

[320] Web 5: A/B Testing

Tyler Caraza-Harter

Source for Examples/Lessons

[Ronny Kohavi](#) Keynote Talk at KDD conference (Knowledge Discovery and Data Mining)

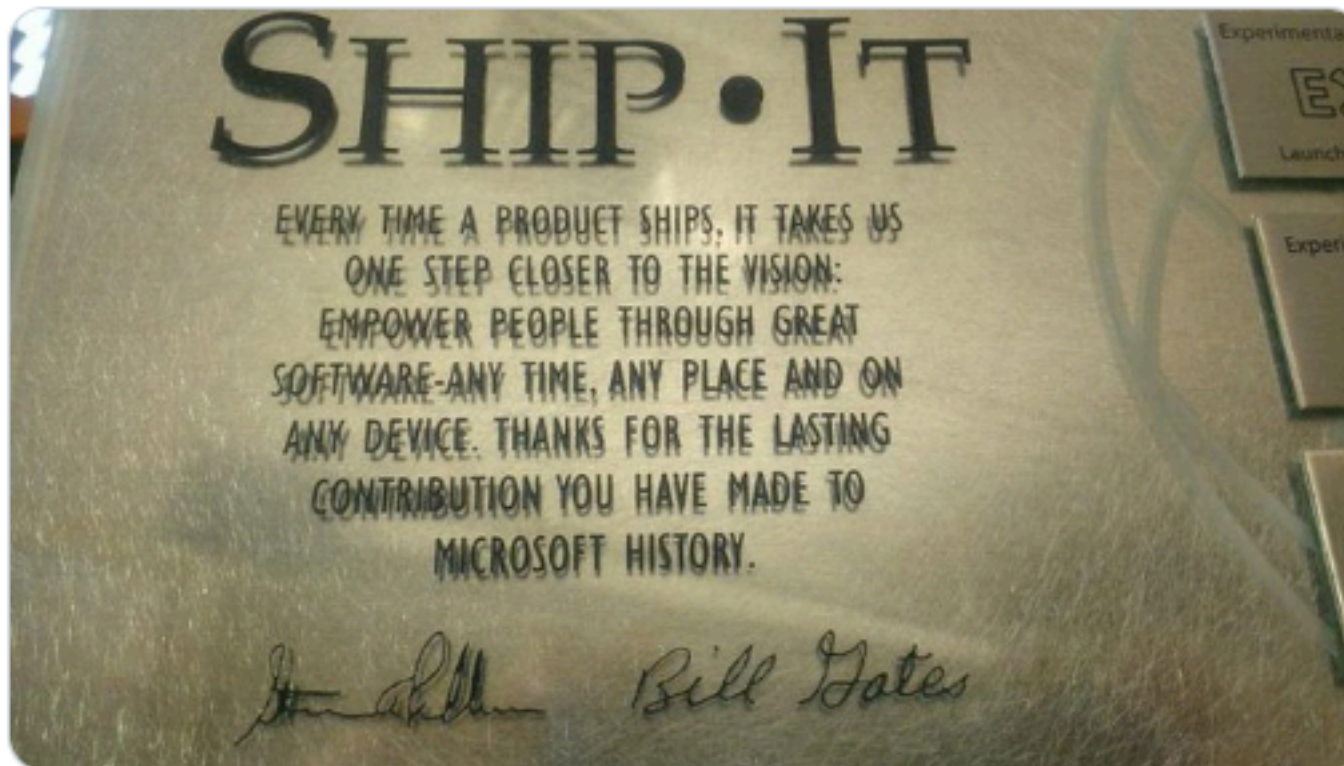
Title: Online Controlled Experiments: Lessons from Running A/B/n Tests for 12 years

