercise. Econometrica 77 (3), 909–931.
Deci, E.L., Koestner, R., Ryan, R.M., 1999. A meta-analytic review of experiments examining the effect of extrinsic rewards on intrinsic motivation. Psychol. Bull. 125 (6), 627–668.
Frey, B., Jegen, R., 2001. Motivational Interactions: effects on behavior. Ann. Econ. Stat. 63, 131–153.
Heckman, J., Moon, S., Pinto, R., Savelyev, P., Yavitz, A., 2010. Analyzing social experiments as implemented: a reexamination of the evidence from the HighScope Perry Preschool program. Quant. Econ. 1 (1), 1–46.
Hung, H., Joshipura, J., Jiang, R., Hu, F., Hunter, D., Smith-Warner, S., Colditz, G., Rosner, B., Spiegelman, D., Willett, W., 2004. Fruit and vegetable intake and risk of major chronic disease. J. Natl. Cancer Inst. 96 (21), 1577–1584.
Just, D., Price, J., 2013. Using incentives to encourage healthy eating in children. J. Hum. Res. 48 (4), 885-872.
John, L., Loewenstein, G., Troxel, A., Norton, L., Fassbender, J., Volpp, K., 2011. Financial incentives for extended weight loss: a randomized, controlled trial. J. Gen. Inter. Med. 26 (6), 621–626.
Kant, A., 2004. Dietary patterns and health outcomes. J. Am. Diet. Assoc. 104 (4), 615–635.
Liang, K.Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. Biometrika 73 (1), 13–22.
List, J.A., Samek, S., 2015. The behavioralist as nutritionist: leveraging behavioral economics to improve child food choice and consumption. J. Health Econ. 39, 135–146.
Prescott, J., Rockoff, J., 2011. Do sex offender registration and notification laws affect criminal behavior? J. Law Econ. 53, 161–206.
Rainville, A., 2001. Nutritional quality of reimbursable school lunches compared to lunches brought from home in elementary schools in two Southeastern Michigan districts. J. Child Nutr. Manag. 25 (1), 13–18.
Royer, H., Stehr, M.F., Sydnor, J.R., 2015. Incentives, commitments and habit formation in exercise: evidence from a field experiment with workers at a Fortune-500 company. Am. Econ. J. Appl. Econ. 7 (3), 51–84.
Schofield, H., Loewenstein, G., Kopsic, J., Volpp, K.G., 2015. Comparing the effectiveness of individualistic, altruistic, and competitive incentives in motivating completion of mental exercises. J. Health Econ. 44, 286–299.
Sweitzer, Sara, Briley, M., Robert-Gray, C., 2009. Do sack lunches provided by parents meet the nutritional needs of young children who attend child care? J. Am. Diet. Assoc. 109 (1), 141–144.
Volpp, K., John, L., Troxel, A., Norton, L., Fassbender, J., Lowenstein, G., 2008. Financial incentive–based approaches for weight loss: a randomized trial. J. Am. Med. Assoc. 300, 2631–2637.
Volpp, K.G., Troxel, A.B., Pauly, M.V., Glick, H.A., Puig, A., Asch, D.A., Galvin, R., Zhu, J., Wan, F., DeGuzman, J., Corbett, E., Weiner, J., Audrain-McGovern, J., 2009. A randomized, controlled trial of financial incentives for smoking cessation. N. Engl. J. Med. 360 (7), 699–709.