Behavior change through field experiments: Evidence from studies on healthy eating and exercise

Eric M. VanEpps
The problem

People want to make healthy decisions but struggle with self-control in the moment
Advance Ordering for Healthier Eating

With Julie Downs and George Loewenstein

(Journal of Marketing Research, 2016)
Advance ordering

Pre-commitment:
Advance orders will have fewer calories
Advance orders ➔ Healthier choices

Past empirical work:
• **Online grocery orders** *(Milkman, Rogers, & Bazerman 2010)*
• **Snack choices** *(Read & van Leeuwen 1998; Bucher-Koenen & Schmidt 2011)*
• **School lunch choices** *(Hanks, Just, & Wansink 2012)*

Present work:
• Delay of minutes/hours, not days/weeks
• Hold decision context constant
• Real-world meal orders
Research outline

Study 1 (University Experiment)
• Advance vs. immediate orders

Study 2 (Field Study/Observational)
• Natural variation in length of delay

Study 3 (Field Experiment)
• Advance vs. lunchtime orders
University study: Participants (N = 195)

- CMU students (undergraduate & graduate)
  - 42% female
  - Median age of 25 (range 19-54)
  - 27% White/Caucasian, 62% Asian
- Classes ended between 11:30 am and 12:30 pm
- Participants agreed to complete 3 surveys: one before class, one after class, and one after lunch
Procedure

Advance: Class

Lunch Order → Dem Survey → Receive Lunch → Exit Survey

Lunchtime: Class

Dem Survey → Lunch Order → Receive Lunch → Exit Survey
Lunch options

• ½ Sandwich (Maple Turkey Club, Turkey & Muenster, Caprese)

• Side 1 (Creamy Potato Salad, Homemade Coleslaw, Garden Veggie Pasta Salad)

• Side 2 (Chocolate Chip Cookie, Banana, Apple, Orange)

• Beverage (Coca-Cola, Capri Sun, Bottled Water)
Calories ordered

$\text{Advance}$  
$t(193) = 2.75, p < .01$; Error Bars +/- 1 SE

$\text{Lunchtime}$
Calories consumed

Possible concern
• May eat different proportions by condition; care about consumption

Measure
• Asked participants in exit survey what proportion of each item they ate
• Calculate total proportion of calories consumed

Takeaway
• No evidence of compensatory consumption

\[ t(149) = 0.23, p = .82; \text{ Error Bars +/- 1 SE} \]
Meal satisfaction

Possible concern
• People only ate less with advance order because they hated lunch order

Measure
• Participants rated meal satisfaction in exit survey

Takeaway
• No significant decline in meal satisfaction when lunch ordered in advance

$ t(157) = .99, p = .33; $ Error Bars +/- 1 SE
Does hunger mediate effect?

\[ \beta = -1.43^{***} \]

\[ \beta = 28.9 \text{ cals}^* \]

\[ \beta = -101.5 \text{ cals}^{**} \]

\[ (\beta = -76.7 \text{ cals}^{†}) \]

Indirect \( \beta = -24.9 \text{ cals} \), 95% CI \([-65.9, 16.2]\)
• Orders placed in advance had 100 fewer calories
• Effect carried through to consumption
• No differences in meal satisfaction
• Hunger alone cannot explain this effect

What happens in a more natural ordering environment?
Field study

- Large corporate headquarters
- On-site cafeteria with sandwiches, salads, wraps, and packaged snacks/drinks
- Created new online ordering system
- Offered as 4-week trial period
- $3 discount on each meal
Field study: Participants (N = 394)

- Full-time Humana employees
  - 58% female
  - 84% white
  - Median age: 39 (range 22-71)
  - 32% on a weight-loss diet
  - Median BMI: 26 (range 16-55)
  - Median household income: $75-$100,000
Field study procedure

- Placed orders between 7am & 1:30pm
- ~30 minutes between placement & pick-up
- Orders picked up between 11am & 2pm
Field study procedure

7:00am Site opens
8:30am Order placed
11:30am Order picked up
2:00pm Lunch period over