Video: <https://exp-platform.com/kdd2015keynotekohavi/>

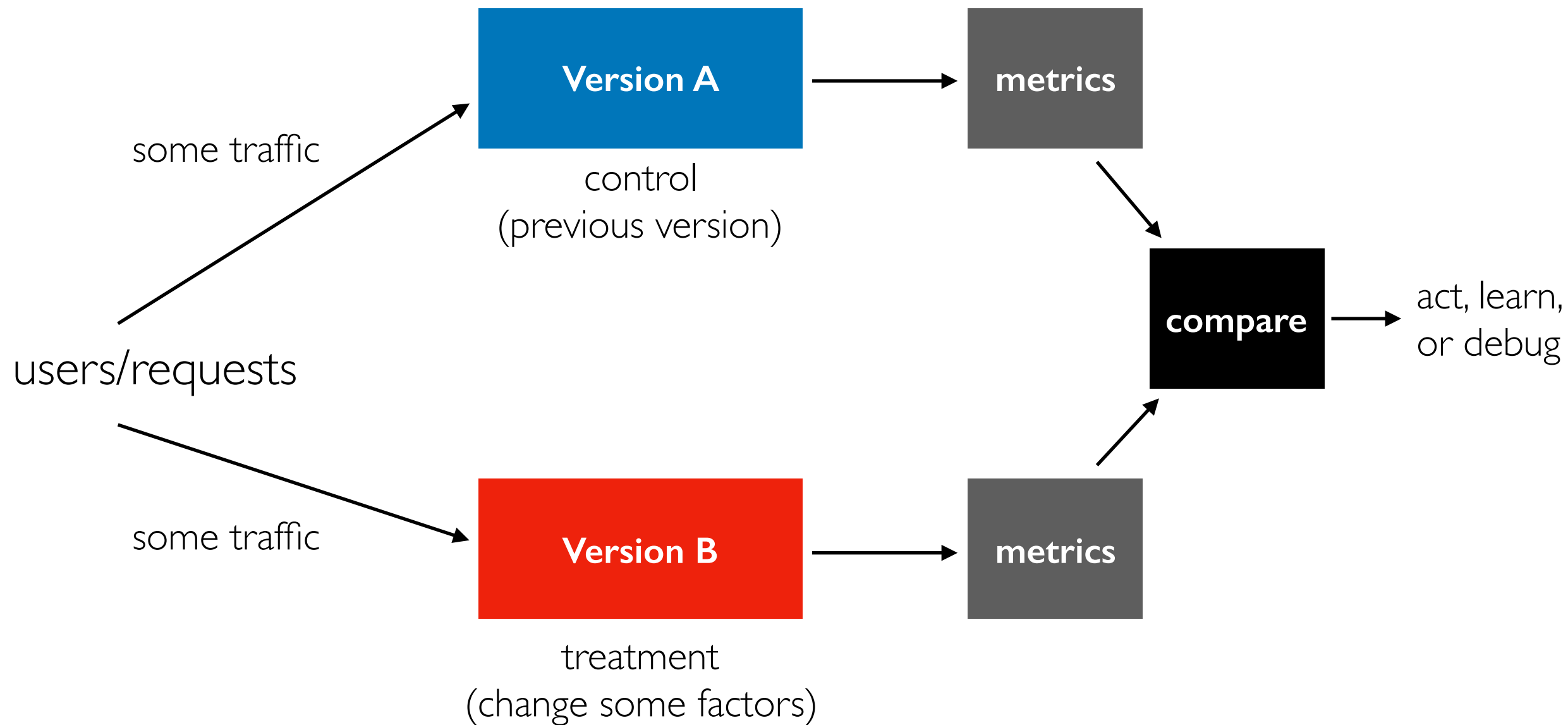


Ronny Kohavi @ronnyk · Nov 7, 2014

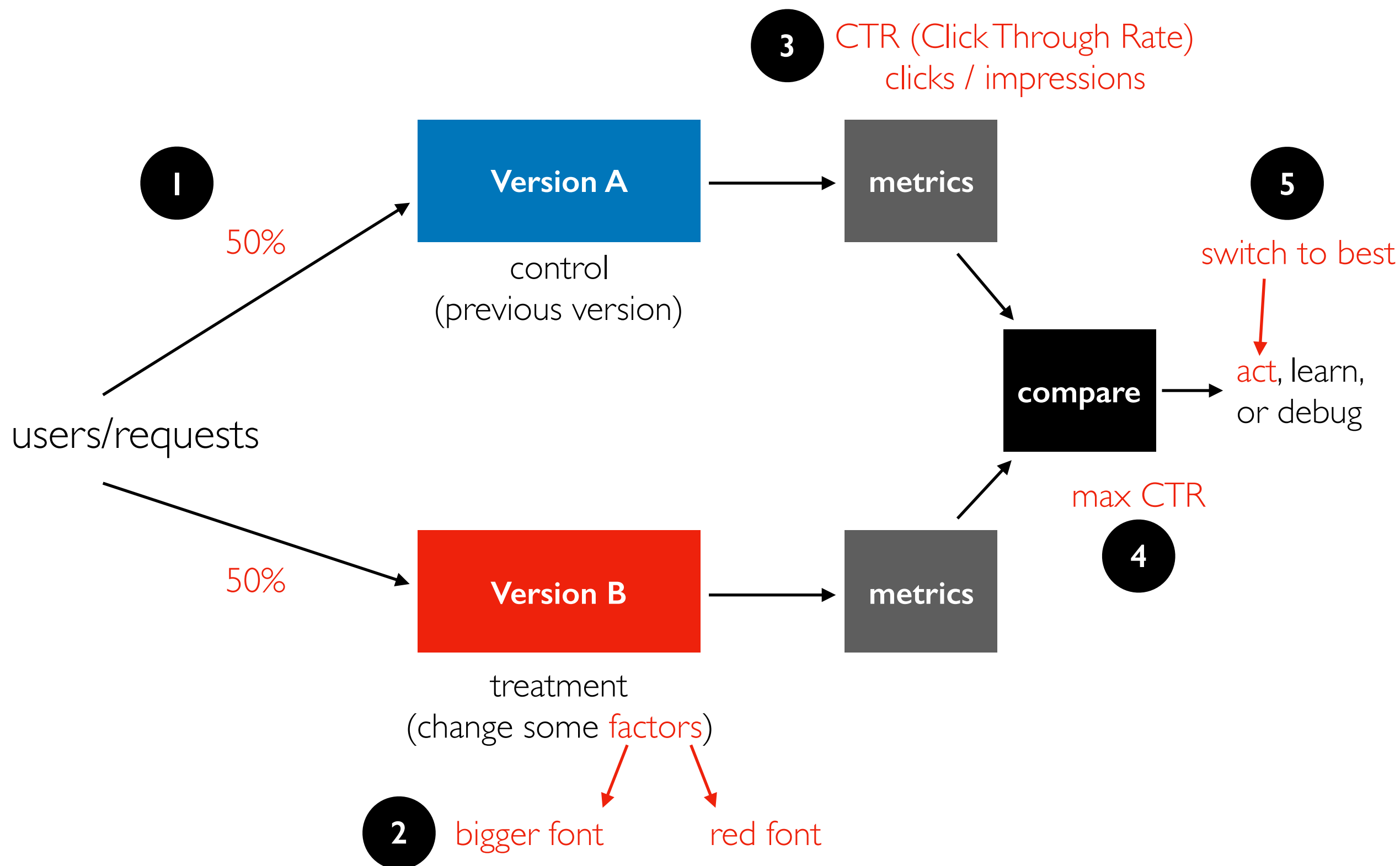
Microsoft stopped ship-it-awards today! With [#abtesting](#), it's about user-impact; NOT shipping is often better!



