Trapped by the present: Estimating long-term impact from A/B experiments

Strata San Jose  Brian Karfunkel  March 8, 2018
Experiments help make decisions

But “decisions are made by those who show up”
Core users show up.

They use the product, so they’re in the experiment data.
Unengaged users are busy

So they’re not in the data

(as much)

(for most in-product experiments)
When we make decisions based (mostly) on the most active users, we risk building a product for the users you have instead of for the users you want to have.
In other words: Engagement bias
Taking a more long-term view of experiments mitigates engagement bias and helps prioritize efforts to engage non-core users.
Help people **discover** and **do** the things they **love**.
Cravings
Omar Seyal

The perfect path to cold brew

Pin
A bookmark someone has saved from the internet.
The perfect path to cold brew
The perfect path to cold brew
Caffeinated Inc.
The perfect path
to cold brew
Caffeinated Inc.
Omar Seyal
Cravings
People on Pinterest each month: 200m+

- 100b+ Pins
- 2b+ Boards
- 10b+ Recommendations/Day
- 450+ Engineers
1. Experiments and decisions
2. Biased population
3. Biased lift
4. Biased impact
5. Predicting the future
How experiments power decision-making
Home feed: a grid of Pins
Closeup experiment:
Test if removing attribution increases engagement
Control
With attribution

Brenna saved to Drinks
Delicious summer cocktail.

Enabled
No attribution

Brenna saved to Drinks
Delicious summer cocktail.
Closeup experiment

**Targeting:** Who is eligible to join?

**Triggering:** What actions do users have to take to join?

**Treatment:** How does enabled differ from control?
Targeting: All users
Triggering: DAU and close up on Pin
Treatment: Removing attribution affordance

Closeup experiment
Why run an experiment at all?

What do we want to know?
1. Is **enabled** better than **control**?

- Use statistics to test hypothesis that a given **metric** is higher/lower for **enabled** than **control**

**Propensity metrics like:**
- share of users who save (savers)
- share of users who close up
- share of users who are active (DAU)
- share of users who are active on iOS

**Volume (avg. per user) metrics like:**
- number of saves
- number of pin impressions
- number of closeups
1. Is enabled better than control?

- Use statistics to test hypothesis that a given metric is higher/lower for enabled than control
- “This treatment made users more likely to save a Pin”
1. **Is enabled better than control?**

- Use statistics to test hypothesis that a given **metric** is higher/lower for **enabled** than **control**
- “*This treatment made users* (in this experiment) *more likely to save a Pin*”
- If **enabled** > **control**, ship **enabled**
2. How much better is enabled?

- Estimate the **lift**: enabled minus control

- "This treatment made users (in this experiment) 3% more likely to save"

- If \( \text{lift} > X \), then we (iterate | stop iterating | ship despite some negative effects)
3. How much should we care?

- Estimate impact: \( \text{(lift)} \times \text{(users affected)} \)

- "This treatment will create 30k more savers over 12 weeks."

- If impact is higher than for other (workstreams | teams), then increase resources
lift
population lift
population \times \text{lift} = \text{impact}
Where engagement bias comes in: population
Levels of engagement: All users

- Core (very active)
- Casual (fairly active)
- Marginal (barely active)

Each block is 1,000 users.

Note: for illustration purposes only, not representative of Pinterest user base.
Levels of engagement: Users active on given day (DAU)

Core

Casual

Marginal
Levels of engagement: Closeuppers on a given day

<table>
<thead>
<tr>
<th>Core</th>
<th>Casual</th>
<th>Marginal</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Core" /></td>
<td><img src="image2" alt="Casual" /></td>
<td><img src="image3" alt="Marginal" /></td>
</tr>
<tr>
<td><img src="image1" alt="Core" /></td>
<td><img src="image2" alt="Casual" /></td>
<td><img src="image3" alt="Marginal" /></td>
</tr>
<tr>
<td><img src="image1" alt="Core" /></td>
<td><img src="image2" alt="Casual" /></td>
<td><img src="image3" alt="Marginal" /></td>
</tr>
<tr>
<td><img src="image1" alt="Core" /></td>
<td><img src="image2" alt="Casual" /></td>
<td><img src="image3" alt="Marginal" /></td>
</tr>
<tr>
<td><img src="image1" alt="Core" /></td>
<td><img src="image2" alt="Casual" /></td>
<td><img src="image3" alt="Marginal" /></td>
</tr>
</tbody>
</table>
Jan. 1 — Experiment launched

Users in closeup experiment:

