

Advertising at Scale

Sophia Tee, Director of Data Science

Alex Eftimiades, Applied Machine Learning Scientist



Introduction – Driving Conversions through Advertising

Penguin
Random
House

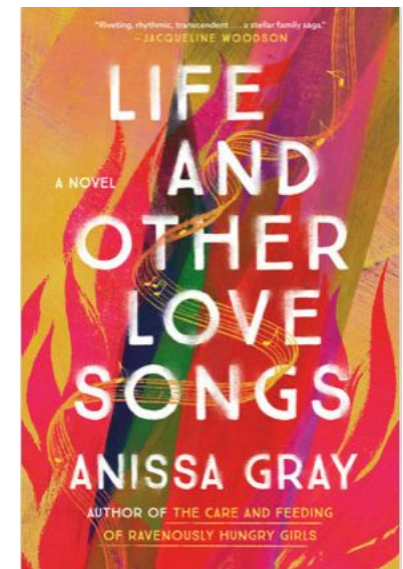
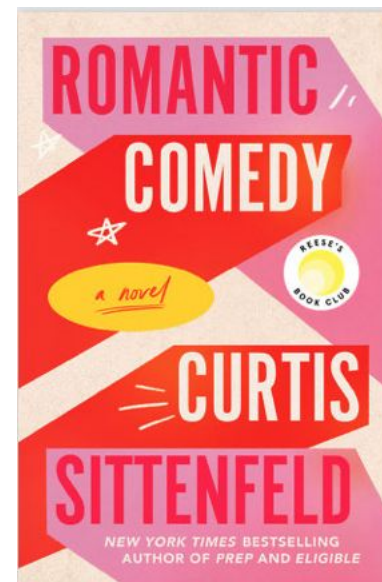
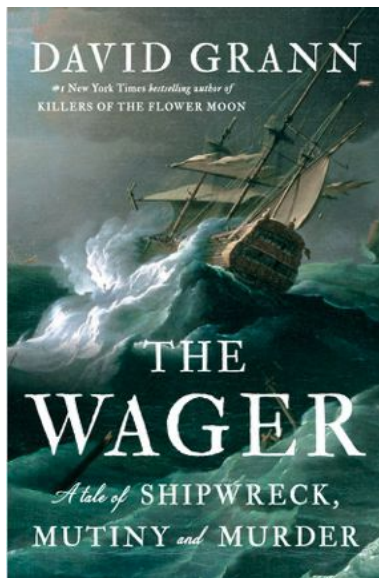
RIGHT
CONTENT



RIGHT
MESSAGE



RIGHT
AUDIENCE



Automation of Campaign Process

Title	Version	Artwork	Copy	Audience
The Disappearing	1	1	1	1
The Disappearing	2	1	2	1
The Disappearing	3	2	1	1
The Disappearing	4	2	2	1
The Disappearing	5	1	1	2
The Disappearing	6	1	2	2
The Disappearing	7	2	1	2
The Disappearing	8	2	2	2

Automated way of picking winner copy, audience and creative combinations and optimizing our budget

Push remaining budget to winning combination

#1 ranked combination of:

- Artwork
- Copy
- Audience

Audience gets updated weekly to ensure latest performance gets taken into account

Audience
Generatio
n

With the ISBNs and PRH first party data as an input, the audience model generates a candidate audience

Creative
Generatio
n

With the ISBNs characteristics (genre, season, trend) as an input, the creative automation model creates a list of candidate backgrounds

Copy
Generatio
n

With the ISBNs, it is possible to leverage AI to create ad copy by scraping product page content and improving it over time (currently working with two vendors)

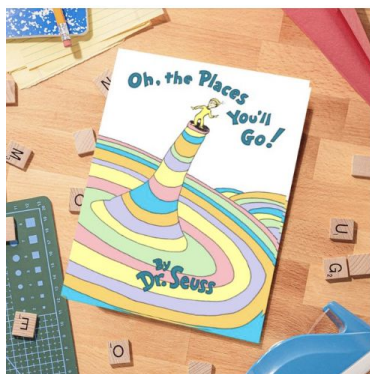
Facebook
UI

Campaign details, audience, creative and copy are pushed automatically to Facebook through the API

