The Science of Product Development

Bringing Causal Inference to Conversion and Retention Metrics

David Robinson, Director of Data Science at Contentsquare
2024 NYR Conference

Source: The Uncomfortable by Katerina Kamprani

When a company designs a product, they want it to be easy to use

Some are designed better than others!

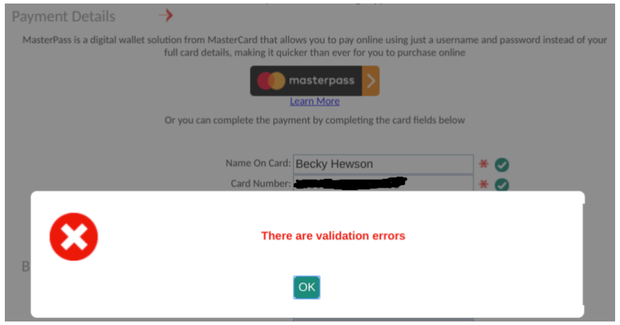
Digital products have an advantage that designers of physical products can only dream of

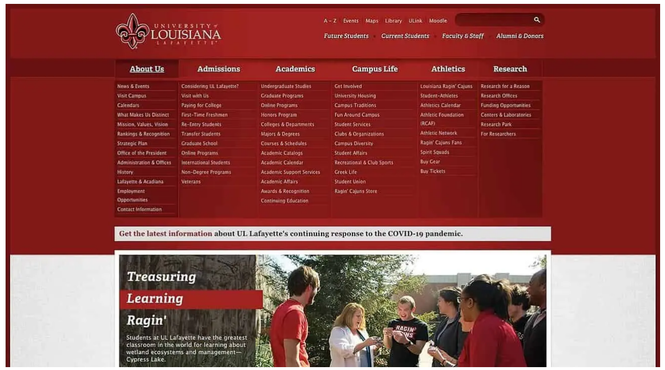
They can collect *data* on how every single customer is using the product: what they're viewing, what they're clicking, and where they're getting stuck.

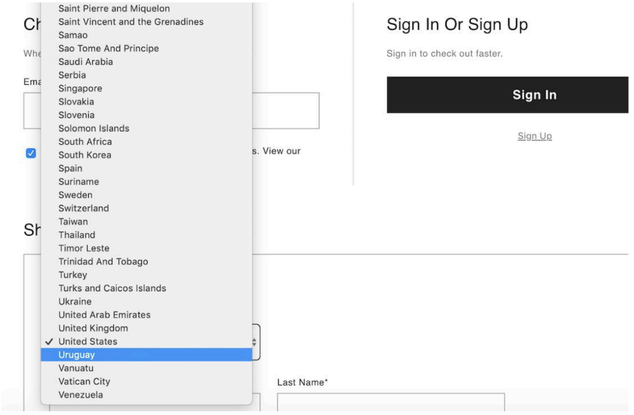
Digital products have an advantage that designers of physical products can only dream of

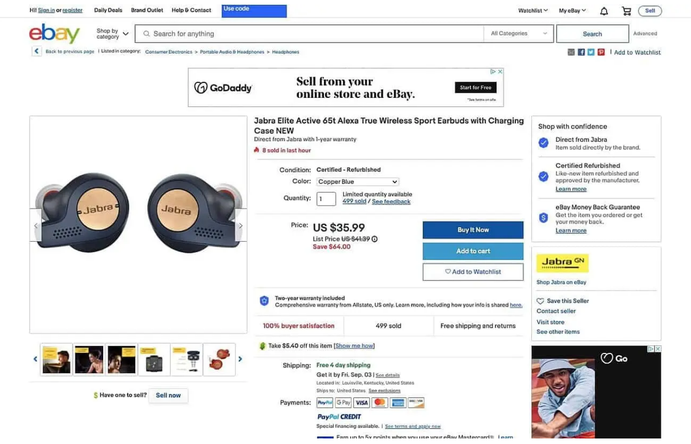
They can collect *data* on how every single customer is using the product: what they're viewing, what they're clicking, and where they're getting stuck.

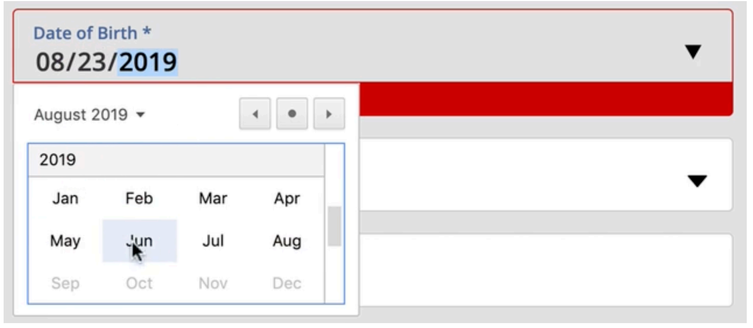
So what's their excuse?











Land on /home
Click "Special Offers"
View /special-offers
Click "Paperclips"
View /paper-clips
Click "Add to Cart"
Click "Check Out"
View /checkout
Click "Save for Later"
Drop Off

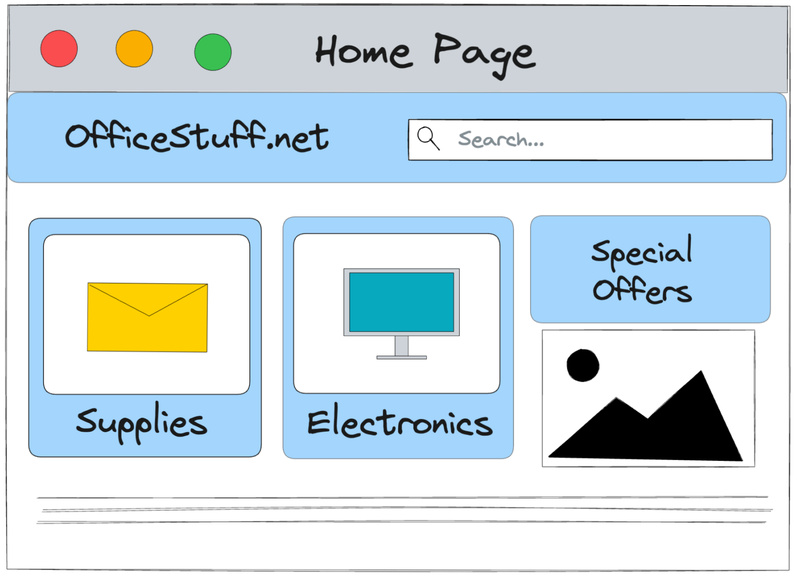
Land on /home
Click "Special Offers"
View /special-offers
Click Back
Click "Search"
Enter "Headphones"
Click - Sort by Price
Click - Sort by Category
Drop Off

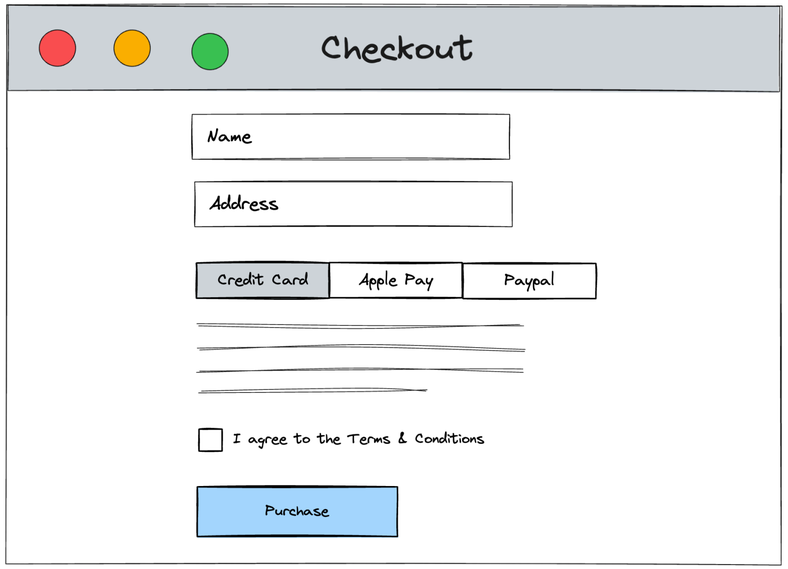
Land on /promo
Click "50% Off Special Offers"
View /special-offers
Click Back
Click - Sign In
Click "Search"
Enter "Keyboard"
Click - Sort by Price
Click - Filter by Price
Click - Mechanical Keyboard
View /products/mechanical-keyboard
Click - "Buy (One Click)"
Purchase

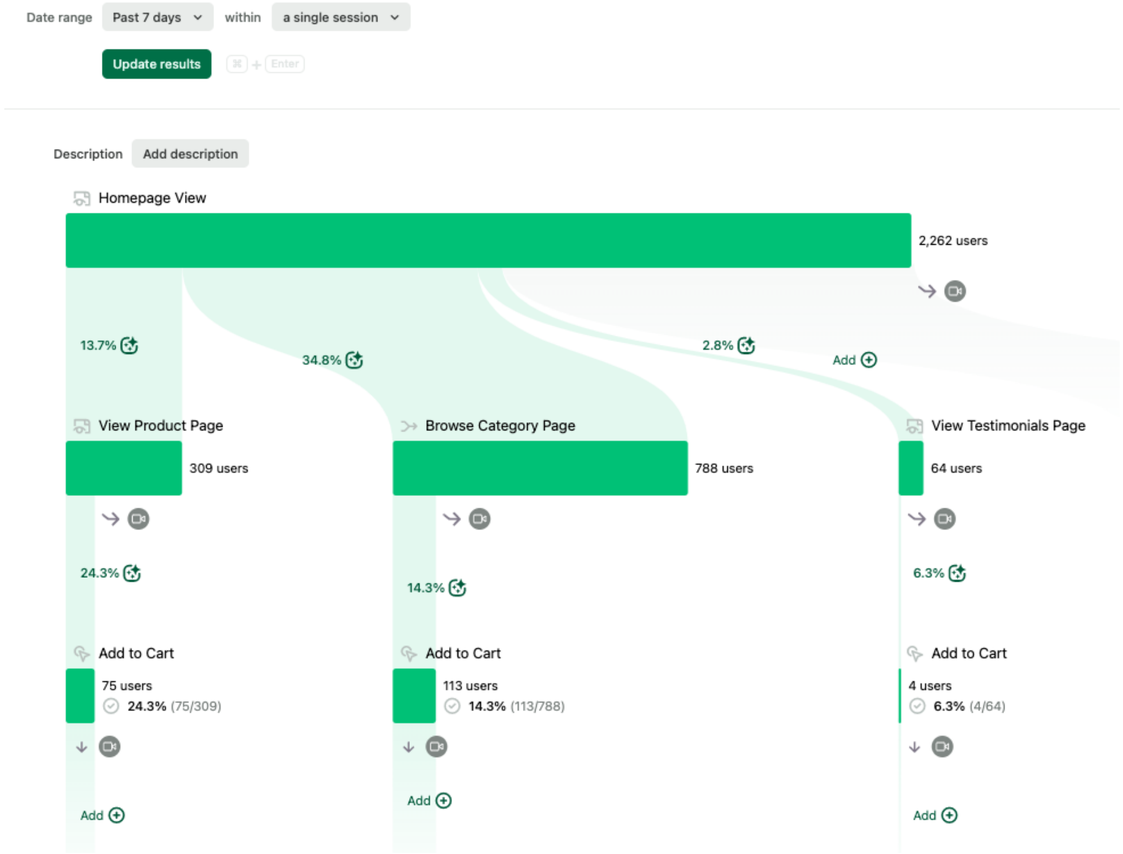
Land on /home
Click "Browse"
View /categories
View /recent
Click Back
Drop Off

View /home
Click "Supplies"
View /supplies
Click "Paper"
View /products/paper
Click "Add to Cart"
View /supplies
Click "Scissors"
View /products/scissors
Click Back
Click "Premium Scissors"
View /products/scissors
Click "Add to Cart"
View - Checkout
Click - Enter Name
Click - Enter Credit Card
Validation Error
Click - Enter Credit Card
Click - Check Terms & Conditions
Purchase

What statistical tools can turn sequences of events into **product insights and improvements**?