Core    Casual    Marginal

[Diagram with user icons for each category]
<table>
<thead>
<tr>
<th>Levels of engagement</th>
<th>Core</th>
<th>Casual</th>
<th>Marginal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Closeuppers on a given day</td>
<td><img src="image1" alt="Core Users" /></td>
<td><img src="image2" alt="Casual Users" /></td>
<td><img src="image3" alt="Marginal Users" /></td>
</tr>
</tbody>
</table>
Jan. 1 — Experiment launched

<table>
<thead>
<tr>
<th>Users in closeup experiment</th>
<th>Core</th>
<th>Casual</th>
<th>Marginal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="image" alt="Users" /></td>
<td><img src="image" alt="Users" /></td>
<td><img src="image" alt="Users" /></td>
</tr>
<tr>
<td>Core</td>
<td>Casual</td>
<td>Marginal</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="Core User" /></td>
<td><img src="image" alt="Casual User" /></td>
<td><img src="image" alt="Marginal User" /></td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="Core User" /></td>
<td><img src="image" alt="Casual User" /></td>
<td><img src="image" alt="Marginal User" /></td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="Core User" /></td>
<td><img src="image" alt="Casual User" /></td>
<td><img src="image" alt="Marginal User" /></td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="Core User" /></td>
<td><img src="image" alt="Casual User" /></td>
<td><img src="image" alt="Marginal User" /></td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="Core User" /></td>
<td><img src="image" alt="Casual User" /></td>
<td><img src="image" alt="Marginal User" /></td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="Core User" /></td>
<td><img src="image" alt="Casual User" /></td>
<td><img src="image" alt="Marginal User" /></td>
<td></td>
</tr>
<tr>
<td>Users in closeup experiment:</td>
<td>Jan. 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core</td>
<td>Casual</td>
<td>Marginal</td>
<td></td>
</tr>
<tr>
<td><img src="image1" alt="Core users" /></td>
<td><img src="image2" alt="Casual users" /></td>
<td><img src="image3" alt="Marginal users" /></td>
<td></td>
</tr>
<tr>
<td><img src="image1" alt="Core users" /></td>
<td><img src="image2" alt="Casual users" /></td>
<td><img src="image3" alt="Marginal users" /></td>
<td></td>
</tr>
<tr>
<td><img src="image1" alt="Core users" /></td>
<td><img src="image2" alt="Casual users" /></td>
<td><img src="image3" alt="Marginal users" /></td>
<td></td>
</tr>
<tr>
<td><img src="image1" alt="Core users" /></td>
<td><img src="image2" alt="Casual users" /></td>
<td><img src="image3" alt="Marginal users" /></td>
<td></td>
</tr>
<tr>
<td><img src="image1" alt="Core users" /></td>
<td><img src="image2" alt="Casual users" /></td>
<td><img src="image3" alt="Marginal users" /></td>
<td></td>
</tr>
<tr>
<td><img src="image1" alt="Core users" /></td>
<td><img src="image2" alt="Casual users" /></td>
<td><img src="image3" alt="Marginal users" /></td>
<td></td>
</tr>
<tr>
<td><img src="image1" alt="Core users" /></td>
<td><img src="image2" alt="Casual users" /></td>
<td><img src="image3" alt="Marginal users" /></td>
<td></td>
</tr>
<tr>
<td><img src="image1" alt="Core users" /></td>
<td><img src="image2" alt="Casual users" /></td>
<td><img src="image3" alt="Marginal users" /></td>
<td></td>
</tr>
<tr>
<td><img src="image1" alt="Core users" /></td>
<td><img src="image2" alt="Casual users" /></td>
<td><img src="image3" alt="Marginal users" /></td>
<td></td>
</tr>
<tr>
<td>Core</td>
<td>Casual</td>
<td>Marginal</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td><img src="image1" alt="Core users" /></td>
<td><img src="image2" alt="Casual users" /></td>
<td><img src="image3" alt="Marginal users" /></td>
<td></td>
</tr>
</tbody>
</table>

**Users in closeup experiment**

**Jan. 5**
Closeup experiment

Users in experiment, by engagement level

Users

0
5,000
10,000
15,000
20,000

Date
1/1
1/2
1/3
1/4
1/5
1/6
1/7
1/8

Core
Casual
Marginal
Closeup experiment

Share of users in each engagement level, as of a given day

- Core
- Casual
- Marginal
<table>
<thead>
<tr>
<th></th>
<th>Core</th>
<th>Casual</th>
<th>Marginal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Users in closeup experiment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jan. 5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Core
Casual
Marginal