Length of Delay
### Calories ordered

<table>
<thead>
<tr>
<th>Model 1</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Length of Delay</strong></td>
<td><strong>-31.49</strong></td>
<td><strong>(11.81)</strong></td>
</tr>
<tr>
<td>(hours)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Intercept</strong></td>
<td><strong>654.99</strong></td>
<td><strong>(26.47)</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Observations</strong></td>
<td>1389</td>
<td></td>
</tr>
<tr>
<td><strong>Participants</strong></td>
<td>394</td>
<td></td>
</tr>
</tbody>
</table>
Field experiment (N = 296)

Random Assignment

Weeks 1 & 2
Advance: 7am – 10am
Lunchtime: 11am – 1pm

Weeks 3 & 4
Advance: 7am – 10am
Lunchtime: 11am – 1pm
Field experiment: Calories ordered

(568 vs. 598), $p = .09$; Fixed effects (572 vs. 603), $p = .07$
Advance ordering summary

• Calorie reduction for advance orders
  • Large benefit in university setting
    • 100 calories, 10% of meal calories
    • Compared advance to immediate orders
  • Modest benefit in field setting
    • 30-40 calories, 5-6% of meal calories
    • Compared short delay to longer delay
What makes a study valid?

There are two main aspects to research:

Internal validity (identifying cause and effect)

– Can we say that variable X caused this outcome?

External validity (generalizability)

– Who are the subjects?
– What is the context of the decision/test?
– What is the setting?
– Is this result applicable to the real world?
Three broad classes of empirical studies

Observational (correlational) studies
- Is the number of children in a family related to marital success/divorce rates?

Quasi-experiments
- Did the removal of the death penalty in one state increase rate of violent crimes? (can compare to a state where death penalty remains, but not random assignment)

Experiments
- “Eliminate all other factors, and the one that remains must be the truth.”
- Considered the best way to determine causal relationship
Why Random Experiments?

Example: We want to know if morning yoga improves productivity.
Yoga and Productivity

Suppose we don’t randomly assign, but instead assign by gender. No matter what happens, we cannot conclude whether incentives increase achievement.
Yoga and Productivity

Suppose we don’t randomly assign, but instead just observe who naturally shows up for yoga.

No matter what happens, we cannot conclude whether yoga increases productivity, or whether those who attend are different in other ways from those who don’t attend.
Yoga and Productivity

Suppose we randomly assign (using a random number generator, for instance) invitations to yoga. The groups, before the intervention are, on average, equal. We can rule out all alternative explanations.
The importance of random assignment cannot be overstated.

It is the best method to establish causality.
What makes a study valid?

There are two main aspects to research:

Internal validity (identifying cause and effect)
- Can we say that variable X caused this outcome?

External validity (generalizability)
- Who are the subjects?
- What is the context of the decision/test?
- What is the setting?
- Is this result applicable to the real world?

Field experiments are considered the gold standard
Goal-Setting for Employee Wellness

Study in collaboration with University of Utah Auxiliary Wellness Challenge
Program Structure

• Employees from across the university can enroll in a wellness challenge that runs through the spring semester
• Earn points for completing wellness activities, points give entries into drawings for prizes
• Weekly emails with health tips and updates
Earning Points

- 30 minutes/day of exercise = 5 points
- 64 oz of water in a day = 5 points
- 7+ hours of sleep = 5 points
- 10 points per high-intensity work-out (running, CrossFit, cycling, intramural sport)
- 5 points for 7,500 steps in a day; 10 points for 10,000 steps in a day
- Team points available for monthly challenges
Research Question

• Do participants perform better when they have a points target/goal in mind?
• How should the goal be selected to make it most motivating?
Past Research

- Proximity to a goal (goals that are “within reach”) increases motivation (Hull, 1932; Kivetz, Urminsky, & Zheng, 2006)
- Yet people often choose more difficult goals for themselves, either to try to “commit” themselves to higher achievement, or out of biased predictions regarding what they can achieve (VanEpps et al., working paper)
- Relatively little is known about whether people care more about goals when set by someone else or when set by the self
After first few weeks in study, we had a sense of how many points per week people would earn, on average. Therefore, we could use that as a baseline to set new goals to motivate increased performance. We sent participants an email, and randomly assigned them to either receive a goal from us or to pick their own goal. We then tracked their points earned for the remainder of the wellness challenge.
Hi John Doe,

We’ve calculated your average personal activity (e.g., sleeping 7 hours per night, exercising each day) points from the first four weeks, and see that you have averaged [40] points per week. That’s great! But we think that’s only the baseline of what you can achieve, and we’d like to encourage you to do even more in the weeks ahead.