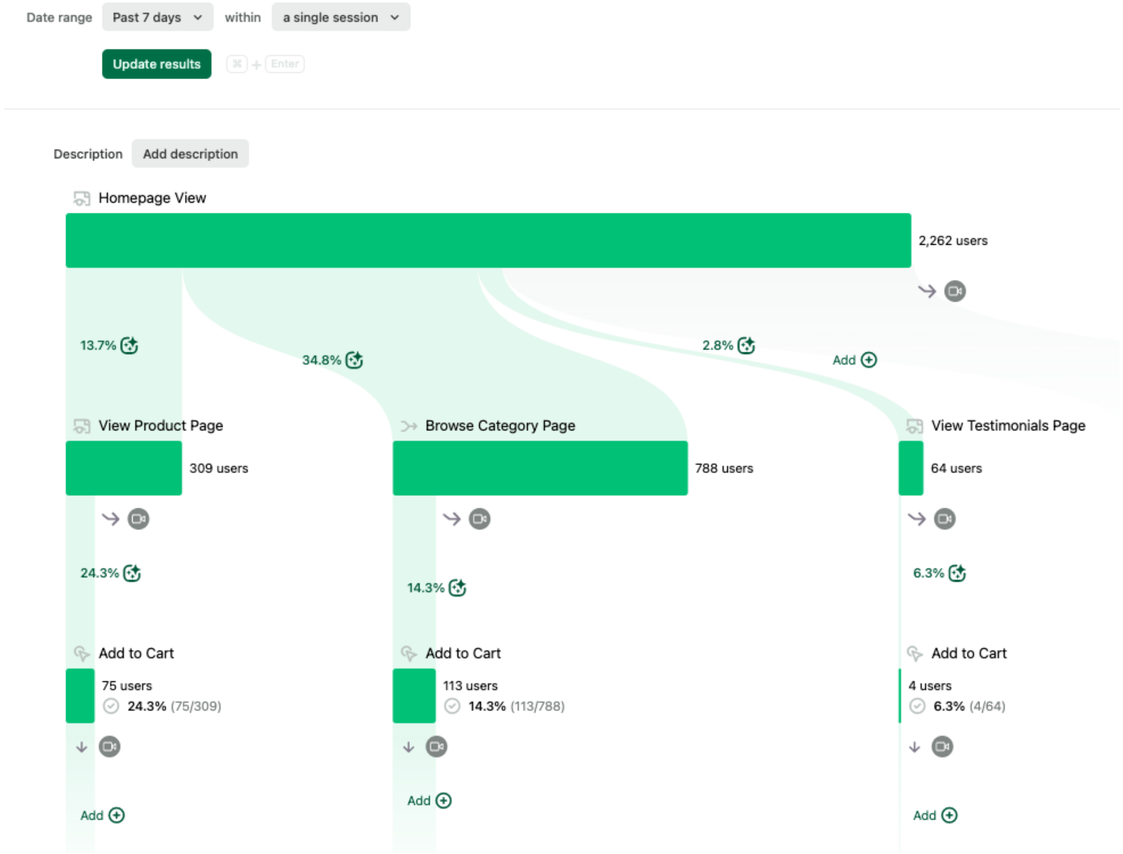




Heap

Contentsquare

We developed these principles at Heap to power the Journeys product (launched 2022)

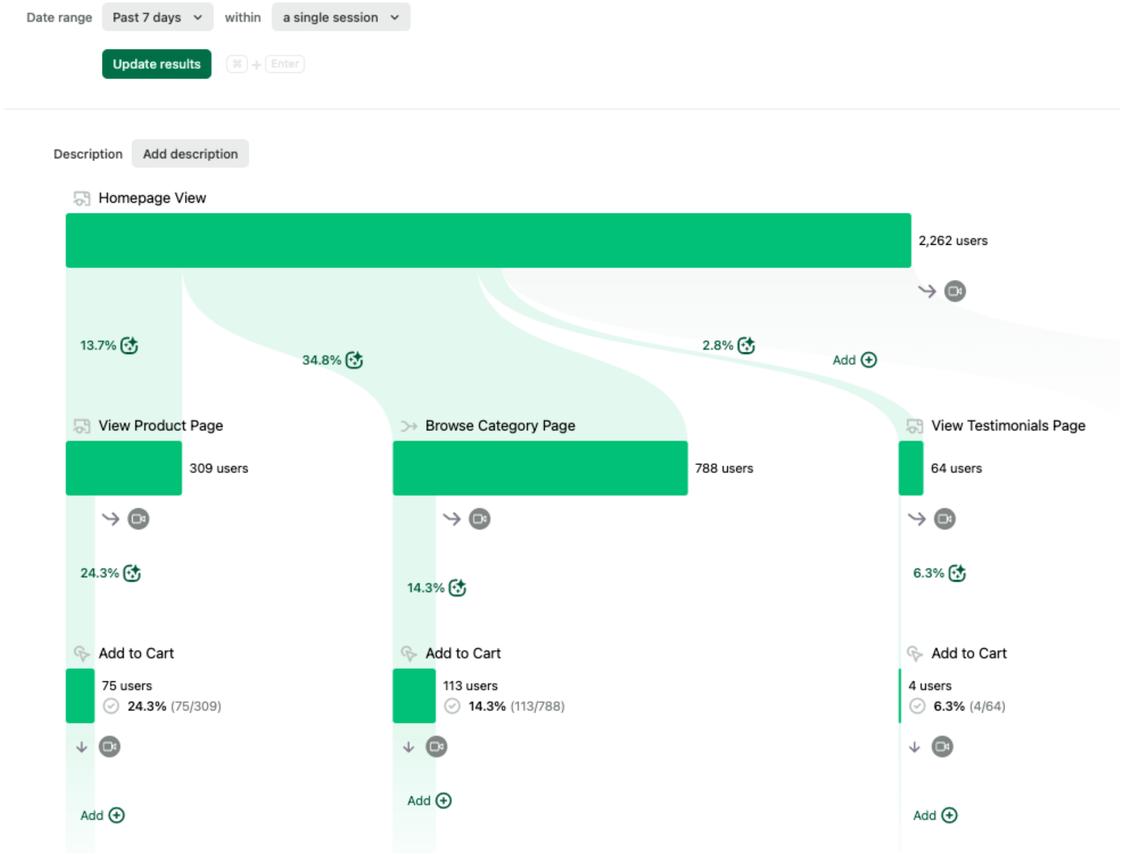


Today, I'm focused more on the theory: on the "missing science" of analyzing and improving products

Heap

Contentsquare

We developed these principles at Heap to power the Journeys product (launched 2022)

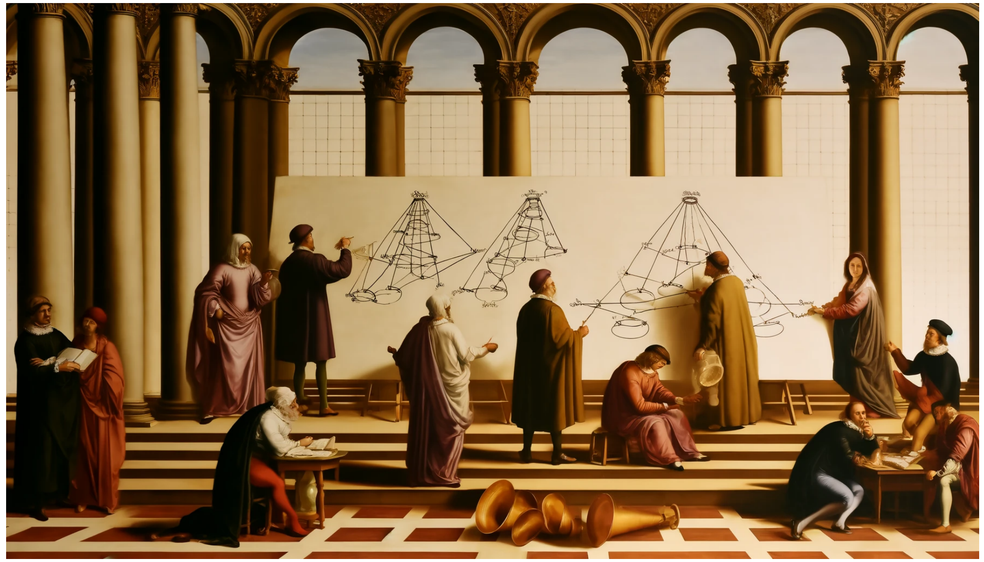


Heap

Contentsquare

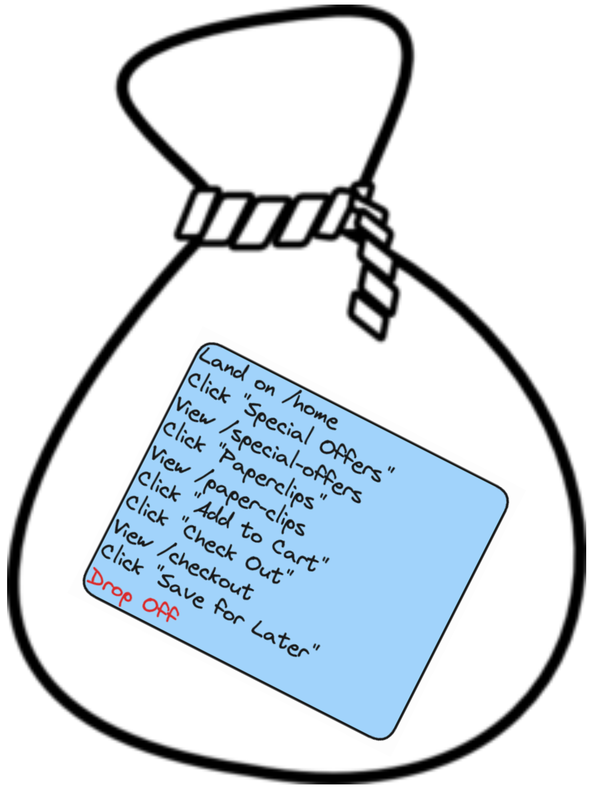
We developed these principles at Heap to power the Journeys product (launched 2022)

Today, I'm focused more on the theory: on the "missing science" of analyzing and improving products

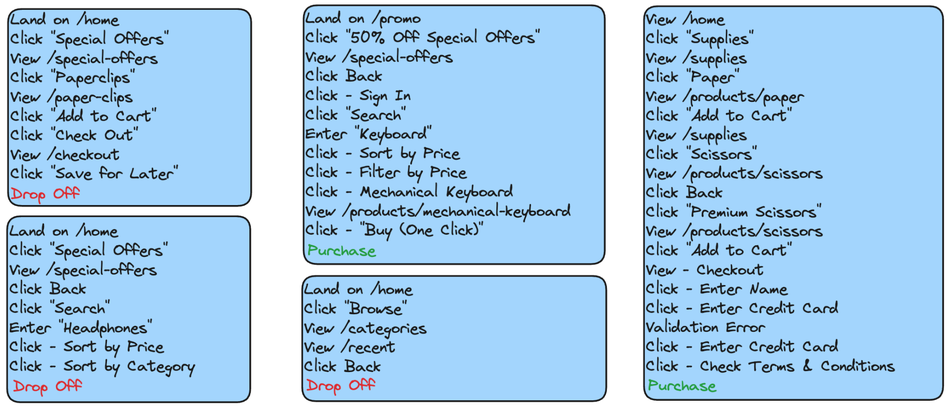


Correlations

The “bag of clicks” model



The “bag of clicks” model



The “bag of clicks” model

user_id	viewed_home	clicked_search	clicked_read_reviews	clicked_learn_more	...yada yada...	purchased
187180	0	1	1	0	...	1
662863	0	0	0	0	...	1
684172	1	1	1	0	...	0
261982	1	1	0	0	...	0
579325	1	1	1	1	...	0
829755	1	1	0	1	...	0
454574	1	1	0	0	...	0
216105	1	1	1	1	...	1
181215	0	1	0	0	...	0
996510	1	1	0	0	...	0
379967	1	1	0	0	...	0
511056	1	1	0	1	...	0
948165	1	1	1	0	...	0
575719	1	1	0	1	...	0
356383	1	1	1	0	...	0
582609	1	1	1	0	...	0
988781	1	0	0	1	...	0

The “bag of clicks” model

Land on /home
Click "Special Offers"
View /special-offers
Click "Paperclips"
View /paper-clips
Click "Add to Cart"
Click "Check Out"
View /checkout
Click "Save for Later"
Drop Off

Land on /home
Click "Special Offers"
View /special-offers
Click Back
Click "Search"
Enter "Headphones"
Click - Sort by Price
Click - Sort by Category
Drop Off

Land on /promo
Click "50% Off Special Offers"
View /special-offers
Click Back
Click - Sign In
Click "Search"
Enter "Keyboard"
Click - Sort by Price
Click - Filter by Price
Click - Mechanical Keyboard
View /products/mechanical-keyboard
Click - "Buy (One Click)"
Purchase

Land on /home
Click "Browse"
View /categories
View /recent
Click Back
Drop Off

View /home
Click "Supplies"
View /supplies
Click "Paper"
View /products/paper
Click "Add to Cart"
View /supplies
Click "Scissors"
View /products/scissors
Click Back
Click "Premium Scissors"
View /products/scissors
Click "Add to Cart"
View - Checkout
Click - Enter Name
Click - Enter Credit Card
Validation Error
Click - Enter Credit Card
Click - Check Terms & Conditions
Purchase