Dynamically switch out book based on title id provided in api call



Dynamically switch out the background based on age group, genre, season



Ide
ntif
y
loc
atio
n
(x, rotation, view angle)
and
orie
ntat
ion
of
boo
ks
in
bac
kgr
oun
ds



- Label books with image segmentation
- Train image segmentation algorithm and use result to identify each corner of book
- Use corner locations to deduce orientation and location

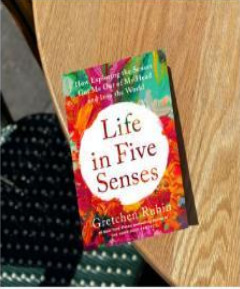


INPUT : product detail page URL



Penguin Random House
Sponsored · 🌐

Discover a life of more energy, creativity, luck, and love by tuning in to the five senses with Gretchen Rubin's *Life in Five Senses*.



penguinrandomhouse.com
Tune Your Life: 5 Senses
#1 New York Times Bests...

Learn more

Penguin Random House
Sponsored · 🌐

Emily joins a reality dating show, only to find herself in a world of love and manipulation. Will she find what she's looking for or is ...See more

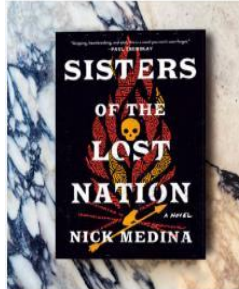


penguinrandomhouse.com
Find Love on TV!
The One is a razor-sharp ...

Learn more

Penguin Random House
Sponsored · 🌐

Anna Horn struggles with her place on the rez, searching for the key to her tribe's past. When her sister goes missing, she'll do ...See more



penguinrandomhouse.com
Unearth Ancient Horrors
Sisters of the Lost Nation ...

Learn more

Penguin Random House
Sponsored · 🌐

Explore the lives of spirited heroines struggling to find a place they belong. Read the classic stories of their unraveling lives in ...See more



penguinrandomhouse.com
Discover Nella Larsen's Brilliance

Learn more

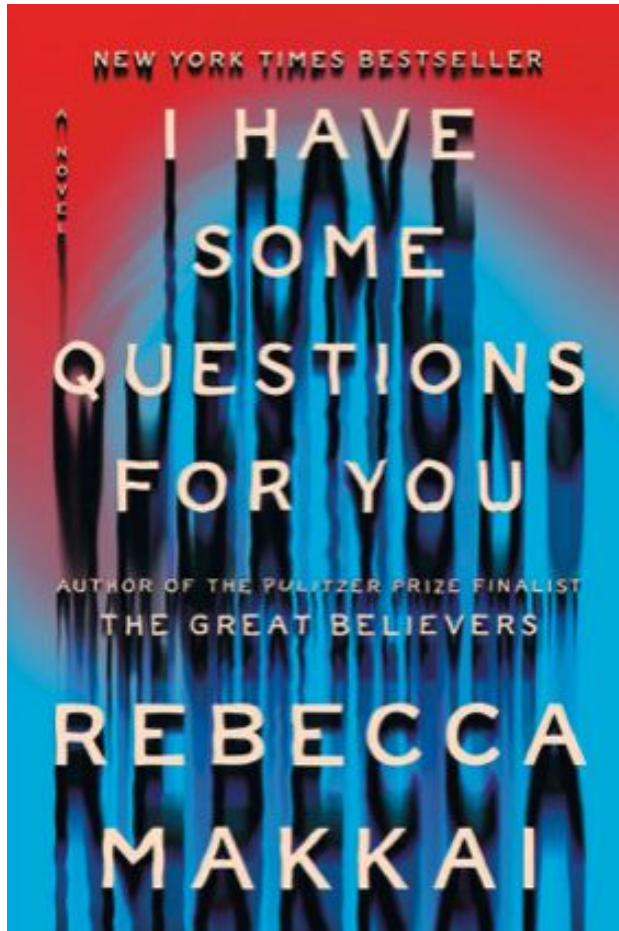
Penguin Random House
Sponsored · 🌐

Acclaimed author Jonathan Rosen's haunting investigation of the forces that led his closest childhood friend from promise to ...See more