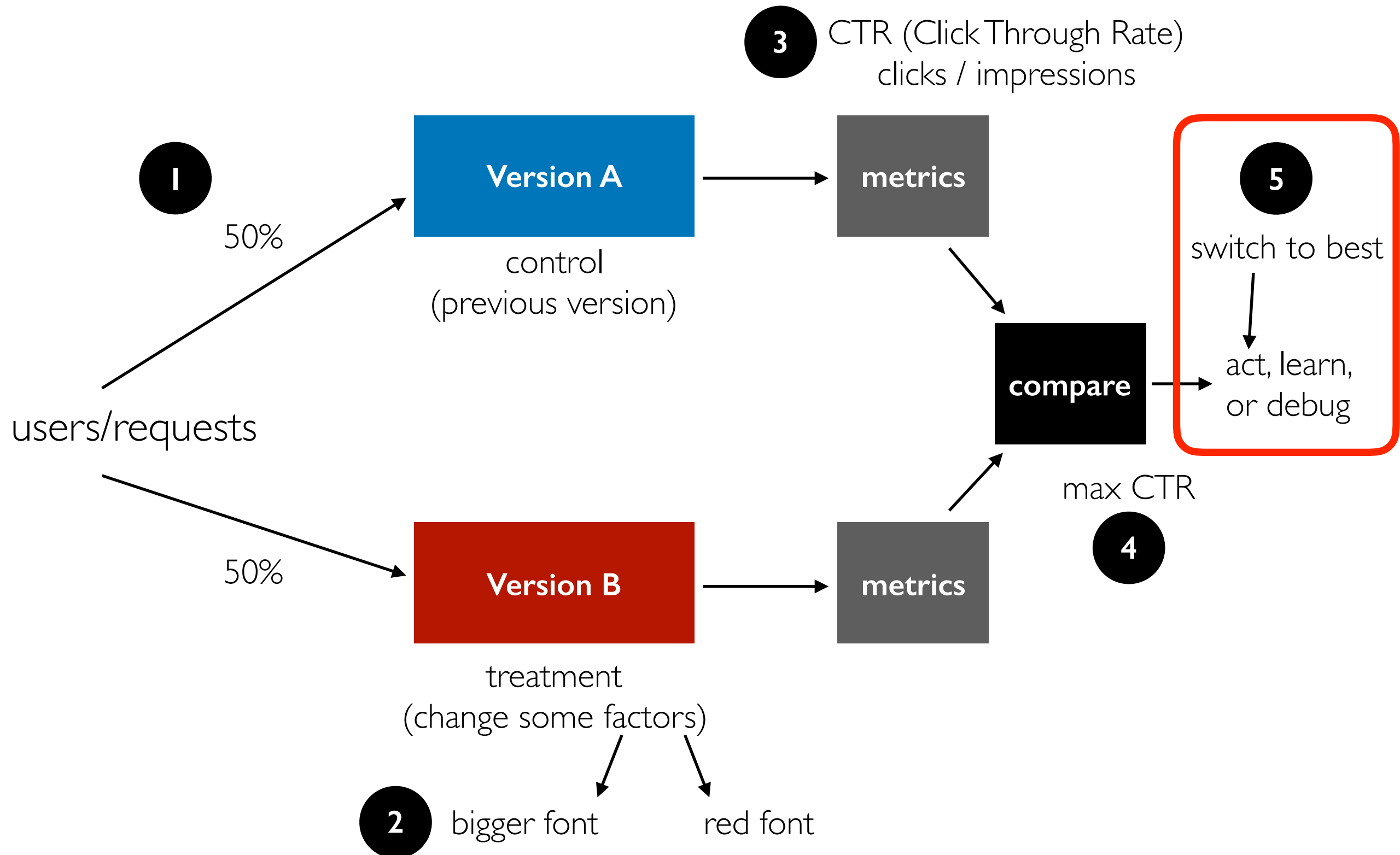
A/B Test Overview



Example I: Link to Donation Page

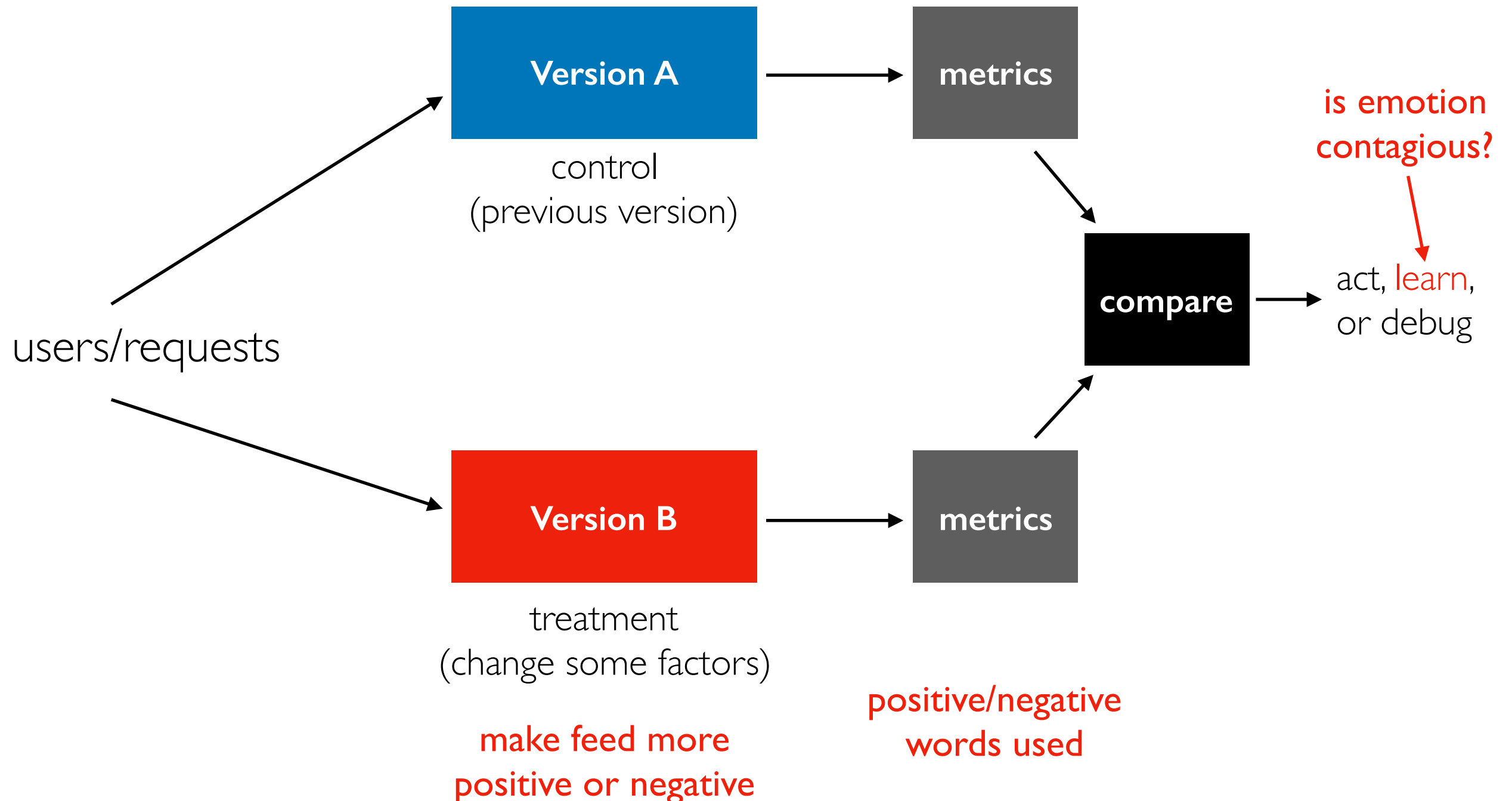


Lecture Outline



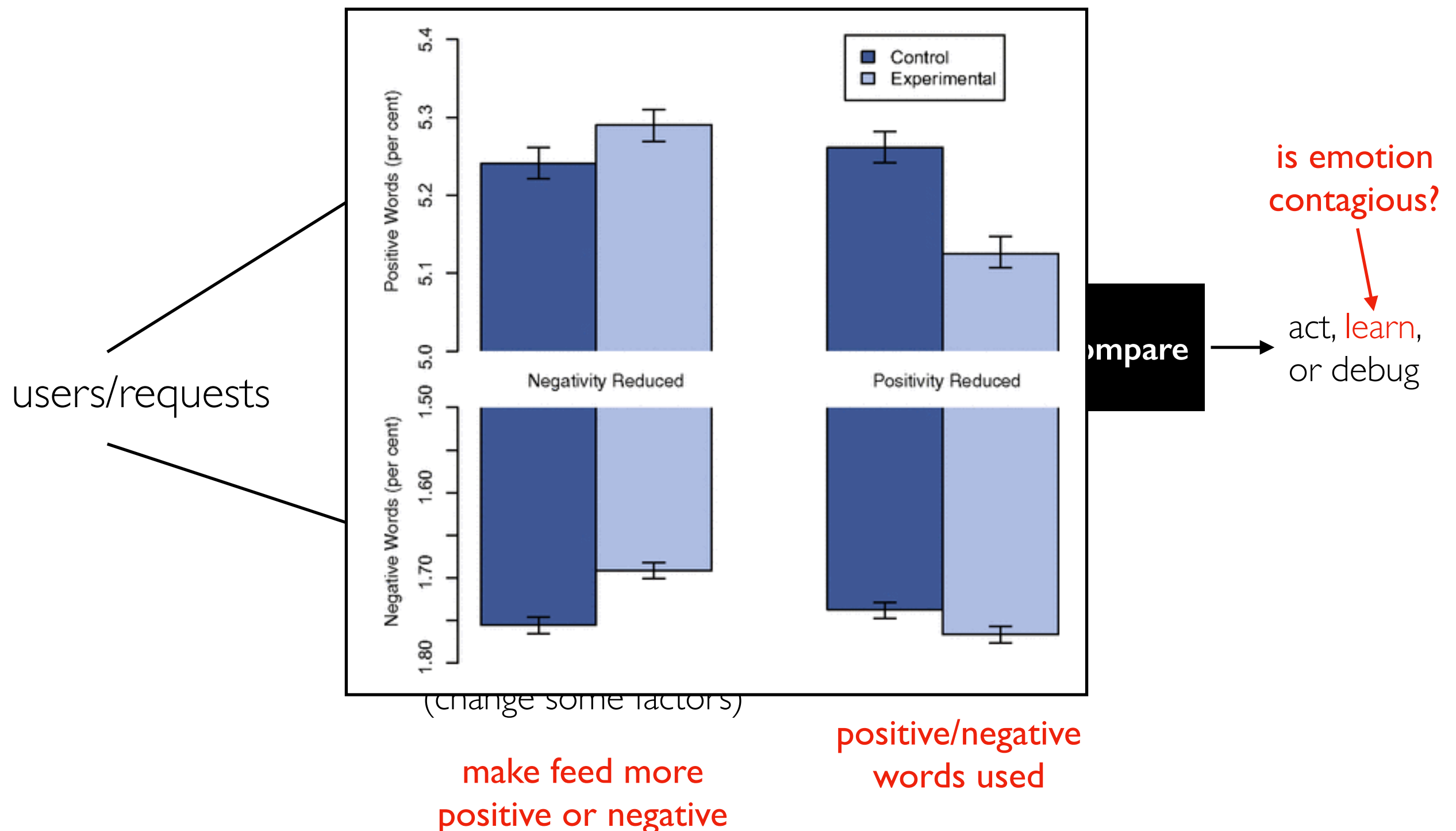