“Users viewed these pages and did these clicks…”

user_id	viewed_home	clicked_search	clicked_read_reviews	clicked_learn_more	…yada yada…	purchased
187180	0	1	1	0	…	1
662863	0	0	0	0	…	1
684172	1	1	1	0	…	0
261982	1	1	0	0	…	0
579325	1	1	1	1	…	0
829755	1	1	0	1	…	0
454574	1	1	0	0	…	0
216105	1	1	1	1	…	1
181215	0	1	0	0	…	0
996510	1	1	0	0	…	0
379967	1	1	0	0	…	0
511056	1	1	0	1	…	0
948165	1	1	1	0	…	0
575719	1	1	0	1	…	0
356383	1	1	1	0	…	0
582609	1	1	1	0	…	0
988781	1	0	0	1	…	0

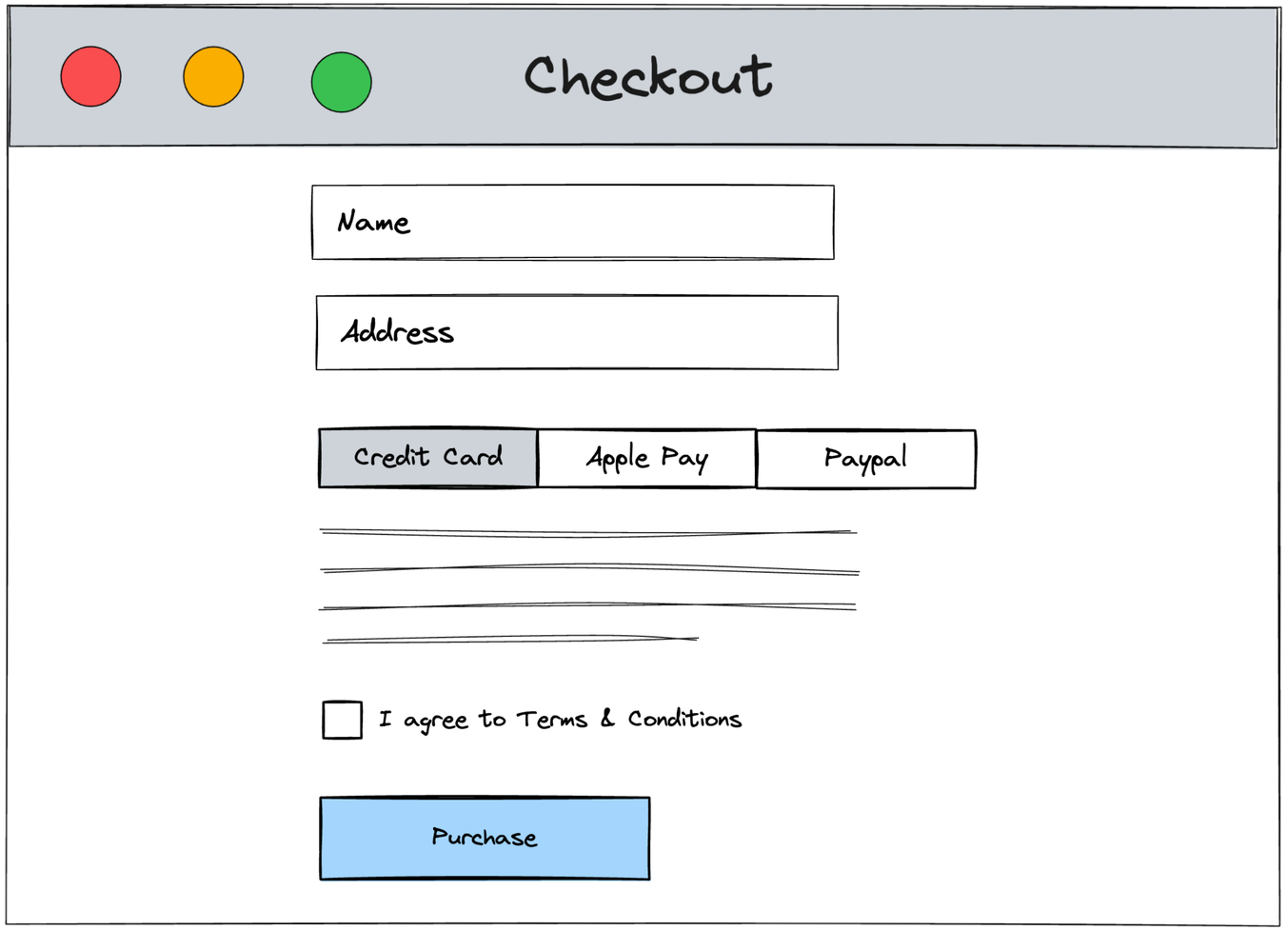
“…which of those led to **purchases**?”

user_id	viewed_home	clicked_search	clicked_read_reviews	clicked_learn_more	...yada yada...	purchased
187180	0	1	1	0	...	1
662863	0	0	0	0	...	1
684172	1	1	1	0	...	0
261982	1	1	0	0	...	0
579325	1	1	1	1	...	0
829755	1	1	0	1	...	0
454574	1	1	0	0	...	0
216105	1	1	1	1	...	1
181215	0	1	0	0	...	0
996510	1	1	0	0	...	0
379967	1	1	0	0	...	0
511056	1	1	0	1	...	0
948165	1	1	1	0	...	0
575719	1	1	0	1	...	0
356383	1	1	1	0	...	0
582609	1	1	1	0	...	0
988781	1	0	0	1	...	0
857058	1	1	0	0	...	0
498438	1	1	0	0	...	0
470140	1	0	1	0	...	0
714748	1	0	0	1	...	0
624947	1	1	1	1	...	0
524455	1	0	0	1	...	1

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

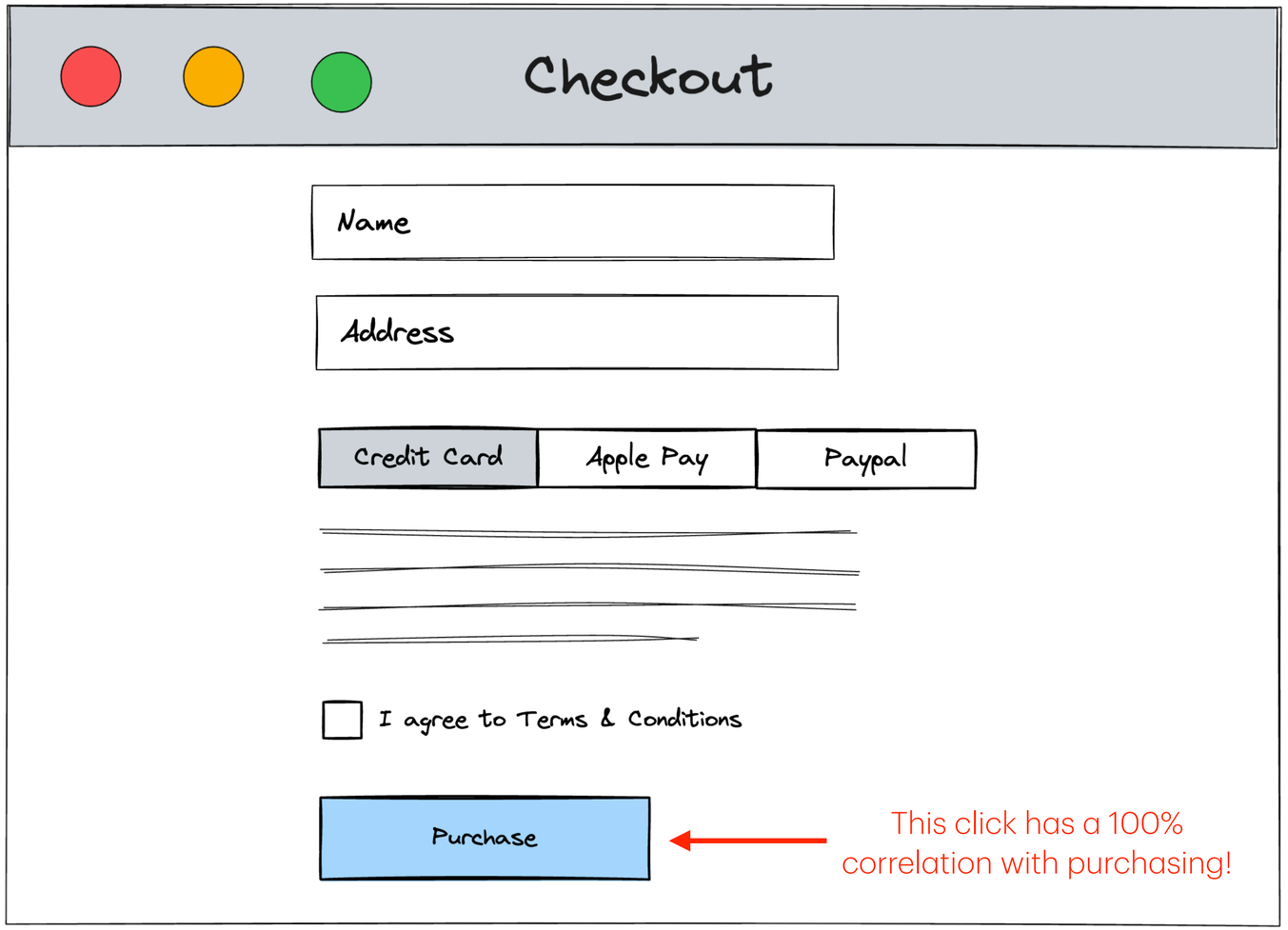
user_id	viewed_home	clicked_search	clicked_read_reviews	clicked_learn_more	...yada yada...	purchased
187180	0	1	1	0	...	1
662863	0	0	0	0	...	1
684172	1	1	1	0	...	0
261982	1	1	0	0	...	0
579325	1	1	1	1	...	0
829755	1	1	0	1	...	0
454574	1	1	0	0	...	0
216105	1	1	1	1	...	1
181215	0	1	0	0	...	0
996510	1	1	0	0	...	0
379967	1	1	0	0	...	0
511056	1	1	0	1	...	0
948165	1	1	1	0	...	0
575719	1	1	0	1	...	0
356383	1	1	1	0	...	0
582609	1	1	1	0	...	0
988781	1	0	0	1	...	0
857058	1	1	0	0	...	0
498438	1	1	0	0	...	0
470140	1	0	1	0	...	0
714748	1	0	0	1	...	0
624947	1	1	1	1	...	0
524455	1	0	0	1	...	1

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$



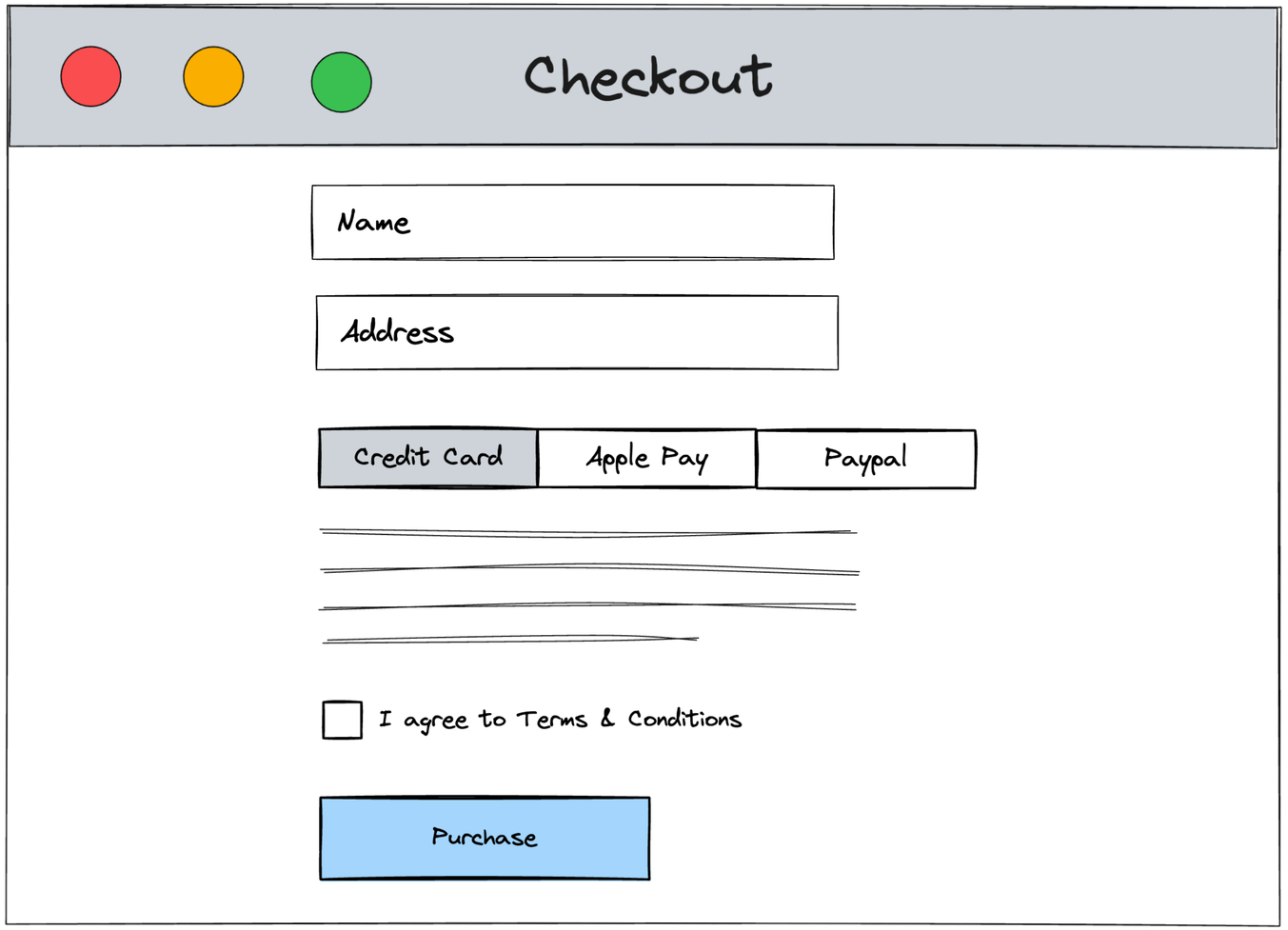
user_id	viewed_home	clicked_search	clicked_read_reviews	clicked_learn_more	...yada yada...	purchased
187180	0	1	1	0	...	1
662863	0	0	0	0	...	1
684172	1	1	1	0	...	0
261982	1	1	0	0	...	0
579325	1	1	1	1	...	0
829755	1	1	0	1	...	0
454574	1	1	0	0	...	0
216105	1	1	1	1	...	1
181215	0	1	0	0	...	0
996510	1	1	0	0	...	0
379967	1	1	0	0	...	0
511056	1	1	0	1	...	0
948165	1	1	1	0	...	0
575719	1	1	0	1	...	0
356383	1	1	1	0	...	0
582609	1	1	1	0	...	0
988781	1	0	0	1	...	0
857058	1	1	0	0	...	0
498438	1	1	0	0	...	0
470140	1	0	1	0	...	0
714748	1	0	0	1	...	0
624947	1	1	1	1	...	0
524455	1	0	0	1	...	1