penguinrandomhouse.com
The Best Minds: Brilliant Promise and Tragedy

Learn more



Copy 1

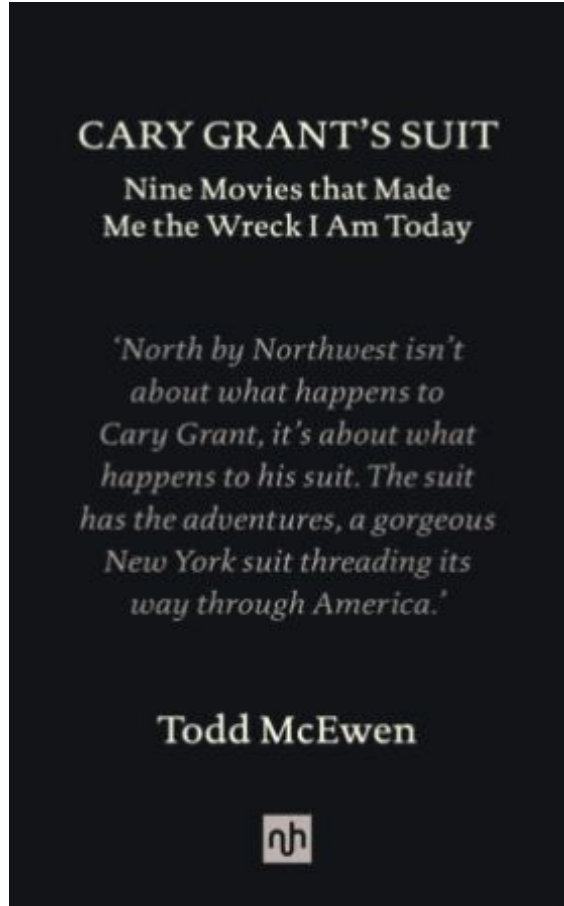
Rebecca Makkai's 'I Have Some Questions For You' is a riveting true-crime page-turner and campus coming-of-age novel. Join Bodie Kane as she dives into a mysterious past.

Copy 2

Unlock a mystery and join Bodie Kane as she dives into a mysterious past with Rebecca Makkai's 'I Have Some Questions For You.'

Multiple references to a character (Bodie Kane) that no one has heard of (it's not like a Jack Ryan or Alex Cross, a recurring series character)

- *Text is too long*



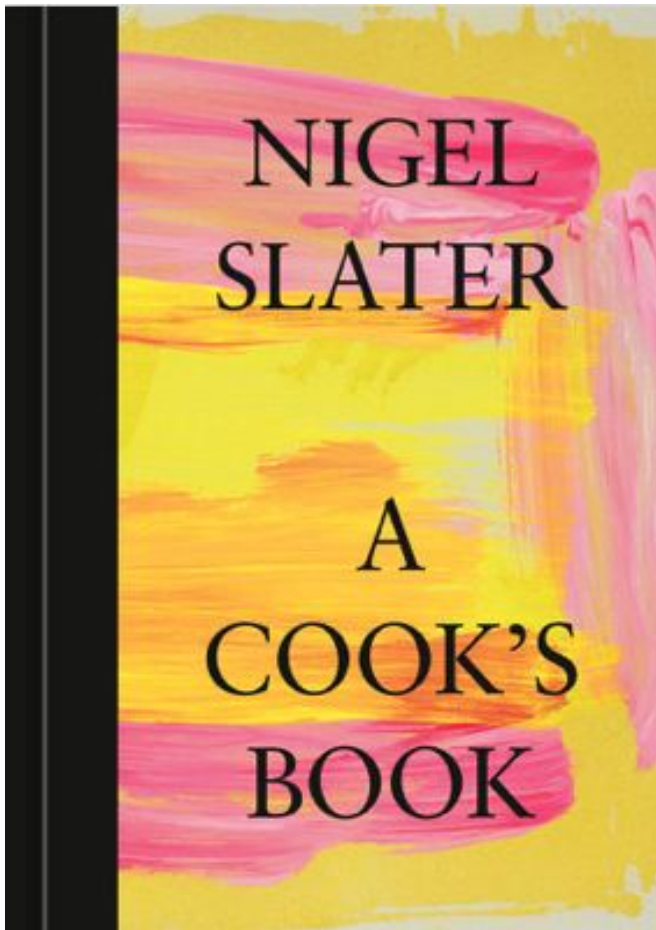
Headline	Text	Description
Experience Old Hollywood	"North by Northwest isn't about what happens to Cary Grant, it's about what happens to his suit." - Ian Sansom, The Telegraph (UK)	A love letter to old Hollywood, this is a book for anyone interested in film.