Users in close-up experiment: Jan. 6
Users in closeup experiment: Jan. 7

Core | Casual | Marginal

The diagram shows the distribution of users across different categories. The number of users in each category is represented by the number of icons in each section.
Users in closeup experiment: Jan. 8

<table>
<thead>
<tr>
<th>Core</th>
<th>Casual</th>
<th>Marginal</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Core users" /></td>
<td><img src="image2" alt="Casual users" /></td>
<td><img src="image3" alt="Marginal users" /></td>
</tr>
</tbody>
</table>

- **Core**: Highest engagement users.
- **Casual**: Average engagement users.
- **Marginal**: Lowest engagement users.
Closeup experiment

Users in experiment, by engagement level

<table>
<thead>
<tr>
<th>Date</th>
<th>Core</th>
<th>Casual</th>
<th>Marginal</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Closeup experiment

Users in experiment, by engagement level

- Core
- Casual
- Marginal
- Total
Closeup experiment

Share of users in each engagement level, as of a given day

Date
1/1 1/8 1/15 1/22 1/28

Share of users in experiment
0% 10% 20% 30% 40% 50% 60%

Core  Casual  Marginal
Closeup experiment

Share of users in each engagement level, as of a given day
Where engagement bias comes in: **lift**
Closeup experiment: For every 100 core users...

control

enabled
Closeup experiment: For every 100 core users...

40% of control users save a pin
Closeup experiment: For every 100 core users...

- 40% of control users save a pin
- 42% of enabled users save a pin

**Absolute lift:**
42% - 40% = 2%

**Relative lift:**
(2 pct. pt.)/40% = 5%
If we shipped the enabled variant, for every 100 core users who see the new closeup we gain 2 savers.
Closeup experiment: Lift in savers

**Core**
For every 100 core users, the enabled group has 2 more savers than control

= 2% absolute lift

**Casual**
For every 100 casual users, the enabled group has 1 more saver than control

= 1% absolute lift

**Marginal**
For every 100 marginal users, the enabled group has 0 more savers than control

= 0% absolute lift
Closeup experiment: Impact from shipping treatment

Engagement distribution of users in exposed to new closeup

Absolute lift in savers

Date

Date
Closeup experiment lift

Changing population means changing overall lift

![Graph showing absolute lift over time with dates 1/1, 1/8, 1/15, 1/22, 1/28. The graph indicates a decline in absolute lift from 2% to 0% over the period. The graph is labeled with Core, Causal, Marginal, and Total categories.]
Closeup experiment lift

Changing population means changing overall lift
Where engagement bias comes in: impact
What we want:

If I ship this experience today, how many (clicks | impressions | DAUs) will I have created (4/13/52 weeks from now | at the end of the quarter/half/year)?
population \times \text{lift} = \text{impact}
Closeup experiment: Impact from shipping treatment

Users exposed to new closeup

- Core
- Casual
- Marginal
- Total

Absolute lift in savers

- Lift
- Date

(population × lift)
Closeup experiment impact

Additional savers created

![Graph showing the number of savers over time with categories: Core, Causal, Marginal, and Total.](image-url)
Problem:

Engagement bias differs from one experiment to the next.
Activation experiment:
Test if suggesting searches activates users who have not searched in last 4 weeks
Control
Normal home feed

Enabled
Search suggestion
Activation experiment

**Targeting:** Users who haven’t searched in 4+ weeks

**Triggering:** DAU and view home feed

**Treatment:** Insert search suggestion in home feed
<table>
<thead>
<tr>
<th>Levels of engagement</th>
<th>Users active on given day (DAU)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core</td>
<td><img src="image1" alt="Core Users" /></td>
</tr>
<tr>
<td>Casual</td>
<td><img src="image2" alt="Casual Users" /></td>
</tr>
<tr>
<td>Marginal</td>
<td><img src="image3" alt="Marginal Users" /></td>
</tr>
</tbody>
</table>
Levels of engagement:

DAUs who rarely search

Core  Casual  Marginal
Users in activation experiment: Jan. 1

- Core
- Casual
- Marginal
Users in activation experiment: Jan. 2

**Core**

**Casual**

**Marginal**

- Users in the core category are represented by red icons.
- Users in the casual category are represented by purple icons.
- Users in the marginal category are represented by gray icons.
Users in activation experiment: Jan. 3

Core  Casual  Marginal
Users in activation experiment: Jan. 4

- Core
- Casual
- Marginal
Activation experiment

Users in experiment, by engagement level

- Core
- Casual
- Marginal

Date
- 1/1
- 1/8
- 1/15
- 1/22
- 1/29

Users
- 0
- 5,000
- 10,000
- 15,000
- 20,000
- 25,000
- 30,000
Comparing experiments

Total users in each experiment

![Graph showing total users in each experiment over time. The graph compares two experiments: Closeup (blue diamonds) and Activation (green triangles). Users increase over time, with Closeup starting at a higher user count on 1/1 and both experiments reaching 50,000 users by 1/29.]
Comparing experiments

Total users in each experiment

Users

Date

Closeup  Activation
For every 100 core users, the enabled group has 0 more savers than control.