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$



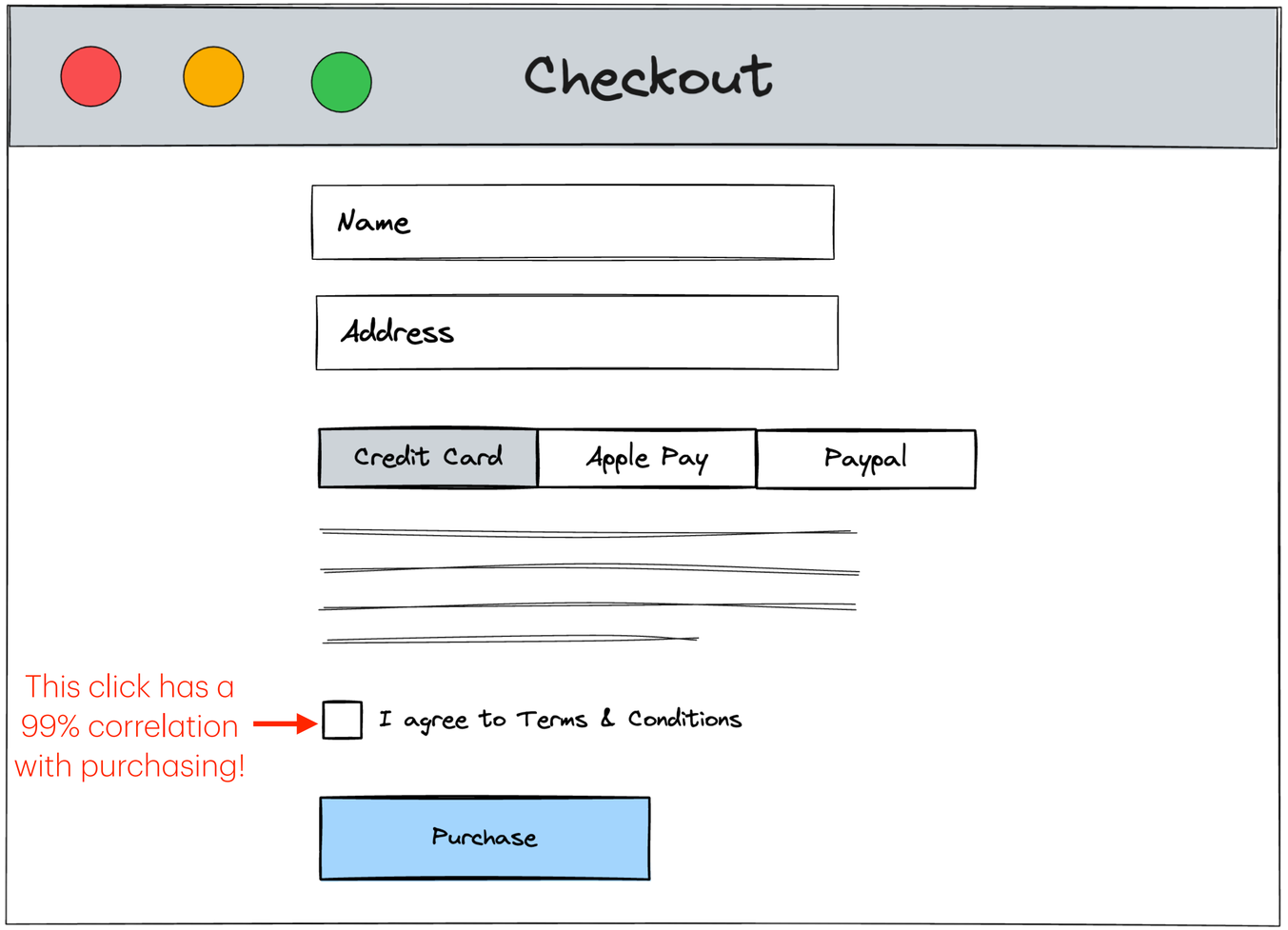
user_id	viewed_home	clicked_search	clicked_read_reviews	clicked_learn_more	...yada yada...	purchased
187180	0	1	1	0	...	1
662863	0	0	0	0	...	1
684172	1	1	1	0	...	0
261982	1	1	0	0	...	0
579325	1	1	1	1	...	0
829755	1	1	0	1	...	0
454574	1	1	0	0	...	0
216105	1	1	1	1	...	1
181215	0	1	0	0	...	0
996510	1	1	0	0	...	0
379967	1	1	0	0	...	0
511056	1	1	0	1	...	0
948165	1	1	1	0	...	0
575719	1	1	0	1	...	0
356383	1	1	1	0	...	0
582609	1	1	1	0	...	0
988781	1	0	0	1	...	0
857058	1	1	0	0	...	0
498438	1	1	0	0	...	0
470140	1	0	1	0	...	0
714748	1	0	0	1	...	0
624947	1	1	1	1	...	0
524455	1	0	0	1	...	1

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$



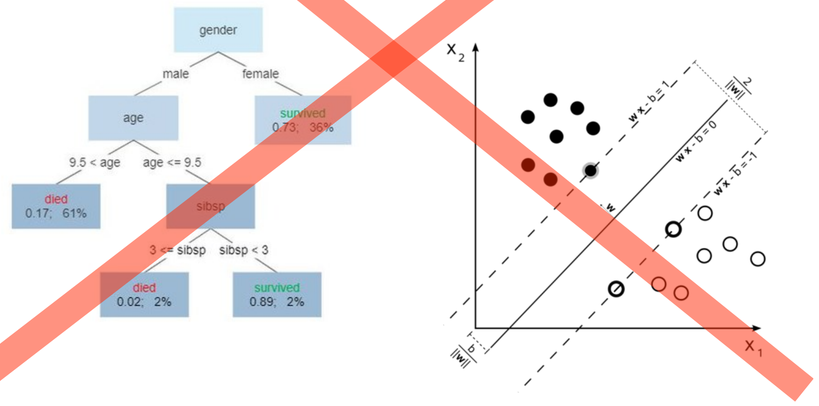
user_id	viewed_home	clicked_search	clicked_read_reviews	clicked_learn_more	...yada yada...	purchased
187180	0	1	1	0	...	1
662863	0	0	0	0	...	1
684172	1	1	1	0	...	0
261982	1	1	0	0	...	0
579325	1	1	1	1	...	0
829755	1	1	0	1	...	0
454574	1	1	0	0	...	0
216105	1	1	1	1	...	1
181215	0	1	0	0	...	0
996510	1	1	0	0	...	0
379967	1	1	0	0	...	0
511056	1	1	0	1	...	0
948165	1	1	1	0	...	0
575719	1	1	0	1	...	0
356383	1	1	1	0	...	0
582609	1	1	1	0	...	0
988781	1	0	0	1	...	0
857058	1	1	0	0	...	0
498438	1	1	0	0	...	0
470140	1	0	1	0	...	0
714748	1	0	0	1	...	0
624947	1	1	1	1	...	0
524455	1	0	0	1	...	1

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$



The most important problems facing product builders aren't predictive, they're **causal**

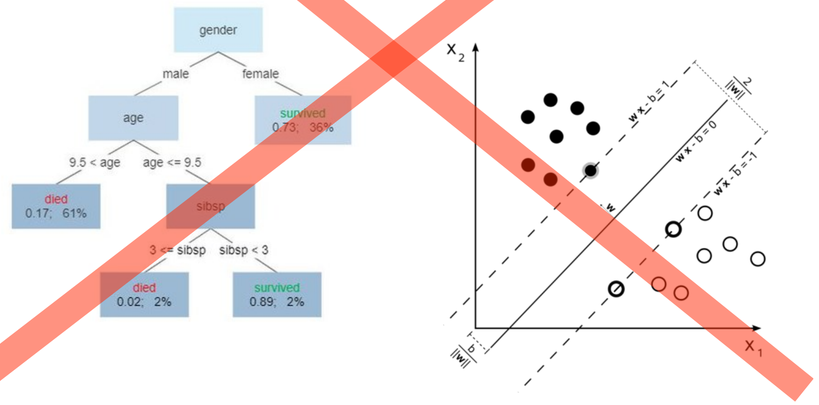
$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$



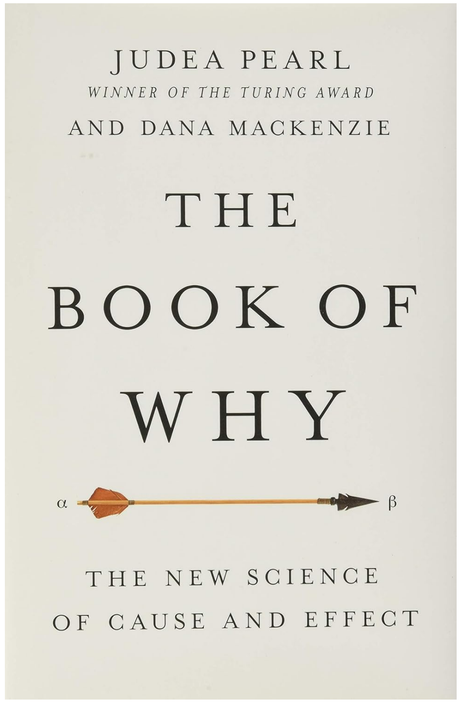
Prediction

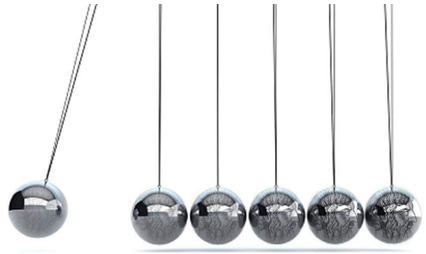
The most important problems facing product builders aren't predictive, they're **causal**

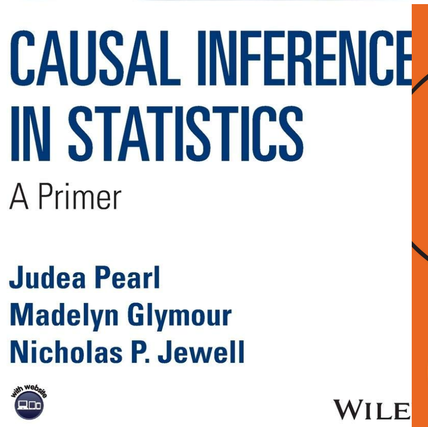
$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$