- *The text is not a review of the book.*
- *And we also would not necessarily want to include a UK publication as an endorsement when targeting readers in the US.*



Headline	Text	Description
Dream Big Adventures!	"A great wedding shower gift or present for an adventurous couple in your life," - Reviewer	Keep track of your goals and memories with Our Bucket List Adventures!

- *The text is not attributed to an actual person/publication*
- *Should not use product description as an endorsement*



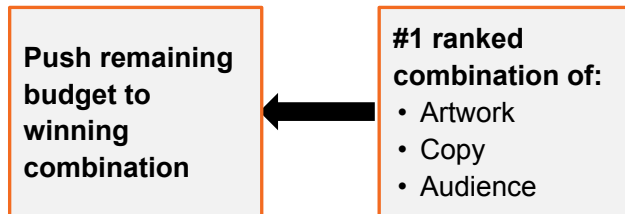
Headline	Text	Description
Nigel Slater's Essential Cookbook	This is by far Slater's most personal book yet," said a reviewer.	A collection of more than 150 delicious, easy, and gratifying plant-based and meat recipes.

- *Text field does not contain an opening quotation mark*
- *The Text is not attributed to an actual person/publication*

Optimization of Campaigns through Testing

Title	Version	Artwork	Copy	Audience
The Disappearing	1	1	1	1
The Disappearing	2	1	2	1
The Disappearing	3	2	1	1
The Disappearing	4	2	2	1
The Disappearing	5	1	1	2
The Disappearing	6	1	2	2
The Disappearing	7	2	1	2
The Disappearing	8	2	2	2

Automated way of picking winner
copy, audience and creative
combinations and optimizing our
budget