Absolute lift = 0% absolute lift

For every 100 casual users, the enabled group has 1 more saver than control.

Absolute lift = 1% absolute lift

For every 100 marginal users, the enabled group has 2 more savers than control.

Absolute lift = 2% absolute lift
Activation experiment: Impact from shipping treatment

### Users exposed to search suggestion

<table>
<thead>
<tr>
<th>Date</th>
<th>Core</th>
<th>Casual</th>
<th>Marginal</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/1</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>1/8</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>1/15</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
</tr>
<tr>
<td>1/22</td>
<td>3%</td>
<td>3%</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>1/29</td>
<td>4%</td>
<td>4%</td>
<td>4%</td>
<td>4%</td>
</tr>
</tbody>
</table>

### Absolute lift in savers

<table>
<thead>
<tr>
<th>Date</th>
<th>Activation experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/1</td>
<td>0%</td>
</tr>
<tr>
<td>1/8</td>
<td>0%</td>
</tr>
<tr>
<td>1/15</td>
<td>0%</td>
</tr>
<tr>
<td>1/22</td>
<td>0%</td>
</tr>
<tr>
<td>1/29</td>
<td>0%</td>
</tr>
</tbody>
</table>
Activ. experiment impact

Additional savers created

![Graph showing the number of savers over time, categorized by Core, Causal, Marginal, and Total. The x-axis represents the dates 1/1, 1/8, 1/15, 1/22, and 1/29, and the y-axis represents the number of savers ranging from 0 to 800. Each category shows an increase in savers over time.]
Comparing experiments

Additional savers created by treatment

Number of savers

Date

Closeup  Activation

Number of savers

Date
Comparing experiments

Additional savers created by treatment

![Graph showing the comparison of experiments with additional savers created by treatment over time.]
population
population  lift
population $\times$ lift = impact
Taking a shortcut to long-term impact: predictive modeling
Closeup experiment

Users in experiment, by engagement level

![Graph showing users in experiment, by engagement level over time from January 1 to January 29. The graph includes lines for Core, Casual, Marginal, and Total users, with the Total line consistently higher than the others.]
Population curve
Total users in experiment

Date
1/1 1/2 1/3 1/4 1/5 1/6 1/7 1/8

Users
0M 0M 0M 0M 0M

Ramped to 10%
Population curve

Total users in experiment

- Ramped to 10%
- Ramped to 50%
Closeup experiment lift

Changing population means changing overall lift
Closeup experiment lift
Assumed constant lift within engagement levels

Date
1/1 1/8 1/15 1/22 28

Core
Casual
Marginal
Activation experiment: Lift in savers

Core:
For every 100 core users, the enabled group has 0 more savers than control.

= 0% absolute lift

Casual:
For every 100 casual users, the enabled group has 1 more saver than control.

= 1% absolute lift

Marginal:
For every 100 marginal users, the enabled group has 2 more savers than control.

= 2% absolute lift
Control
Normal home feed

Enabled
Search suggestion
Lift curve
Lift curve
Lift curve

Treatment effects wear off over time

Days since triggering experiment (days in)

Absolute lift
Users exposed to treatment

- Ramped to 10%
- Ramped to 50%
Closeup experiment

Users in experiment, by engagement level

Users

Date

0

1/1

1/2

1/3

1/4

1/5

1/6

1/7

Core
Closeup experiment

Users triggering in each day, by engagement level
Closeup experiment

Lift curve for core users

Days since triggering experiment (days in)

Absolute lift
Closeup experiment: Impact from shipping treatment

Users exposed to new closeup

Date

1/1 1/2 1/3 1/4 1/5 1/6 1/7

Absolute lift in savers

Days since triggering

0 7 14 21 28

0% 0.5% 1% 1.5% 2%
Closeup experiment impact

Additional savers created, Jan. 1 cohort
Shipping closeup experiment

Users exposed to new closeup

Days since triggering

Absolute lift in savers

Date

1/1 1/2 1/3 1/4 1/5 1/6 1/7

0 800 1,600 2,400 3,200 4,000

0% 0.5% 1% 1.5% 2%
Closeup experiment impact

Additional savers created, Jan. 1 + Jan. 2 cohorts
Shipping closeup experiment

Users exposed to new closeup

<table>
<thead>
<tr>
<th>Date</th>
<th>1/1</th>
<th>1/2</th>
<th>1/3</th>
<th>1/4</th>
<th>1/5</th>
<th>1/6</th>
<th>1/7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Users exposed</td>
<td>800</td>
<td>1,600</td>
<td>2,400</td>
<td>3,200</td>
<td>4,000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Absolute lift in savers