Prediction





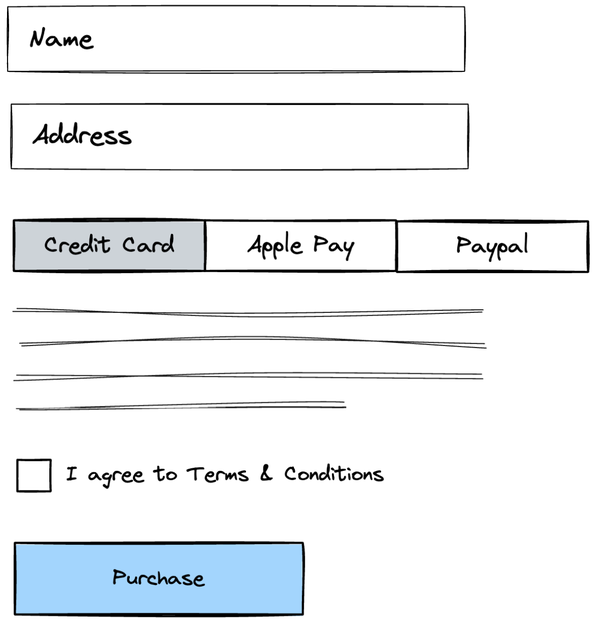


Intervention

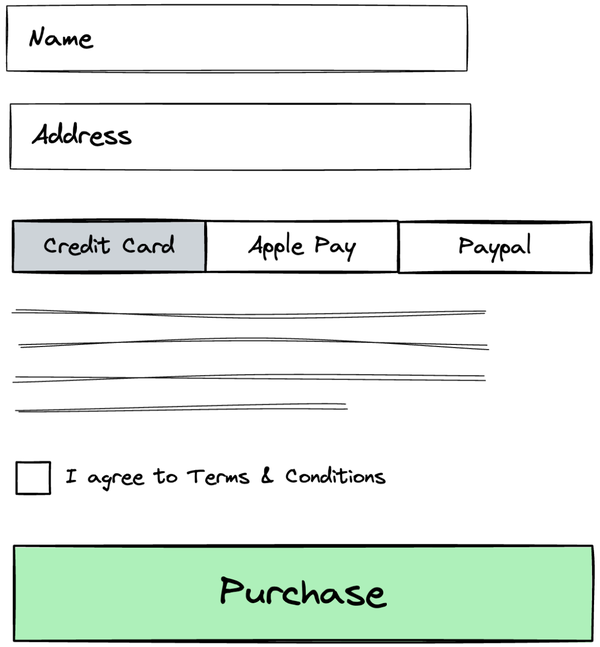
Maybe it *is* actually true that making the purchase button bigger would improve the conversion rate. How can we determine that?

Maybe it *is* actually true that making the purchase button bigger would improve the conversion rate. How can we determine that?

A/B Test!

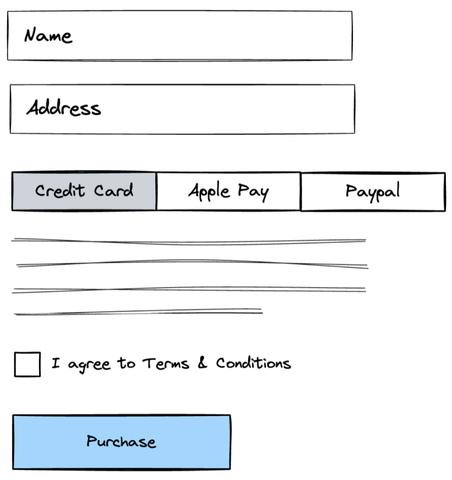


Control

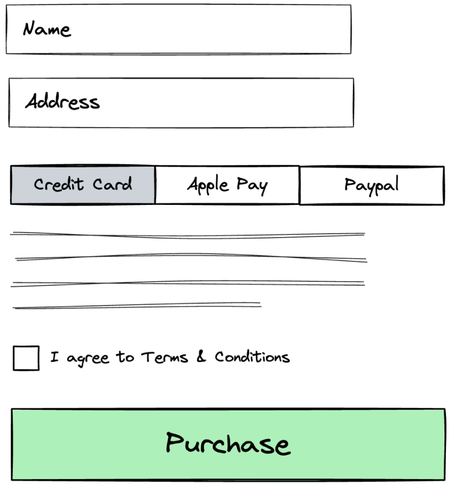


Experiment

A/B tests are the gold standard for determining the impact of an intervention

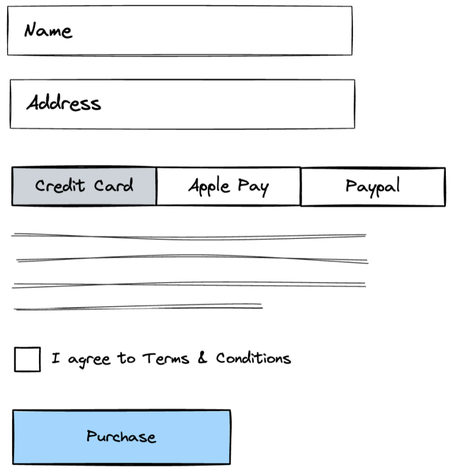


Control

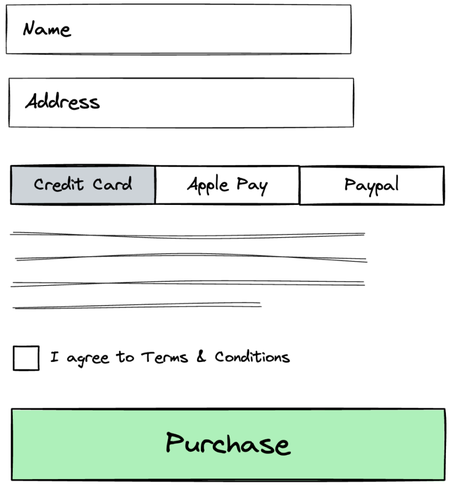


Experiment

A/B tests are the gold standard for determining the impact of an intervention



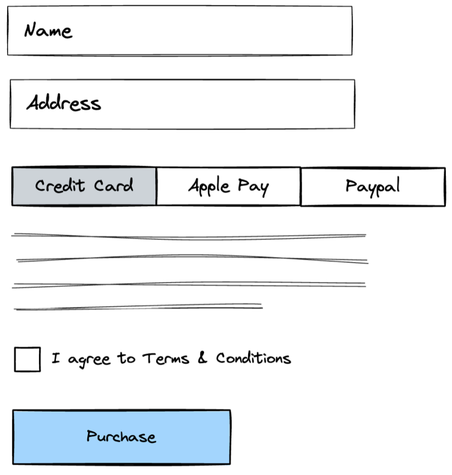
Control



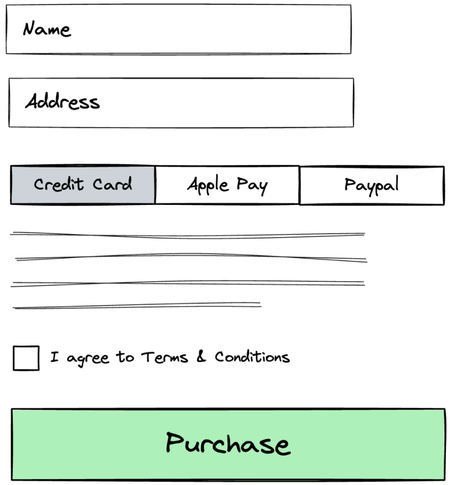
Experiment

- It's no surprise it works: they're randomized controlled trials (RCTs), which scientists have studied for a century!

A/B tests are the gold standard for determining the impact of an intervention



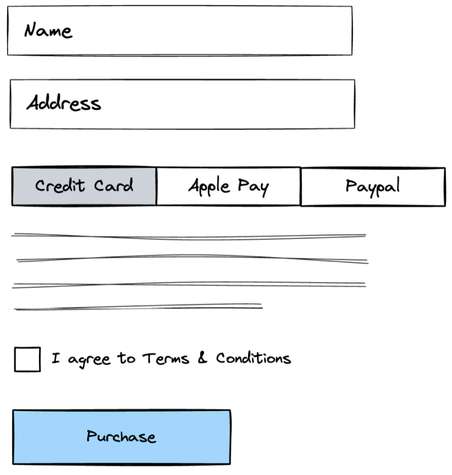
Control



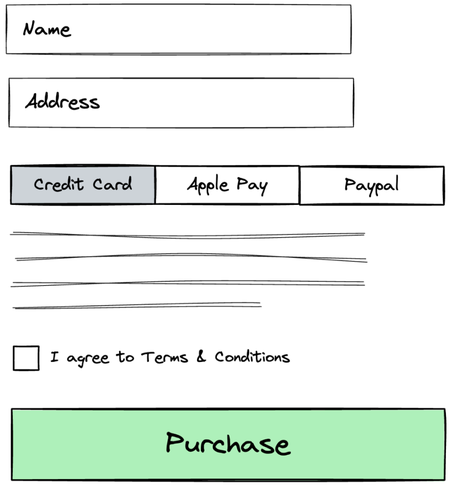
Experiment

- It's no surprise it works: they're randomized controlled trials (RCTs), which scientists have studied for a century!
- But...

A/B tests are the gold standard for determining the impact of an intervention



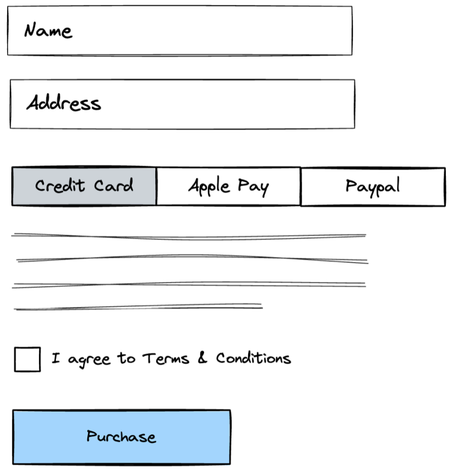
Control



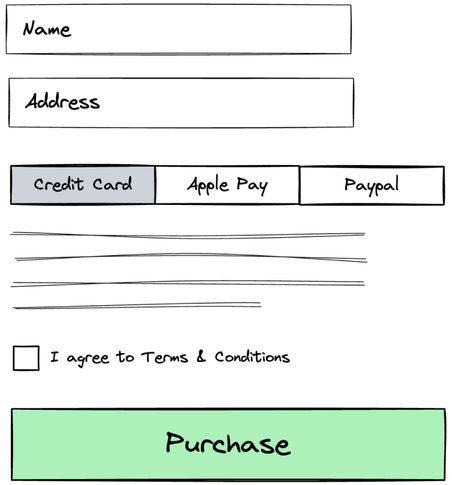
Experiment

- It's no surprise it works: they're randomized controlled trials (RCTs), which scientists have studied for a century!
- But...
  - they are **not** the only way to reason about causal effects