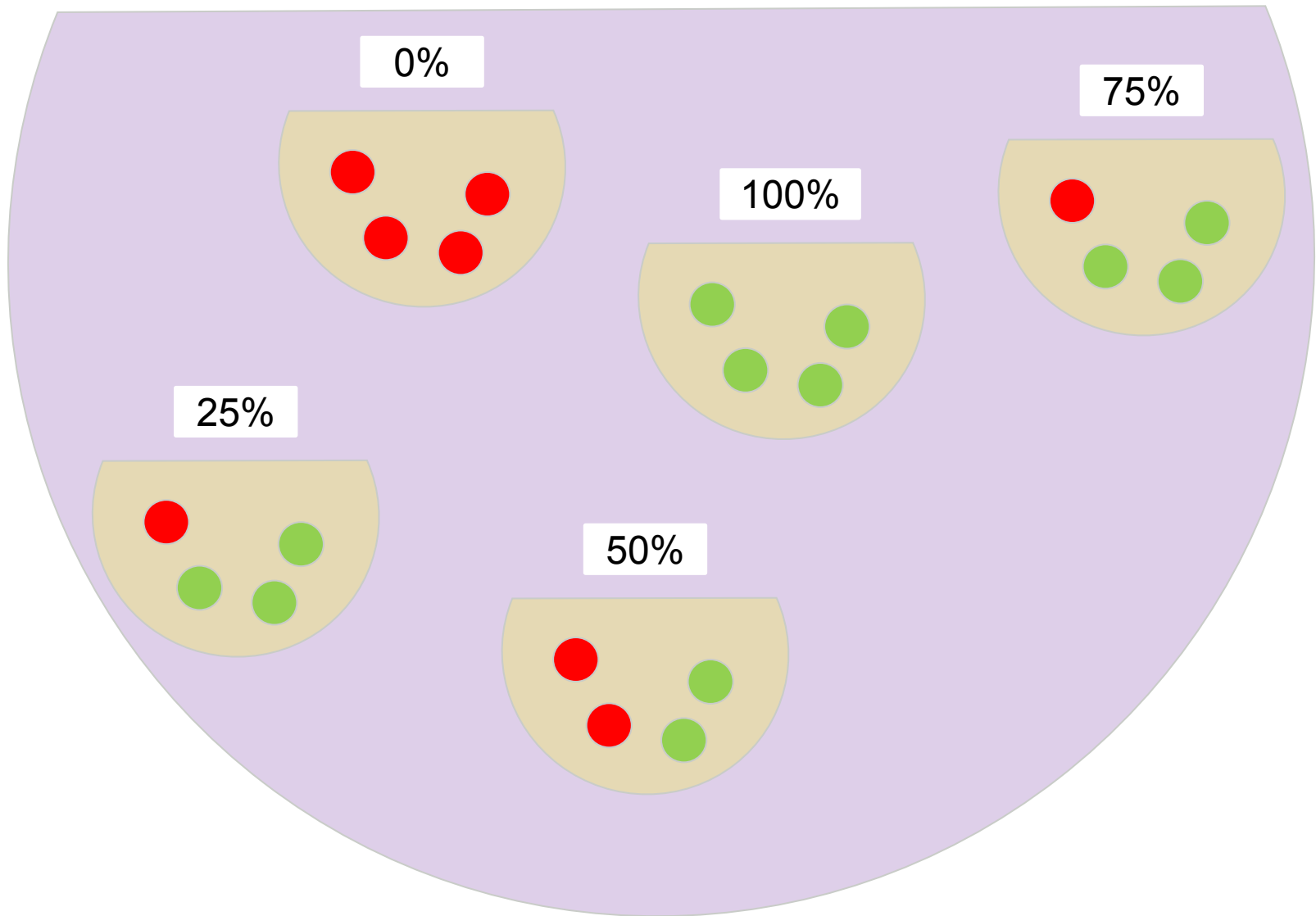
Audience gets updated weekly to
ensure latest performance gets taken
into account

Is the data representative?

- But it was just the time of year!
- No it was underperforming ads!
- It was the combination of ads and audience!

If it were... what can we say about the results with the data we do have?