Example 2: Facebook Emotional Contagion Study

Reading: <https://techcrunch.com/2014/06/29/ethics-in-a-data-driven-world/>



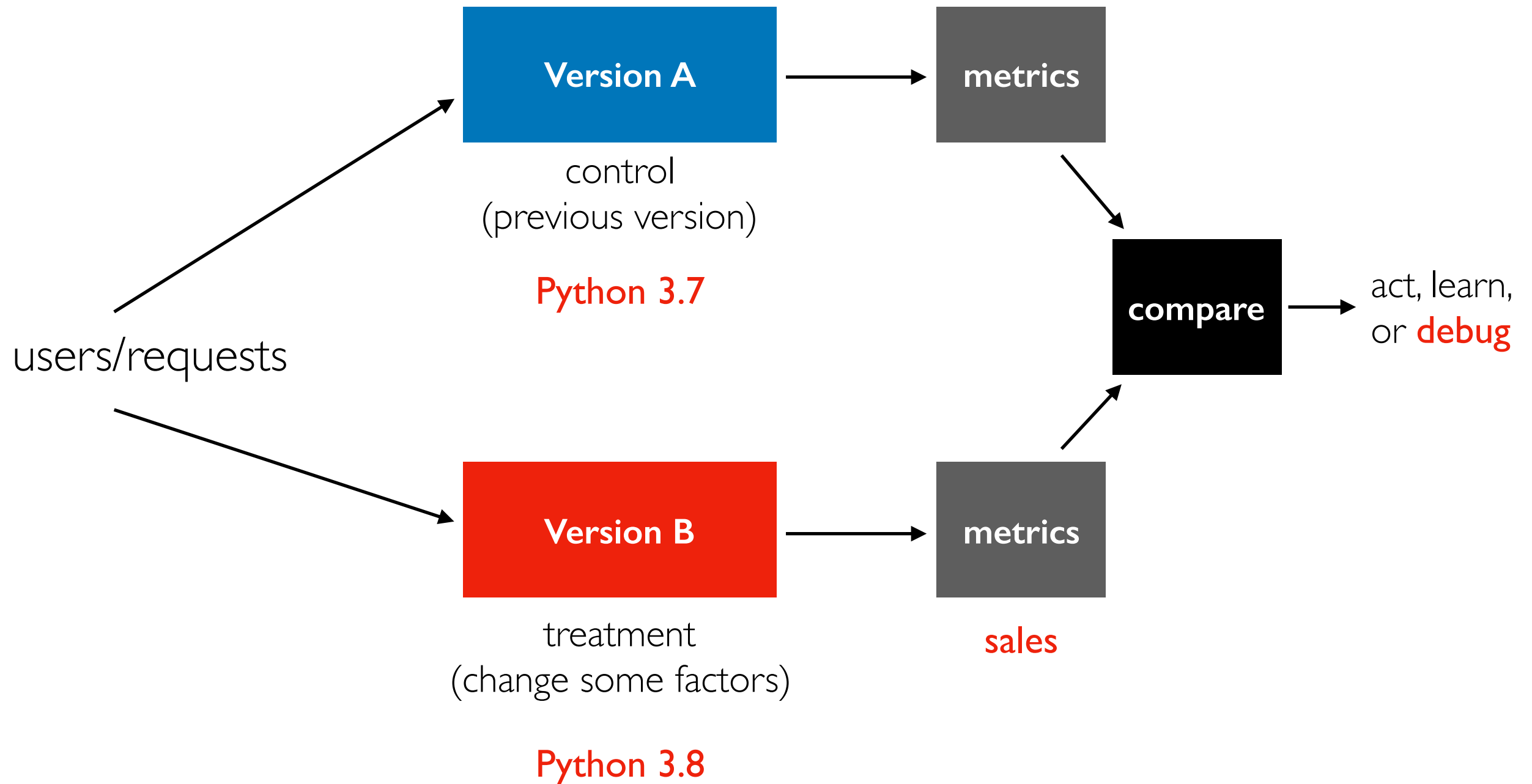
Example 2: Facebook Emotional Contagion Study