A/B tests are the gold standard for determining the impact of an intervention



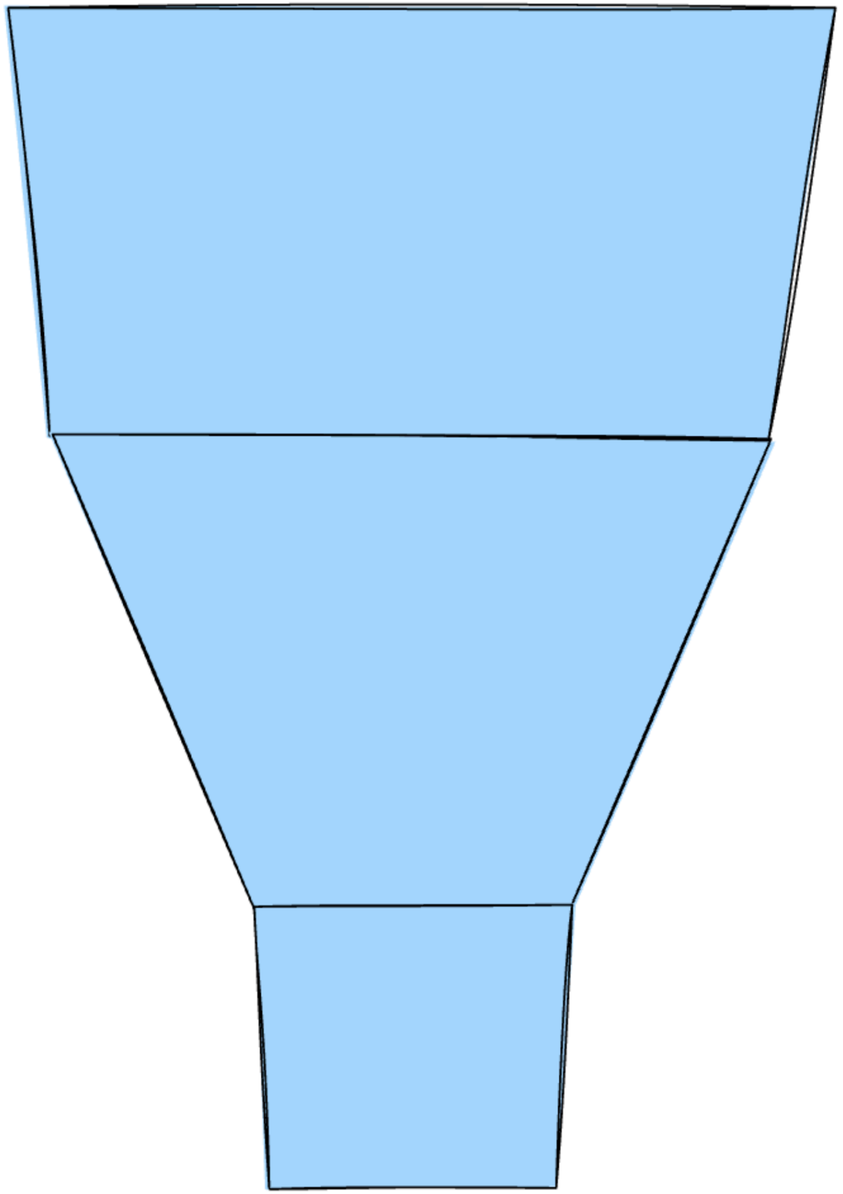
Control

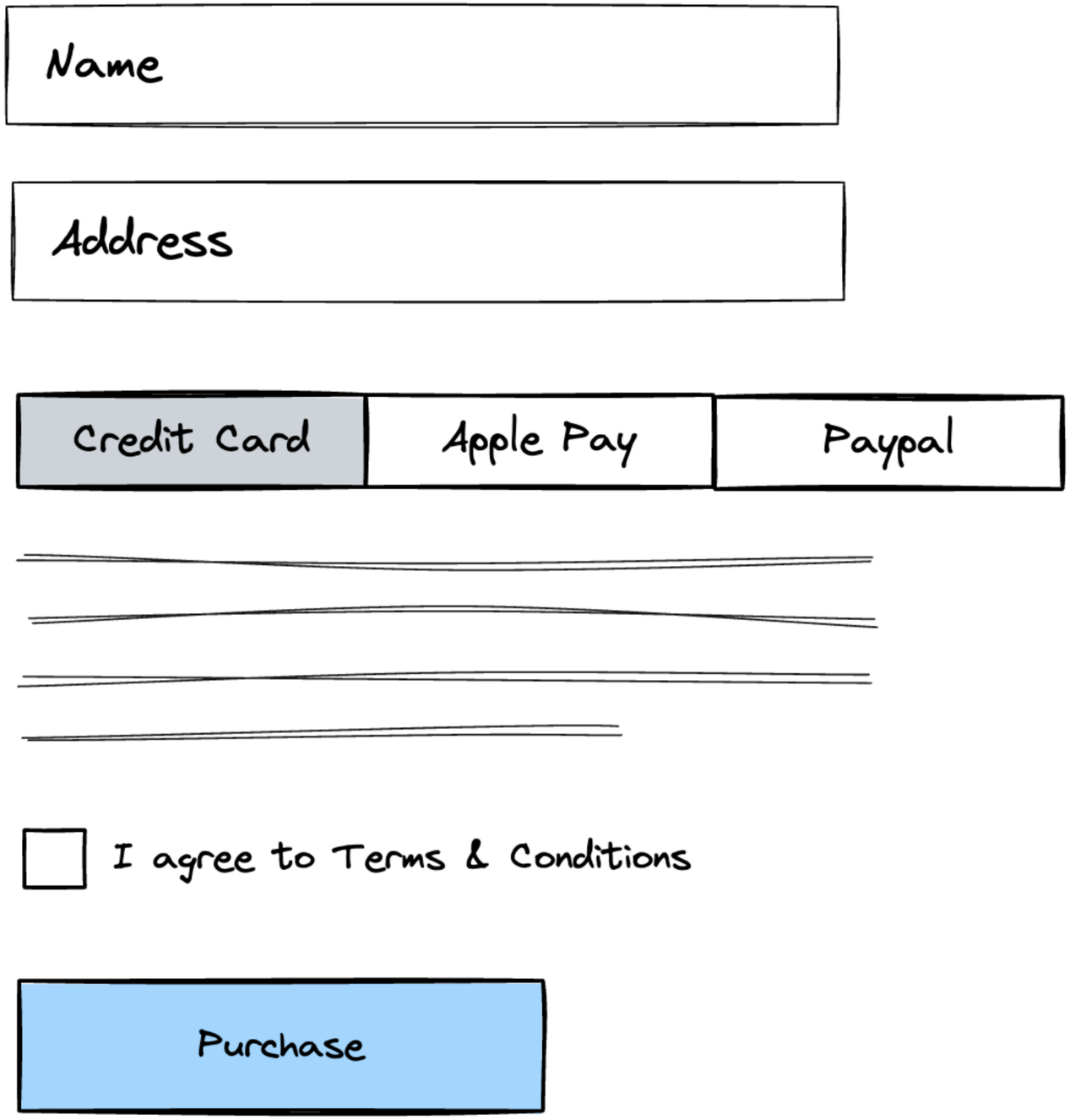


Experiment

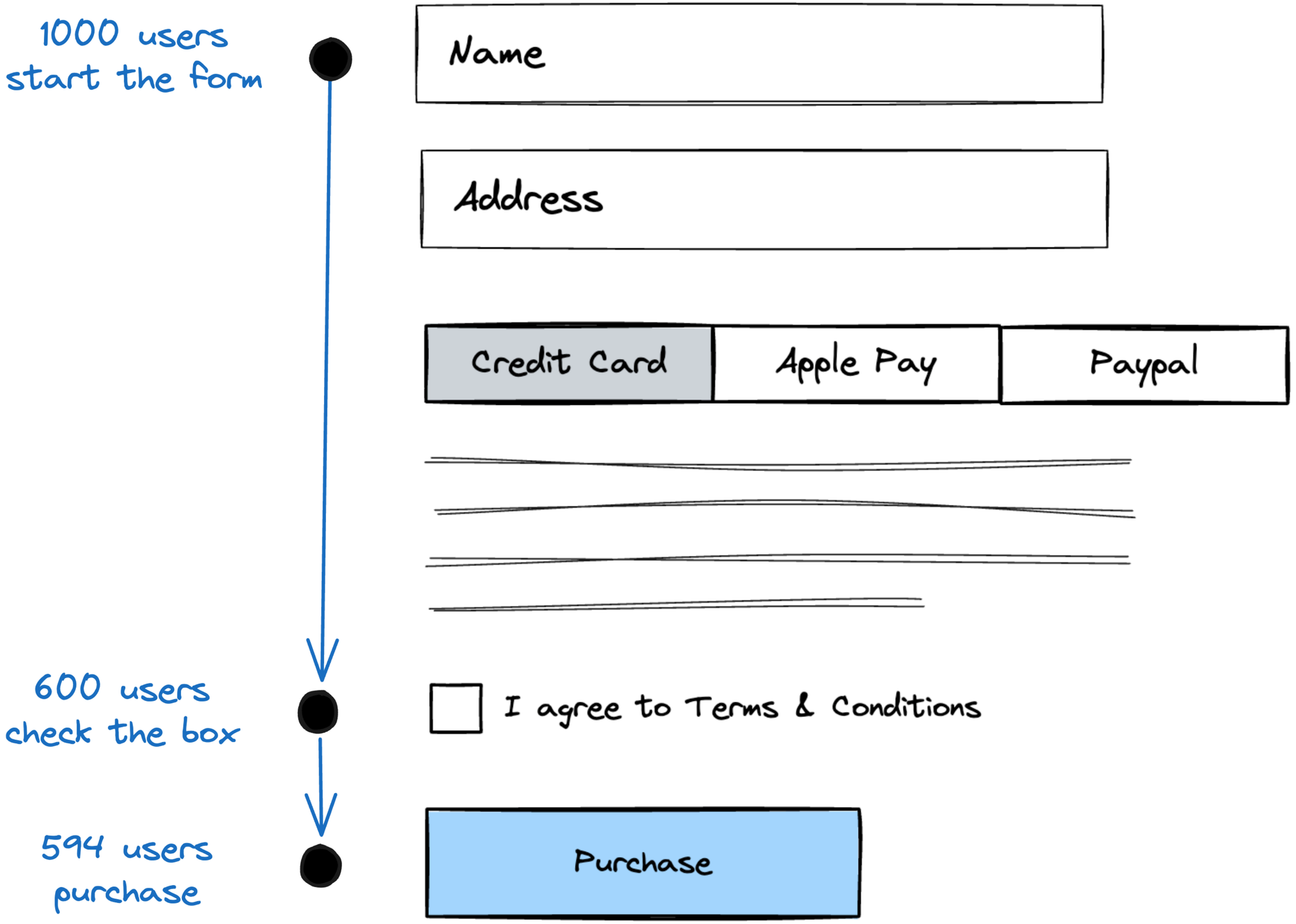
- It's no surprise it works: they're randomized controlled trials (RCTs), which scientists have studied for a century!
- But...
  - they are **not** the only way to reason about causal effects
  - it does **not** mean that all observational data is useless