Simplified Click Through Rates and Urns



posterior

Likelihood

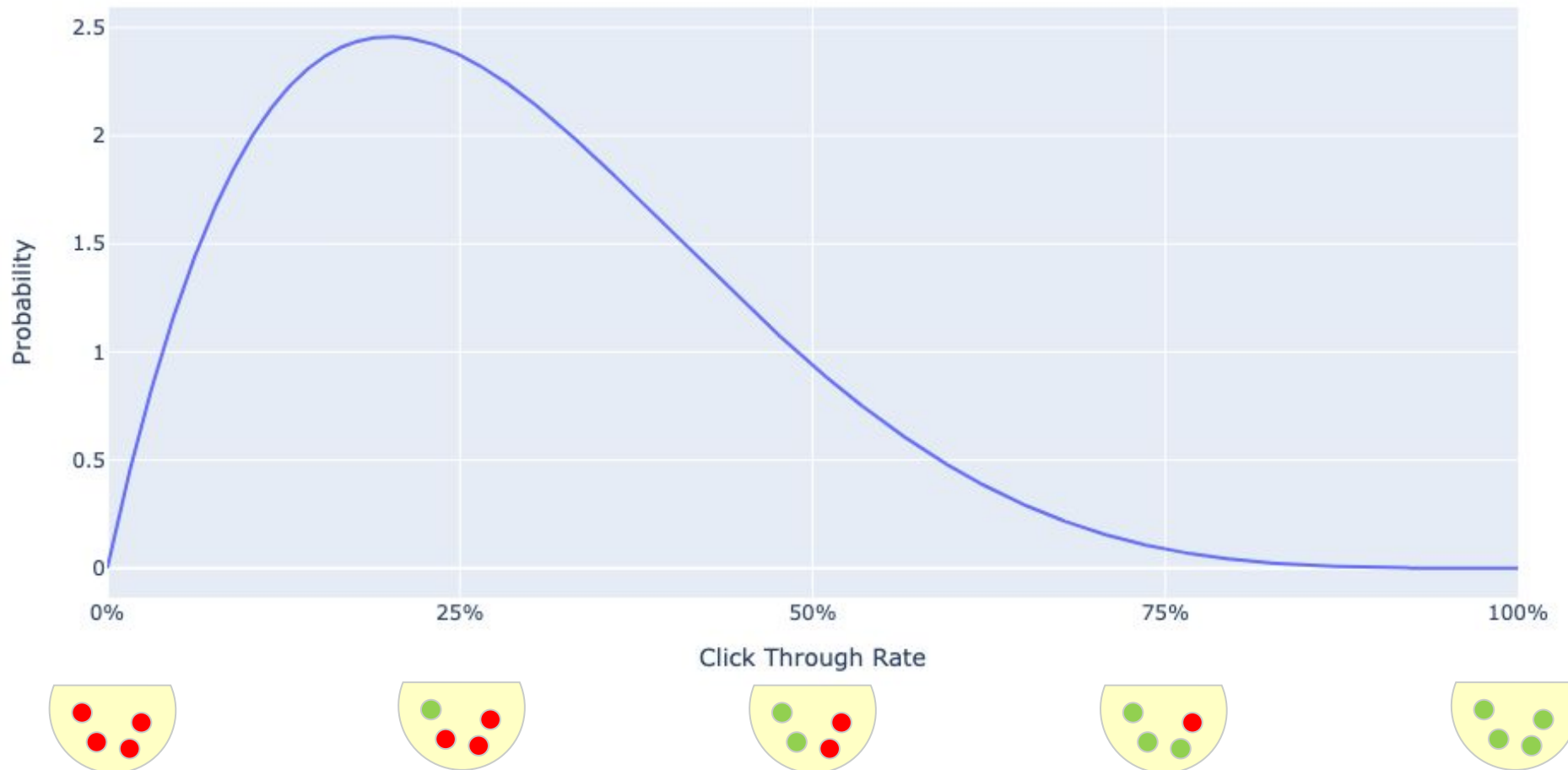
prior

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

Hypothesis	Prior	Likelihood	Product	Posterior
	$p(H)$	$p(D H)$	$p(H) p(D H)$	$p(H D)$
Urn 0	1/5	0	0	0
Urn 1	1/5	1/4	1/20	1/10
Urn 2	1/5	2/4	2/20	2/10
Urn 3	1/5	3/4	3/20	3/10
Urn 4	1/5	4/4	4/20	4/10
Total $p(D)$			10/20	

Continuous Extension with Beta distributions

Beta Distribution ($\alpha=2, \beta=5$)



- Show ad 12 times, of which 8 were clicked...
- Click through rate is a beta distribution (assuming e.g. a uniform prior)
- $P(\text{ROI})$ is given by $P(\text{click through rate})$
- This allows us to view outcomes in a probabilistic way

Epsilon Greedy

- Given each distribution over possible ROIs *per* campaign
- Sample from them to determine $P(\text{Campaign}_{\{i\}} \text{ has highest ROI})$
- Eliminate/retire campaigns that have a $p < \epsilon$ chance of maximizing ROI
- This is equivalent to Epsilon-Greedy strategy in reinforcement learning
- Note there is no single answer to “how much data” is enough.
- More data will be needed for close calls.