Reading: <https://techcrunch.com/2014/06/29/ethics-in-a-data-driven-world/>

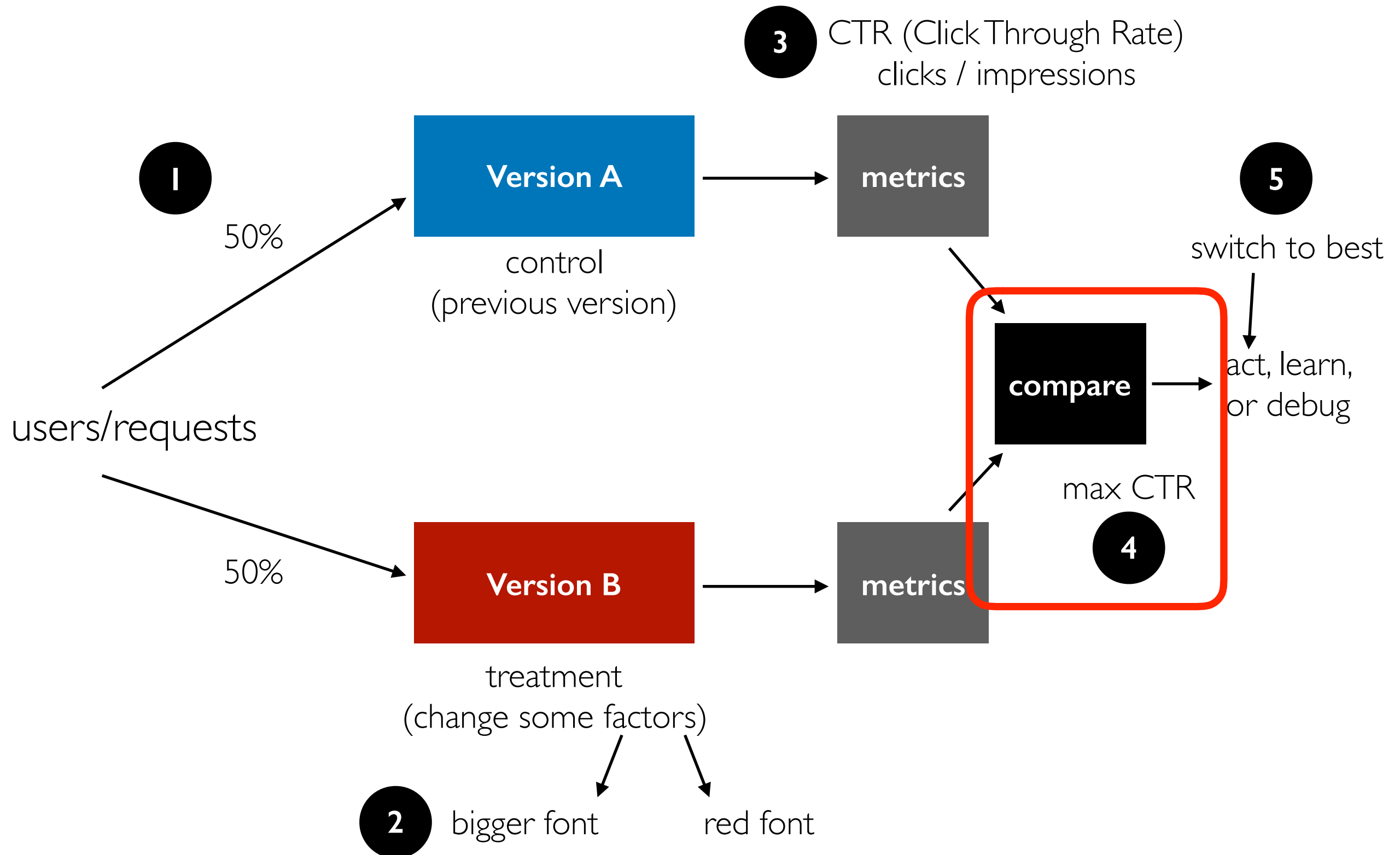


didn't need to submit to the IRB (Institutional Review Board) -- *when should it be required?*

Example 3: Update Python Version



Lecture Outline



Comparisons

Example Metric: **CTR** (Click-Through Rate)

$\text{CTR} = \text{clicks} / \text{impressions}$

"Impression" means user saw it

	click	no-click
A	12	68
B	6	14

df: contingency table

how many B **impressions** were there?
what was B's **CTR**?

Comparisons

Example Metric: **CTR** (Click-Through Rate)

$\text{CTR} = \text{clicks} / \text{impressions}$

"Impression" means user saw it

	click	no-click
A	12	68
B	6	14

df: contingency table

how many B **impressions** were there? 20
what was B's **CTR**? $6/20 = 30\%$

Comparisons

Example Metric: **CTR** (Click-Through Rate)


$\text{CTR} = \text{clicks} / \text{impressions}$

"Impression" means user saw it

	click	no-click
A	12	68
B	6	14

df: contingency table

```
1 df["click"] / (df["click"] + df["no-click"])
A    0.15
B    0.30
dtype: float64
```



is the improvement noise?

Comparisons

Example Metric: **CTR** (Click-Through Rate)

CTR = clicks / impressions

"Impression" means user saw it

	click	no-click
A	12	68
B	6	14

```
1 df["click"] / (df["click"] + df["no-click"])
```

A	0.15
B	0.30

dtype: float64

df: contingency table

pip3 install scipy

```
1 import scipy.stats as stats
2 _, pvalue = stats.fisher_exact(df)
3 pvalue
```

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.fisher_exact.html

0.1886443478471497

Comparisons

Example Metric: **CTR** (Click-Through Rate)

CTR = clicks / impressions

"Impression" means user saw it

	click	no-click
A	12	68
B	6	14

df: contingency table

```
1 import scipy.stats as stats
2 _, pvalue = stats.fisher_exact(df)
3 pvalue
```

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.fisher_exact.html

0.1886443478471497

p-value is probability of seeing a difference this extreme (or more) if both ratios were generated by the same underlying model (the one most likely to generate this)

"significant" means p-value is less than some threshold (e.g., 5%)

false positive means it is significant even though underlying model is same

Comparisons

Example Metric: **CTR** (Click-Through Rate)

CTR = clicks / impressions

"Impression" means user saw it

	click	no-click
A	12	68
B	6	14

df: contingency table

```
1 import scipy.stats as stats
2 _, pvalue = stats.fisher_exact(df)
3 pvalue
```

