Steering User Behavior with Badges

Ashton Anderson

Joint work with Dan Huttenlocher, Jon Kleinberg, and Jure Leskovec
People will work amazingly hard to earn badges
“Give me enough medals and I’ll win you any war.”

– Napoleon
**BADGE TYPES**

- **Meteorite** badges are common and easy to earn when just getting started.
- **Moon** badges are uncommon and represent an investment in learning.
- **Earth** badges are rare. They require a significant amount of learning.
- **Sun** badges are epic. Earning them is a true challenge, and they require impressive dedication.
- **Black Hole** badges are legendary and unknown. They are the most unique Khan Academy awards.
- **Challenge Patches** are special awards for completing topic challenges.
Badges play multiple roles:

– can recognize a wide range of types of activities
– serve both as credentials and create incentives

Despite surface-level simplicity, badges are complex

? How do badge criteria translate into effects on user behavior?

? How should site designers define badges to achieve a particular outcome?
Need a model of user behavior in the presence of badges
This talk:

1. Develop a model of user behavior in the presence of badges (theory)

2. Validate qualitative model predictions with real-world data (empirical analysis)

3. Investigate how to optimally design badges (algorithms and simulations)
1. Develop a model of user behavior in the presence of badges (theory)

2. Validate qualitative model predictions with real-world data (empirical analysis)

3. Investigate how to optimally design badges (algorithms and simulations)
Model Goals

✦ Assume that badges have value to users

✦ A user trades off between her preferred mix of activities and the goal of winning badges

✦ We’d like to see this produce effects on both overall engagement and “steering” - balancing activities differently
Our Model

There is a population of users and a site designer.
The Action Space

There are action types $A_1, A_2, \ldots, A_{n+1}$ that form an action space.
The Action Space

There are action types $A_1, A_2, \ldots, A_{n+1}$ that form an action space.
The Action Space

There are action types $A_1, A_2, \ldots, A_{n+1}$ that form an action space.
The Action Space

There are action types $A_1, A_2, \ldots, A_{n+1}$ that form an action space.
The Action Space

There are action types $A_1, A_2, \ldots, A_{n+1}$ that form an action space.
The Action Space

There are action types $A_1, A_2, \ldots, A_{n+1}$ that form an action space.
The Action Space

There are action types $A_1, A_2, ..., A_{n+1}$ that form an action space.
The Action Space

There are action types $A_1, A_2, ..., A_{n+1}$ that form an action space.
The Action Space

There are action types $A_1, A_2, \ldots, A_{n+1}$ that form an action space.
Probabilistic Actions

At each step, choose a probability distribution \( p_a \) to draw next action from (and exit with probability \( 1 - \theta \))
Preferred Distribution

Users prefer certain actions over others: each user is born with a preferred distribution over actions $p$. 

![Graph showing the distribution of type $A_1$ and type $A_2$ actions.](image)
Cost for deviating

If the user picks probability distribution $p'$, then he incurs a utility penalty $g(p, p') = \|p - p'\|_2^2$.
Badges

- Set of badges $B$
- Each badge $b$ is a subset of cells in action space and has value $V_b$
User’s Utility Function

User’s utility is composed of three parts:
1. Value from badges won
2. Cost for deviating from \( p \)
3. (Recursively) Utility from next state

Utility from won badges

\[
f(a) = \sum_{b \text{ won}} V_b + \theta [p_a^1 \cdot f(a_1 + 1, a_2) + p_a^2 \cdot f(a_1, a_2 + 1)] - g(p, p_a)
\]

Cost

Expected utility of next state
User’s Utility Function

User’s utility is composed of three parts:
1. Value from badges won
2. Cost for deviating from $p$
3. (Recursively) Utility from next state

User’s optimization problem: pick $p_a$ for each state to maximize $f(0)$
User’s Optimization Problem

Example badge: $A_1 \geq 7$
Use dynamic programming to solve this problem:

Past badge boundary, no incentive to deviate from \( p \)
Use dynamic programming to solve this problem:

Past badge boundary, no incentive to deviate from $p$

User sets $p_a = p$ for all states $a$ past boundary

and value of each such state is $V_b$
Before badge boundary, select $p_a$ to maximize expected utility.
In this case, collapse along $A_2$ dimension since $f(a_1, a_2) = f(a_1, a'_2)$ for all $a_1, a'_2$. 
Problem becomes one-dimensional, so we can solve from badge boundary back to origin

# of site action 1 actions
Problem becomes one-dimensional, so we can solve from badge boundary back to origin

# of site action 1 actions
Problem becomes one-dimensional, so we can solve from badge boundary back to origin

# of site action 1 actions
Problem becomes one-dimensional, so we can solve from badge boundary back to origin

# of site action 1 actions
Problem becomes one-dimensional, so we can solve from badge boundary back to origin

# of site action 1 actions
Example: badge at 25 type $A_1$ actions

Canonical behavior: user “steers” in $A_1$ direction as he approaches the badge boundary, then resets
Two-dimensional scenario:
a badge on each dimension
Two-dimensional scenario:
a badge on each dimension

Given
Two-dimensional scenario:
a badge on each dimension

1-dimensional case

# of type $A_1$ actions

# of type $A_2$ actions
Two-dimensional scenario: a badge on each dimension.
Two-dimensional scenario:
a badge on each dimension

Solve directly
1. Develop a model of user behavior in the presence of badges (theory)

2. Validate qualitative model predictions with real-world data (empirical analysis)

3. Investigate how to optimally design badges (algorithms and simulations)
- Programming-related Q&A answering site
- Heavy use of badges

Peter Mortensen
5,214
6 27 62

Badge counts
Won when a user votes on 600 questions

Won when a user votes 300 times
Users accelerate as they approach the badge boundary
1. Develop a model of user behavior in the presence of badges (theory)

2. Validate qualitative model predictions with real-world data (empirical analysis)

3. Investigate how to optimally design badges (algorithms and simulations)
Badge placement problem:

General question: how should the site designer “place” badges in action space to achieve desired effects?
Concrete question: If the site designer can place one badge and wants to maximize actions on a particular dimension, where should she put it?

Optimal badge placement at internal optimum
Given two badges of fixed value, the designer should place them approximately evenly apart for maximum effect.
Given a fixed amount of value, an even split of value optimizes yield.
Conclusions

– We introduced a model of user behavior in the presence of badges

– Model predicts that users steer between actions and engage more

– Validated the model’s predictions against real-world Stack Overflow data

– Introduced and investigated the badge placement problem
Thanks
Overjustification Effect

when an expected external incentive decreases a person’s intrinsic motivation to perform a task

e.g. paying for blood donations reduces the number of donors

Very possible for badges to backfire!
Our Model

Badge boundary

Example badge: $A_1 \geq 8$