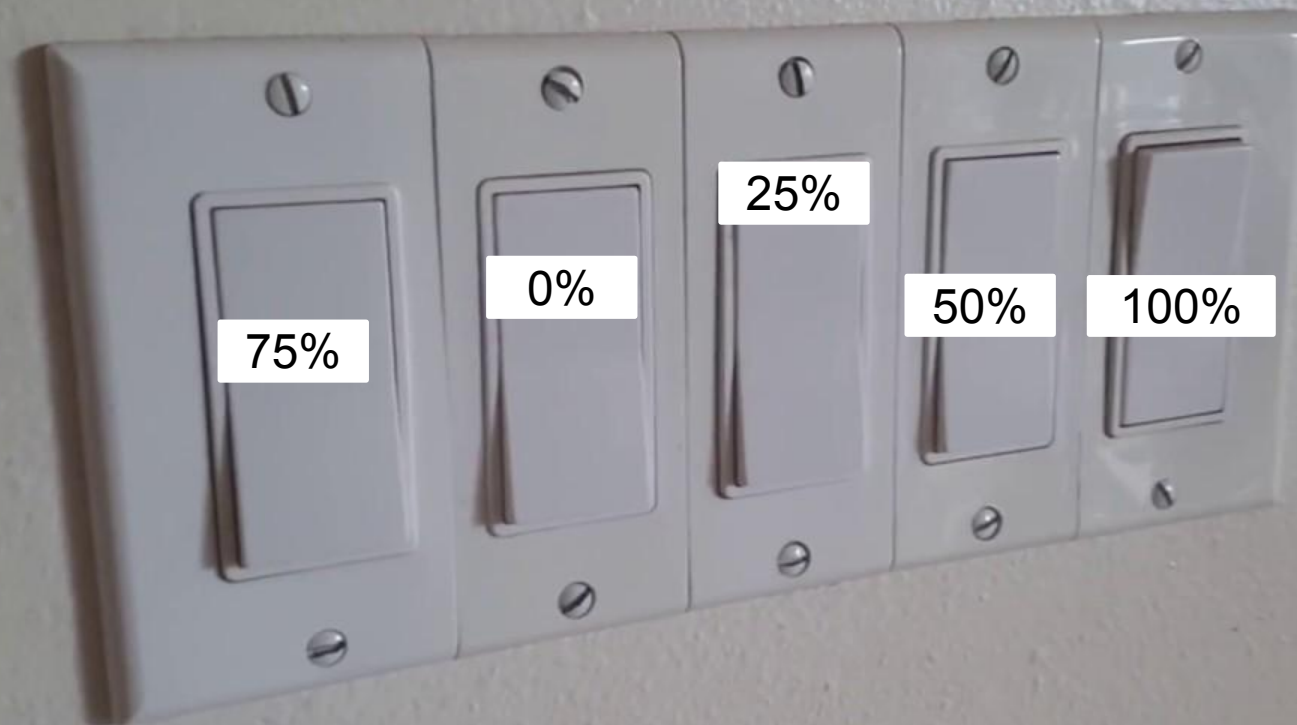
Thompson Sampling

- Given each distribution over possible ROIs *per* campaign
- Sample from them to determine $P(\text{Campaign}_{\{i\}} \text{ has highest ROI})$
- Generate variant $\text{Campaign}_{\{i\}}$ in accordance with P
- If $\text{Campaign}_{\{i\}}$ has a 0.001% chance of having the highest ROI, don't retire it...
- ...just launch it 0.001% of the time

- Opportunity to expand to other ad platforms where automation is needed.
- With widespread adoption and help from marketers, creative and copy can be elevated to improve the quality of the output.
- Ability to expand to multi title campaigns
- Testing with different seasonality, trend conditions

Questions??