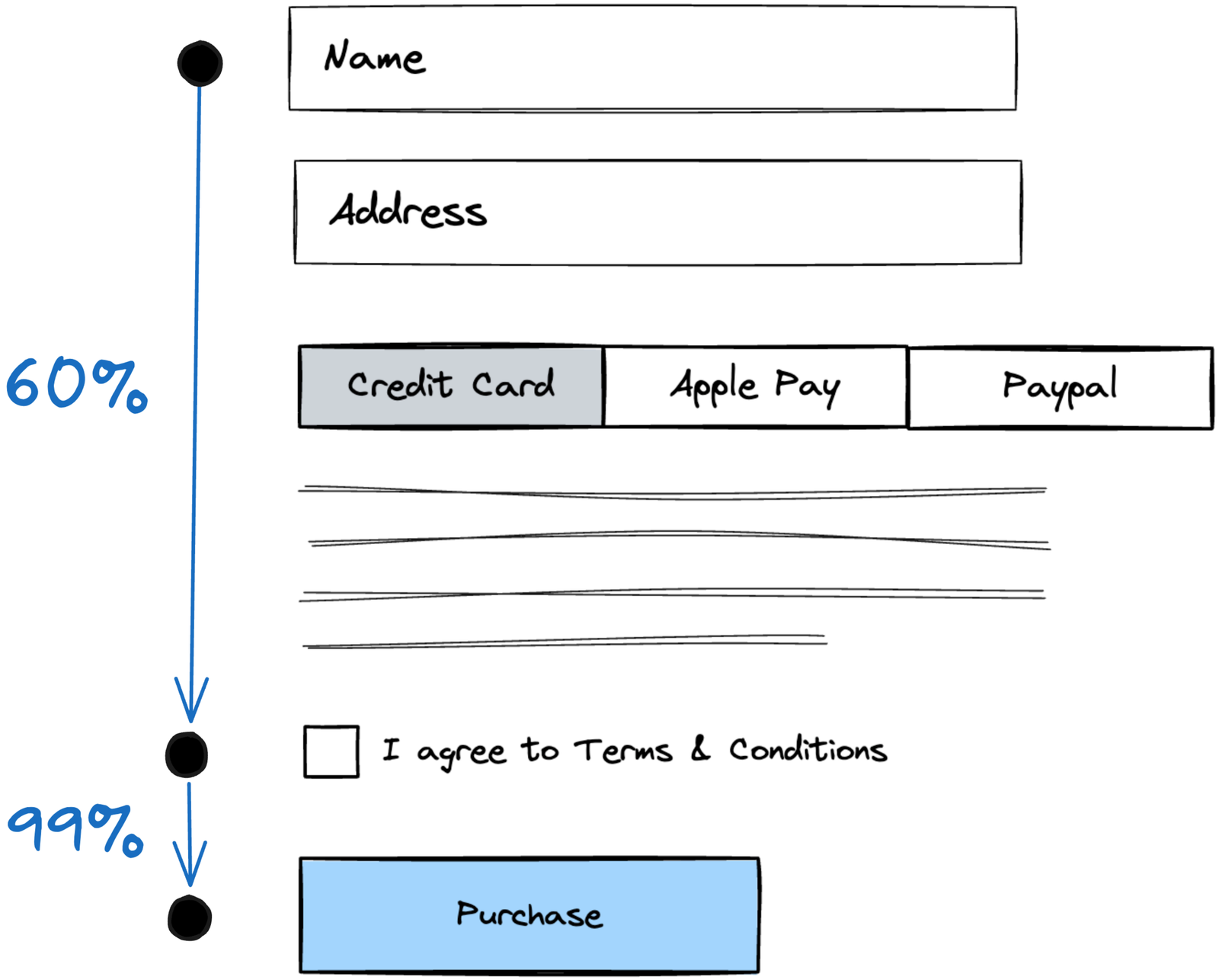
Funnels

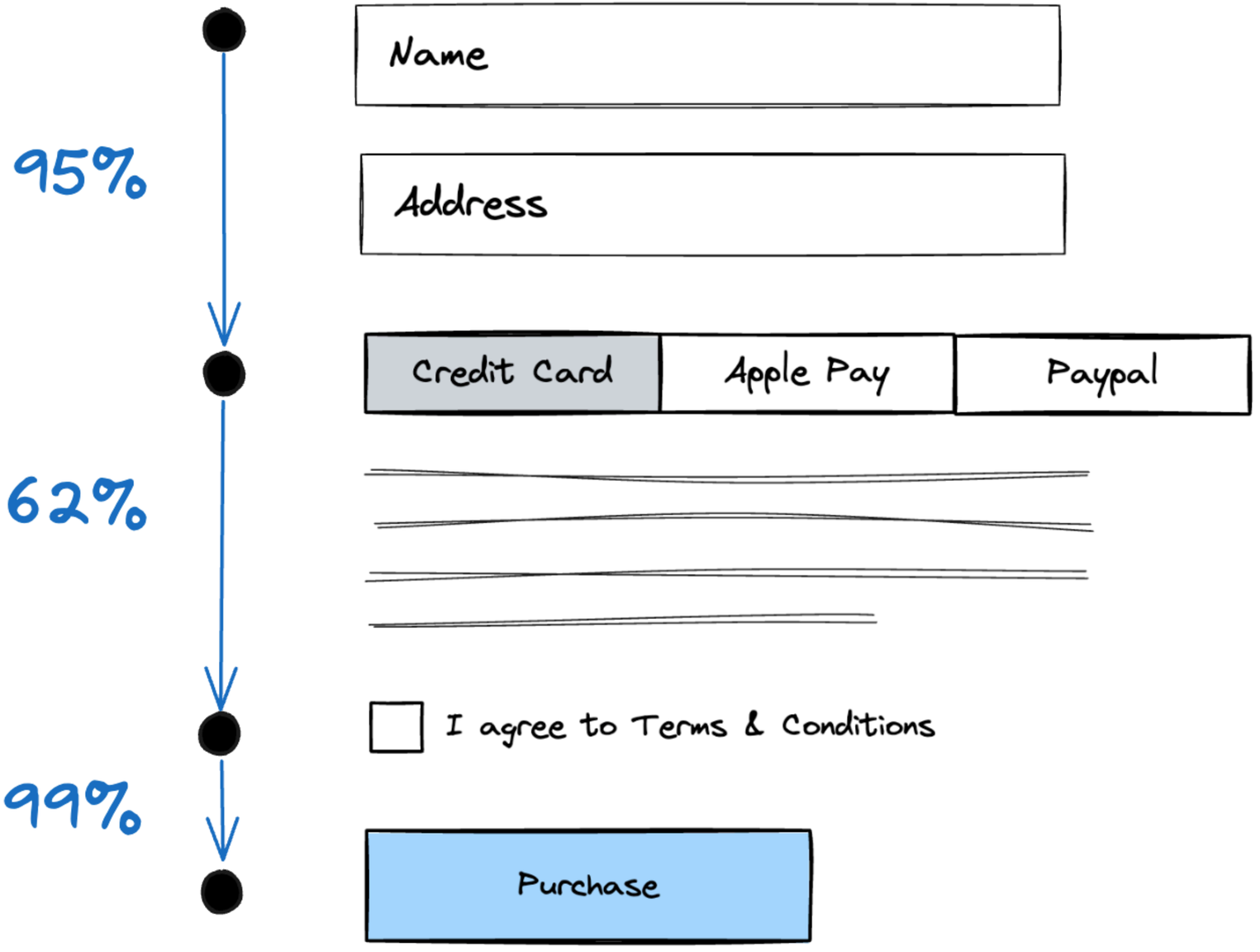


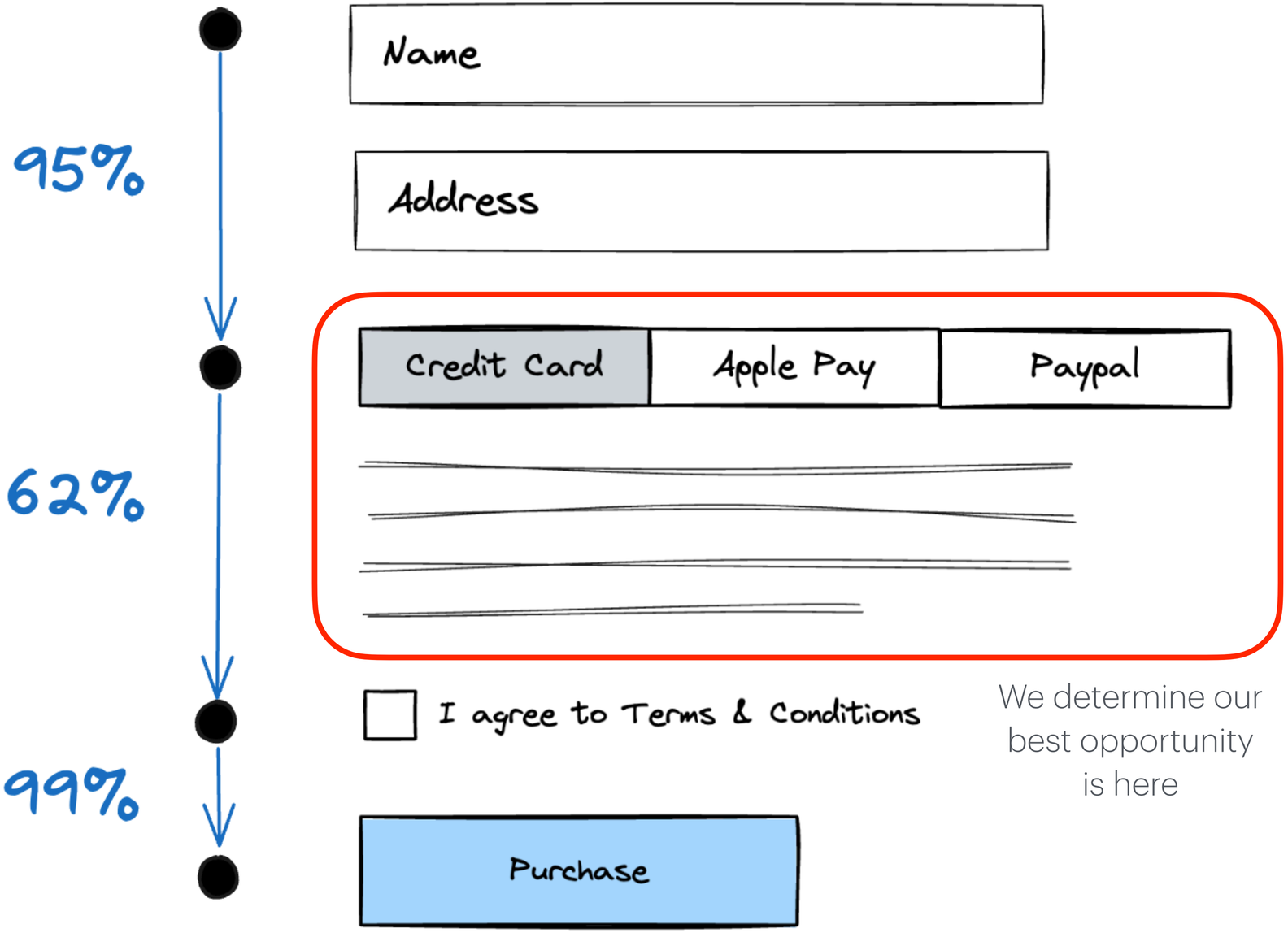

Name
Address
Credit Card
Apple Pay
Paypal
I agree to Terms & Conditions
Purchase

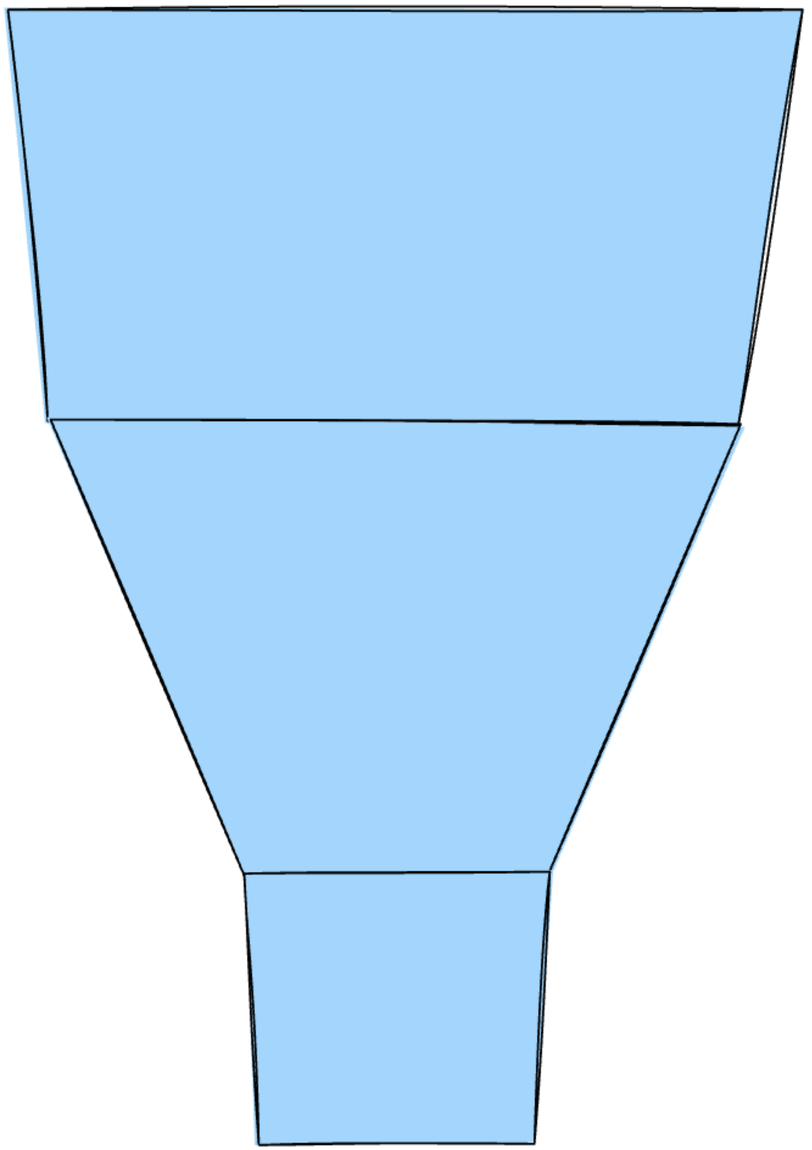
Name
Address
Credit Card
Apple Pay
Paypal
I agree to Terms & Conditions
Purchase


1000 users
start the form
Name
Address
Credit Card
Apple Pay
Paypal
600 users
check the box
I agree to Terms & Conditions
594 users
purchase
Purchase


Name
Address
60%
Credit Card
Apple Pay
Paypal
I agree to Terms & Conditions
99%
Purchase


95%
62%
99%
Name
Address
Credit Card
Apple Pay
Paypal
I agree to Terms & Conditions
Purchase


95%
62%
99%
Name
Address
Credit Card
Apple Pay
Paypal
I agree to Terms & Conditions
Purchase
We determine our best opportunity is here