<table>
<thead>
<tr>
<th>Days since triggering</th>
<th>0%</th>
<th>0.5%</th>
<th>1%</th>
<th>1.5%</th>
<th>2%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>7</td>
<td>14</td>
<td>21</td>
<td>28</td>
</tr>
</tbody>
</table>
Closeup experiment impact

Additional savers created, Jan. 1 + Jan. 2 + Jan. 3 cohorts

[Graph showing the number of savers over time for Jan. 1 Cohort, Jan. 2 Cohort, and Jan. 3 Cohort]
Closeup experiment impact

Additional savers created, all cohorts

<table>
<thead>
<tr>
<th>Date</th>
<th>Jan. 1 Cohort</th>
<th>Jan. 2 Cohort</th>
<th>Jan. 3 Cohort</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/1</td>
<td>120</td>
<td>100</td>
<td>80</td>
<td>300</td>
</tr>
<tr>
<td>1/8</td>
<td>60</td>
<td>40</td>
<td>20</td>
<td>120</td>
</tr>
<tr>
<td>1/15</td>
<td>30</td>
<td>20</td>
<td>10</td>
<td>60</td>
</tr>
<tr>
<td>1/22</td>
<td>20</td>
<td>15</td>
<td>10</td>
<td>45</td>
</tr>
<tr>
<td>1/29</td>
<td>15</td>
<td>10</td>
<td>5</td>
<td>30</td>
</tr>
</tbody>
</table>
Lift as function of days in

$\mathcal{L}(t) =$

![Graph showing the decrease of absolute lift over days since triggering the experiment.](image-url)
Users triggering on date $d$

$$U(d) =$$

![Bar chart showing users by date]
Ideal definition of impact

If we knew everything, we’d want to calculate:

$$Impact_Q = \sum_{d=0}^{Q} U(d) \times L(Q - d)$$

where:

$$U(d):$$ Users triggering into experience on day $d$

$$L(t):$$ The absolute lift $(treatment - control)$ $t$ days after triggering
But how do we know $L$ and $U$?

$L(t) = ?$

$U(d) = ?$
Lift curve
Population curve

Total users in experiment

- Ramped to 10%
- Ramped to 50%
Triggering curve
Users triggering in each day

- Ramped to 10%
- Ramped to 50%
But how do we know $L$ and $U$?

$L(t) = \ldots$

$U(d) = \ldots$
Closeup experiment lift

Changing population means changing overall lift
But...what if ideal isn’t possible?

If we can’t come up with a good model for $L(t)$ for large $t$, we can at least estimate:

$$\text{Impact}_Q \propto \left( \sum_{d=0}^{Q} U(d) \right) \times L(W), \text{ for some } W$$
Assume a constant lift

If \( W = 14 \),
\[
L(W) = \_\_\_
\]
Predicting triggering curve

Users entering experiment each day.

Users:
0 25,000 50,000 75,000 100,000 125,000 150,000 175,000 200,000

Weeks since launching experiment:
1/1 1/8 1/15 1/22 1/29

Graph:
- Ramped to 10%
- Ramped to 40%
- Holdout
Predicting triggering curve

Users entering experiment each day.

\[ U(d) \sim f(\text{days since ramp, ramp percentage, day of week}) \]
Predicting triggering curve

Users entering experiment each day.
Population curve

Total users in experiment, predicted to 12 weeks

- Users: 0M, 2M, 3M, 5M, 6M
- Date: 1/1, 1/8, 1/15, 1/22, 1/29, 2/5, 2/12, 2/19, 2/26, 3/5, 3/12, 3/19
- Ramped to 100%
Population curve varies based on
Lift curve varies based on
What we want:

If I ship this experience today, how many (clicks | impressions | DAUs) will I have created (4/13/52 weeks from now | at the end of the quarter/half/year)?
Comparing experiments

Total users in each experiment

Users

Date

Closeup  Activation
Comparing experiments

Additional savers created by treatment
Taking a more long-term view of experiments mitigates engagement bias and helps prioritize efforts to engage non-core users.