Estimating the probability of hypotheses: Clickthrough rates and switches

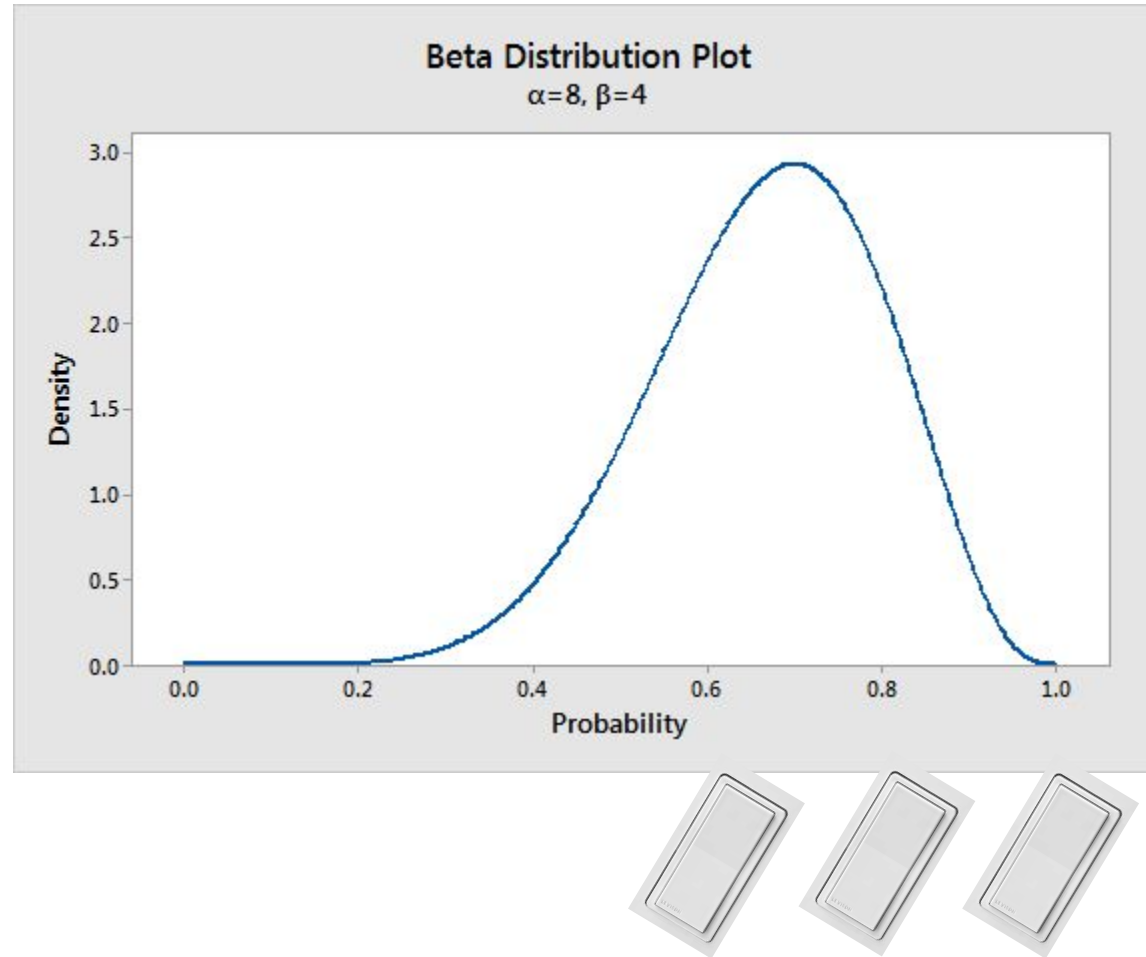


But what about all the confounding factors?

- $P(\text{Purchase} \mid \text{Billed}, \text{background}=\text{christmas}, \text{moon}=\text{full}, \dots)$
- $P(\text{Purchase} \mid \text{Billed}, \text{confounders}=\text{state})$
- $P(\text{Purchase} \mid \text{Billed}) = E_{\text{state} \sim \text{universe}} P(\text{Purchase} \mid \text{Billed}, \text{confounders}=\text{state})$
- Random sampling over confounding variables is sufficient
-but expensive
- We can sometimes trade off time for variety (try a variation on a single ad in summer and winter, or try that variation on summer and winter ads in the summer).
- Probably need to be agreed upon ahead of time



Continuous Clickthrough rates: Beta beta distributions



ROI = Revenue / Spend

= (purchases x Sales price) / (views x Cost)

= purchases / views x Sales Price / Cost

= P(Purchase | viewed) x Sales Price / Cost

~ P(Click | viewed) P(Purchase | Click) x Sales Price / Cost

So a distribution