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.fisher_exact.html

0.1886443478471497

out of 200 neutral changes, how many will falsely show up as significant if we set our p-value threshold to 5%?

p-value is probability of seeing a difference this extreme (or more) if both ratios were generated by the same underlying model (the one most likely to generate this)

"significant" means p-value is less than some threshold (e.g., 5%)

false positive means it is significant even though underlying model is same

Comparisons

Example Metric: **CTR** (Click-Through Rate)

CTR = clicks / impressions

"Impression" means user saw it

	click	no-click
A	12	68
B	6	14

df: contingency table

```
1 import scipy.stats as stats
2 _, pvalue = stats.fisher_exact(df)
3 pvalue
```

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.fisher_exact.html

0.1886443478471497

*out of 200 neutral changes, how many
will falsely show up as significant if we
set our p-value threshold to 5%?*

10

p-value is probability of seeing a difference
this extreme (or more) if both ratios were
generated by the same underlying model
(the one most likely to generate this)

"significant" means p-value is less
than some threshold (e.g., 5%)

false positive means it is significant
even though underlying model is same

Comparisons

out of 200 neutral changes, how many will falsely show up as significant if we set our p-value threshold to 5%?

10

Example Metric: **CTR** (Click-Through Rate)

CTR = clicks / impressions

"Impression" means user saw it

occasionally run A/A tests

	click	no-click
A	12	68
B	6	14

df: contingency table

```
1 import scipy.stats as stats
2 _, pvalue = stats.fisher_exact(df)
3 pvalue
```

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.fisher_exact.html

0.1886443478471497

Comparisons

Example Metric: **CTR** (Click-Through Rate)

$\text{CTR} = \text{clicks} / \text{impressions}$

"Impression" means user saw it

	click	no-click
A	12	68
B	6	14

df: contingency table

3 outcomes, based on CTRs and significance

- A is significantly better
- B is significantly better
- *neither wins*

what to do?

- ideas???

Comparisons

Example Metric: **CTR** (Click-Through Rate)

$\text{CTR} = \text{clicks} / \text{impressions}$

"Impression" means user saw it

	click	no-click
A	12	68
B	6	14

df: contingency table

3 outcomes, based on CTRs and significance

- A is significantly better
- B is significantly better
- *neither wins*



what to do?

- collect more data
- ignore significance, just look at CTR
(indecision may be the worst decision)
- choose previous version A (probably fewer bugs)
- choose new version B (for simplicity or other merits)

Which Version Has Higher Whole-page CTR?

Version A

amazon

ALL SHOPPING IMAGES VIDEOS MAPS NEWS

196,000,000 Results Any time ▾

[Amazon.com - Amazon.com® Official Site](#)
<https://www.amazon.com> ▾

(Ad) Earth's biggest selection of books, electronics, apparel & more at low prices.
amazon.com has been visited by 1M+ users in the past month
Fast Shipping · Explore Amazon Devices · Shop Prime Wardrobe · Try Prime for Free

[Shop Echo & Alexa Devices](#)
Play music, get news, control your smart home & more using your voice.

[Learn More About Alexa](#)
Hands-free voice control for music, calling, smart home devices & more.

[Meet the Fire TV Family](#)
See our devices for streaming your favorite content and live TV.
[See results only from amazon.com](#)

[Amazon.com: Online Shopping for Electronics, Apparel ...](#)
<https://www.amazon.com> ▾

Free One-Day Delivery on millions of items with Prime. Low prices across earth's biggest selection of books, music, DVDs, electronics, computers, software, apparel & accessories, shoes, jewelry, tools & hardware, housewares, furniture, sporting goods, beauty & ...

5/5 ★★★★★ (1) Price: \$21.06

[Sign In](#)
This site won't let us show the description for ...
[How to Use Account Switching](#)

[Books](#)
Books at Amazon. The Amazon.com Books homepage helps you explore Earth's Biggest ...

[See more](#) ▾

Version B

amazon

ALL SHOPPING IMAGES VIDEOS MAPS NEWS

196,000,000 Results Any time ▾

[Amazon.com - Amazon.com® Official Site](#)
<https://www.amazon.com> ▾

(Ad) Earth's biggest selection of books, electronics, apparel & more at low prices.
amazon.com has been visited by 1M+ users in the past month
Fast Shipping · Explore Amazon Devices · Shop Prime Wardrobe · Try Prime for Free

[Shop Echo & Alexa Devices](#)
Play music, get news, control your smart home & more using your voice.

[Learn More About Alexa](#)
Hands-free voice control for music, calling, smart home devices & more.

[Meet the Fire TV Family](#)
See our devices for streaming your favorite content and live TV.
[See results only from amazon.com](#)

[Amazon.com: Online Shopping for Electronics, Apparel ...](#)
<https://www.amazon.com> ▾

Free One-Day Delivery on millions of items with Prime. Low prices across earth's biggest selection of books, music, DVDs, electronics, computers, software, apparel & accessories, shoes, jewelry, tools & hardware, housewares, furniture, sporting goods, beauty & ...

5/5 ★★★★★ (1) Price: \$21.06

[Sign In](#)
This site won't let us show the description for ...
[How to Use Account Switching](#)

[Books](#)
Books at Amazon. The Amazon.com Books homepage helps you explore Earth's Biggest ...

[See more](#) ▾

Which Version Has Higher Whole-page CTR?