Going forward, we’d like for you to pursue a goal to increase your points per week.

If randomly assigned to an assigned goal:
Specifically, we want you to aim to record [55] points per week.

If randomly assigned to set their own goal:
Specifically, we want you to select a goal for the number of points you’ll record each week. Select from the options below.

[50 ; 55 ; 60]

Reaching this goal may be challenging, but we know you can do it. Good luck meeting your new goal this week and in the weeks ahead, and thanks for your participation!
Results

Points Earned

Chose Own Goal (n = 66)

Assigned Goal (n = 65)

$t(129) = 1.45$, $p = .15$; error bars +/- 1 SE
So what studies will you run?

Points to keep in mind:

1. Are you already collecting data? Tracking exercise, conducting surveys, measuring attendance at wellness activities?

2. Do you have multiple ideas about the best way to implement a given program? If so, consider testing both approaches

And remember, academics are always looking for real-world data for collaborations! You have a resource that people in my field go nuts for!
Behavior change through field experiments: Evidence from studies on healthy eating and exercise

Eric M. VanEpps
SUPPLEMENTAL SLIDES
Numeric and Traffic Light Calorie Labels for Online Orders

With Julie Downs and George Loewenstein

(Journal of Public Policy & Marketing, 2016)
Nutrition information at point of purchase

• Limited information processing capacity (Bettman 1979; Malhotra 1982, 1984; Jacoby 1984)

• Calorie labels simplify nutrition to single attribute
Numeric calorie labeling

- First implemented in New York City in 2007
- Mandated by Patient Protection and Affordable Care Act
- Mixed effectiveness *(Swartz, Braxton, & Viera 2011; Kiszko et al. 2014; VanEpps et al. 2016)*
Empirical Tests

- Reduced purchase of red-light items in hospital cafeteria (Levy et al., 2012)
- Combined with numeric labels to reduce calorie intake at full-service restaurant (Ellison et al., 2013)
Numeric vs. Traffic light calorie labels

Numeric calorie labels

• Descriptive information
• Precise, allows for exact calculations
• Requires facility with numbers (Burton & Kees 2012; Hawley et al. 2013; Rothman et al. 2006)

Traffic light calorie labels

• Prescriptive information
• Easier to interpret (Andrews et al. 2014; Hawley et al. 2013; Hersey et al. 2013)
Participants (N = 249)

• Full-time Humana employees
  • 61% female
  • Median age: 40 (range 22-67)
  • 82% white
  • Median BMI: 26 (range 16-55)
  • 34% on weight loss diet
  • Median household income $75-$100,000
Procedure

- Order lunches online over 4 weeks
- Menu condition randomly assigned for full 4-week period (2x2 design):
  - Numeric calorie label (absent or present)
  - Traffic light calorie label (absent or present)
Calorie Labels p<.05; Light Labels p<.05; Interaction p=0.15; Error Bars +/- 1 SE
Patterns remain when model includes covariates
Does numeracy matter?

* = different from Control at $p<.05$; Error Bars +/- 1 SE
• Calorie labels help consumers order 10% fewer calories for online lunch orders
  • Effect similar in magnitude for numeric, traffic light, and combined labels
• Label formats have different effects on numerate and non-numerate consumers
Field Study: Distribution of Time

![Histogram showing the distribution of time delays. The x-axis represents time delay in hours, ranging from 0 to 5. The y-axis represents density, ranging from 0 to 0.8. The histogram peaks around the 0 to 1 hour range and gradually decreases as the time delay increases.]
Field Study: Calories ordered