Version A

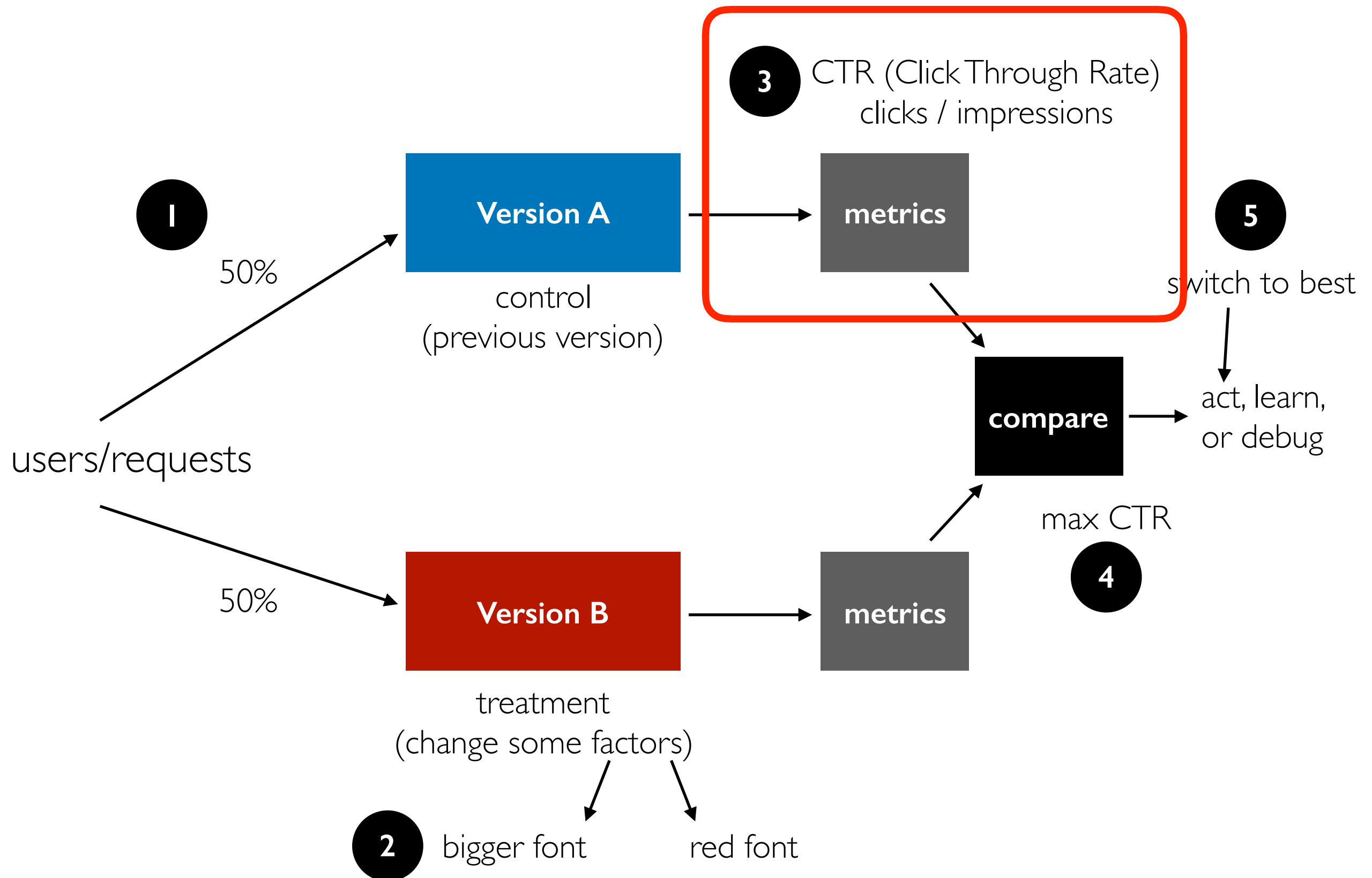
Version A search results for 'amazon' show a clean layout with a search bar at the top. Below the search bar, there are tabs for 'ALL', 'SHOPPING', 'IMAGES', 'VIDEOS', 'MAPS', and 'NEWS'. The 'ALL' tab is selected. The results show 196,000,000 Results and 'Any time' filter. The first result is 'Amazon.com - Amazon.com® Official Site' with a URL 'https://www.amazon.com'. Below the title, there is an advertisement snippet: 'Earth's biggest selection of books, electronics, apparel & more at low prices. amazon.com has been visited by 1M+ users in the past month. Fast Shipping · Explore Amazon Devices · Shop Prime Wardrobe · Try Prime for Free'. Below the ad, there are four links: 'Shop Echo & Alexa Devices', 'Amazon Prime Benefits', 'Learn More About Alexa', and 'Shop Amazon Fire Tablets'. Each link has a short description. At the bottom, there is a link 'Meet the Fire TV Family' and a link 'See results only from amazon.com'. Below that, there is a link 'Amazon.com: Online Shopping for Electronics, Apparel ...' with a URL 'https://www.amazon.com'. Below the title, there is a description: 'Free One-Day Delivery on millions of items with Prime. Low prices across earth's biggest selection of books, music, DVDs, electronics, computers, software, apparel & accessories, shoes, jewelry, tools & hardware, housewares, furniture, sporting goods, beauty & ...'. Below the description, there is a rating '5/5 ★★★★★ (1)' and a price 'Price: \$21.06'. At the bottom, there are two links: 'Sign In' and 'Books'. Each link has a short description. Below the 'Sign In' link, there is a link 'How to Use Account Switching'. At the very bottom, there is a link 'See more' with a downward arrow.

Version B

Version B search results for 'amazon' show a clean layout with a search bar at the top. Below the search bar, there are tabs for 'ALL', 'SHOPPING', 'IMAGES', 'VIDEOS', 'MAPS', and 'NEWS'. The 'ALL' tab is selected. The results show 196,000,000 Results and 'Any time' filter. The first result is 'Amazon.com - Amazon.com® Official Site' with a URL 'https://www.amazon.com'. Below the title, there is an advertisement snippet: 'Earth's biggest selection of books, electronics, apparel & more at low prices. amazon.com has been visited by 1M+ users in the past month. Fast Shipping · Explore Amazon Devices · Shop Prime Wardrobe · Try Prime for Free'. Below the ad, there are four links: 'Shop Echo & Alexa Devices', 'Amazon Prime Benefits', 'Learn More About Alexa', and 'Shop Amazon Fire Tablets'. Each link has a short description. At the bottom, there is a link 'Meet the Fire TV Family' and a link 'See results only from amazon.com'. Below that, there is a link 'Amazon.com: Online Shopping for Electronics, Apparel ...' with a URL 'https://www.amazon.com'. Below the title, there is a description: 'Free One-Day Delivery on millions of items with Prime. Low prices across earth's biggest selection of books, music, DVDs, electronics, computers, software, apparel & accessories, shoes, jewelry, tools & hardware, housewares, furniture, sporting goods, beauty & ...'. Below the description, there is a rating '5/5 ★★★★★ (1)' and a price 'Price: \$21.06'. At the bottom, there are two links: 'Sign In' and 'Books'. Each link has a short description. Below the 'Sign In' link, there is a link 'How to Use Account Switching'. At the very bottom, there is a link 'See more' with a downward arrow.

Lesson: metrics should inform humans, not directly determine decisions

Lecture Outline



Metrics

Things to measure:

- clicks -- when are they bad?

Metrics

Things to measure:

- clicks
- scroll (did they read it?)
- subscribe/unsubscribe
- other ideas? discuss with your neighbour

Metrics

Things to measure:

- clicks
- scroll (did they read it?)
- subscribe/unsubscribe
- purchases/returns
- hover (did they think about it?)
- shares
- likes/upvotes
- comments

Metrics

Things to measure:

- clicks
- scroll (did they read it?)
- subscribe/unsubscribe
- purchases/returns
- hover (did they think about it?)
- shares
- likes/upvotes
- comments

combos: Bing measures how often people click a result link and don't hit back within 30 seconds

Metrics

Things to measure:

- clicks
- scroll (did they read it?)
- subscribe/unsubscribe
- purchases/returns
- hover (did they think about it?)
- shares
- likes/upvotes
- comments

combos: Bing measures how often people click a result link and don't hit back within 30 seconds

what is the effect of B?

B is **send twice as many spammy emails**

Metrics

Things to measure:

- clicks
- scroll (did they read it?)
- subscribe/unsubscribe
- purchases/returns
- hover (did they think about it?)
- shares
- likes/upvotes
- comments

combos: Bing measures how often people click a result link and don't hit back within 30 seconds

what is the effect of B?

B is **send twice as many spammy emails**

Metrics

Things to measure:

- clicks
- scroll (did they read it?)
- subscribe/unsubscribe
- purchases/returns
- hover (did they think about it?)
- shares
- likes/upvotes
- comments

combos: Bing measures how often people click a result link and don't hit back within 30 seconds

what is the effect of B?

B is **remove price from product page link**

what is the effect of B?

B is **send twice as many spammy emails**

Metrics

Things to measure:

- clicks
- scroll (did they read it?)
- subscribe/unsubscribe
- purchases/returns
- hover (did they think about it?)
- shares
- likes/upvotes
- comments

combos: Bing measures how often people click a result link and don't hit back within 30 seconds

what is the effect of B?

B is **remove price from product page link**

Lesson: it's easy to shift clicks

what is the effect of B?

B is **send twice as many spammy emails**

Lesson: it's hard to measure long-term effects (noisy!), so use common sense

Metrics

Things to measure:

- clicks
- scroll (did they read it?)
- subscribe/unsubscribe
- purchases/returns
- hover (did they think about it?)
- shares
- likes/upvotes
- comments

combos: Bing measures how often people click a result link and don't hit back within 30 seconds

what is the effect of B?

B is **remove price from product page link**

what is the effect of B?

B is **send twice as many spammy emails**

Decide beforehand on one **OEC** metric: Overall Experiment Criterion

- Bing has thousands of debug metrics, but only 4 OECs.

Metrics

Things to measure:

- clicks
- scroll (did they read it?)
- subscribe/unsubscribe
- purchases/returns
- hover (did they think about it?)
- shares
- likes/upvotes
- comments

combos: Bing measures how often people click a result link and don't hit back within 30 seconds

what is the effect of B?

B is **send twice as many spammy emails**

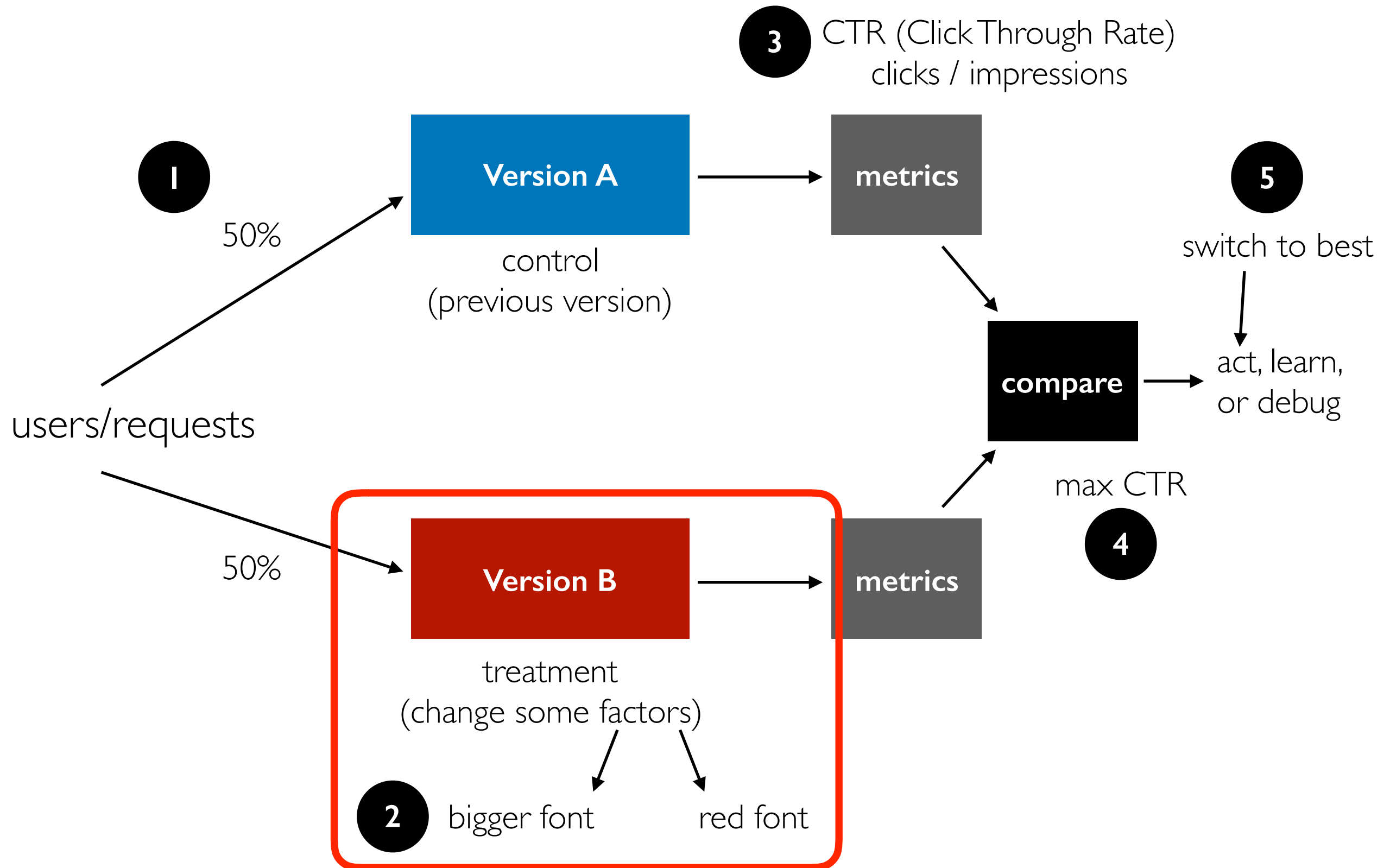
what is the effect of B?

B is **remove price from product page link**

Decide beforehand on one **OEC** metric: Overall Experiment Criterion

- Bing has thousands of debug metrics, but only 4 OECs. Try to consider cost as well as benefit!
- As a rule of thumb, *"if you make something bigger, more people will click on it"* ~ Ron Kovani
- Making part of the site better could hurt other parts if you have a naive OEC

Lecture Outline



Lecture Outline