95%

62%

99%

Name

Address

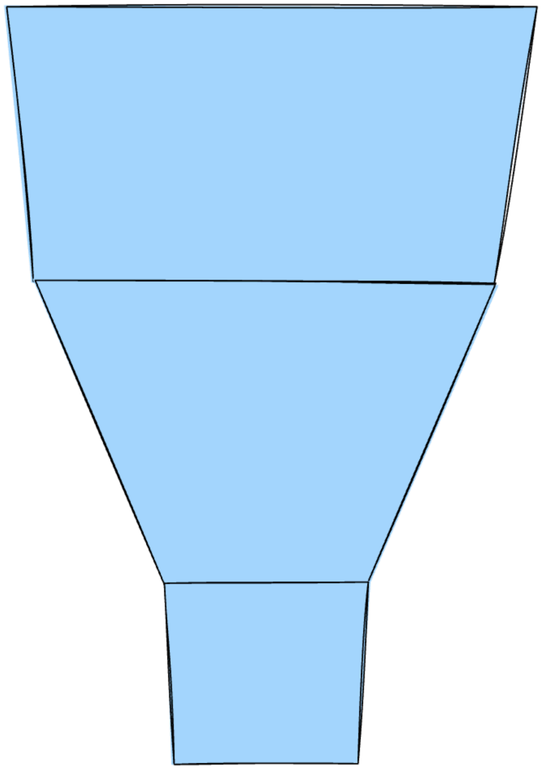
Credit Card | Apple Pay | Paypal

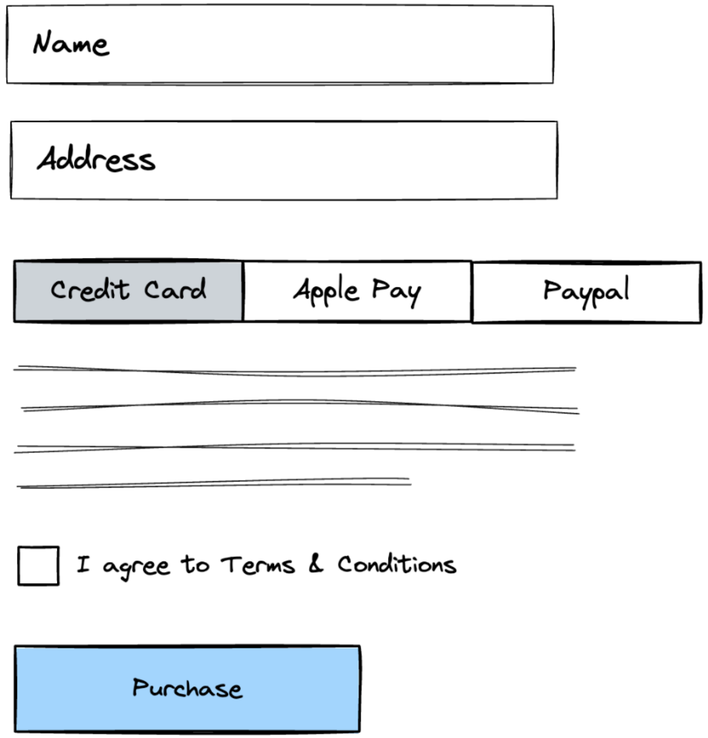
I agree to Terms & Conditions

Purchase

We determine our best opportunity is here

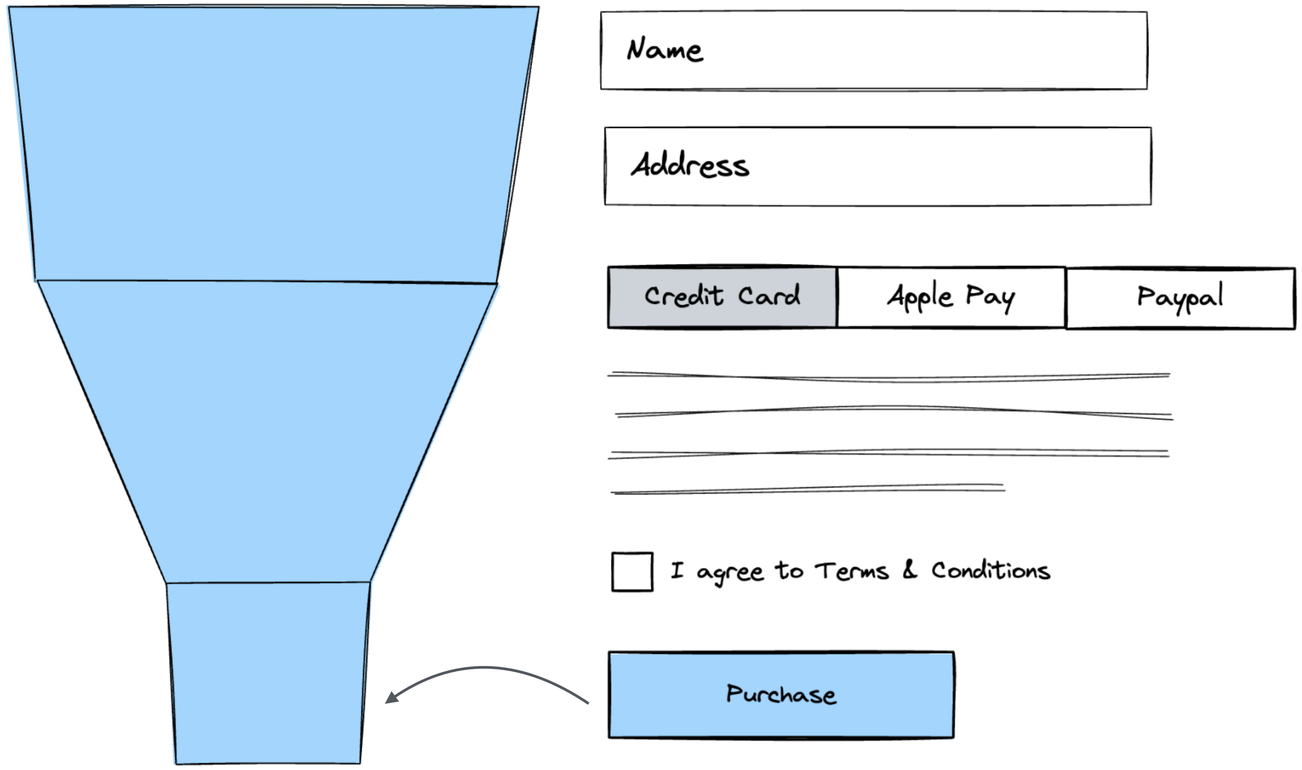
Why are we able to do causal reasoning with funnels?





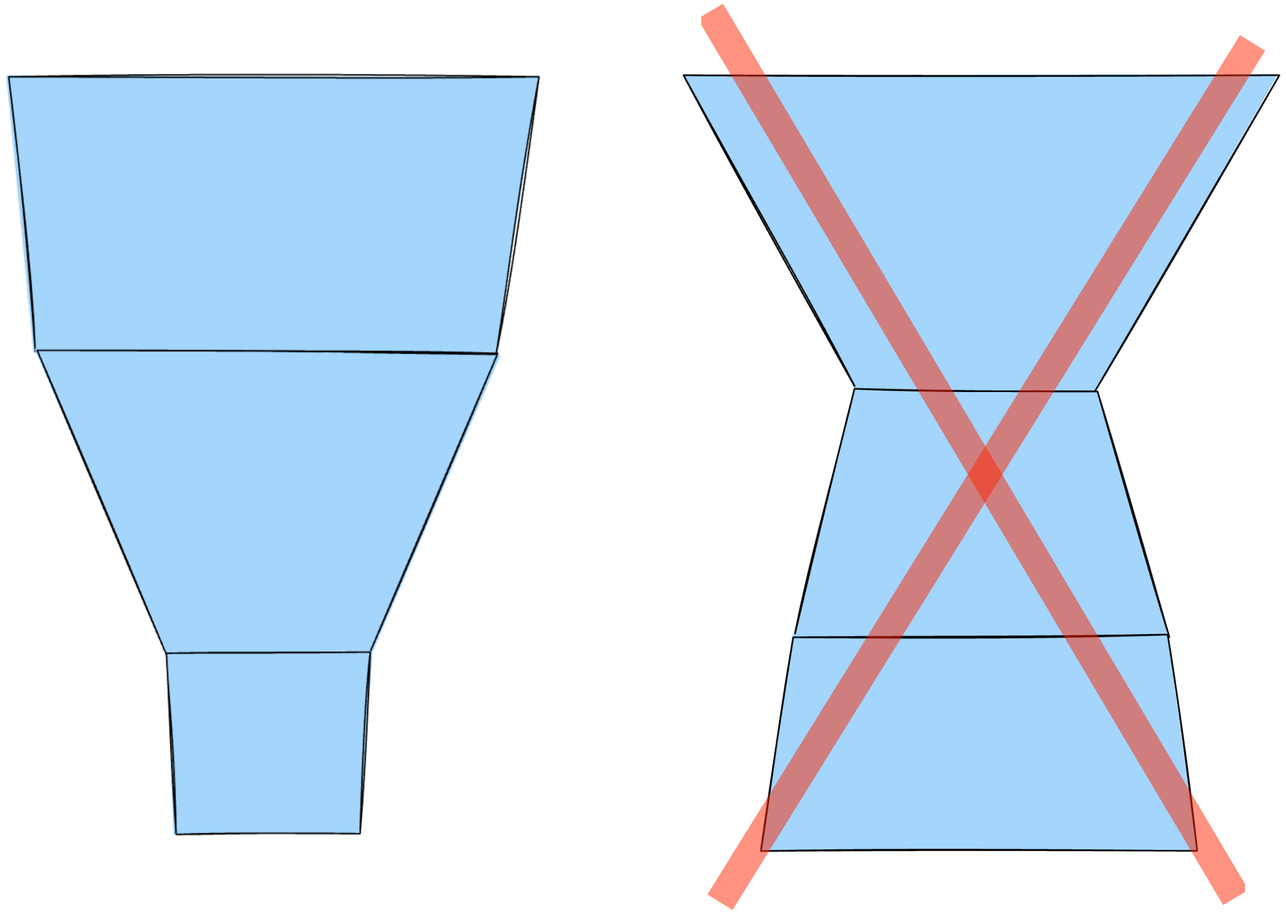
Why are we able to do causal reasoning with funnels?

- **Locality:** We assume modifications to one part of a flow affect only that step in the funnel



Why are we able to do causal reasoning with funnels?

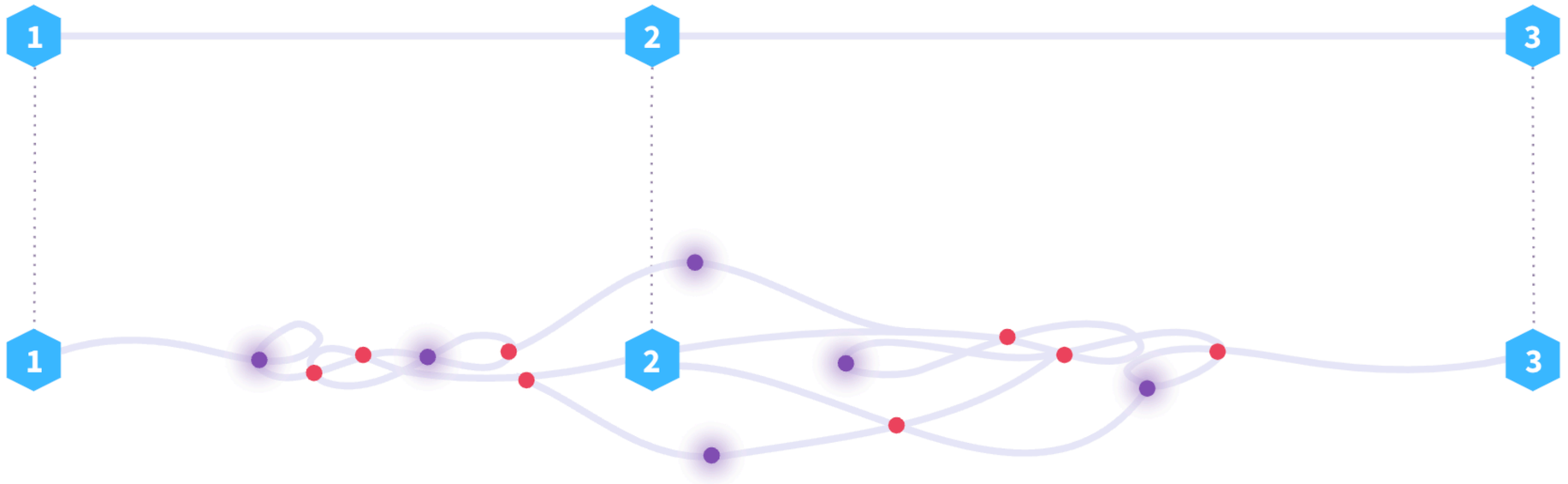
- **Locality:** We expect modifications to one part of a flow to affect only that step in the funnel
- **Chronology:** Users are measured as they progress forward in the funnel; not if they skip steps



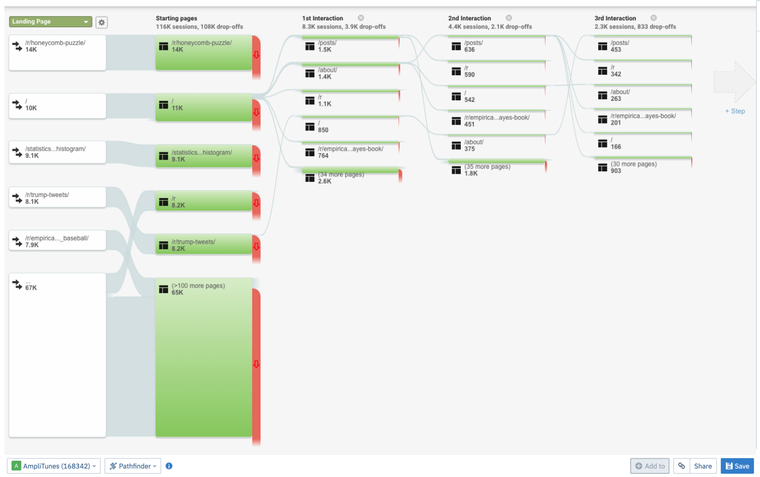
The weakness of a funnel is that it's **linear**

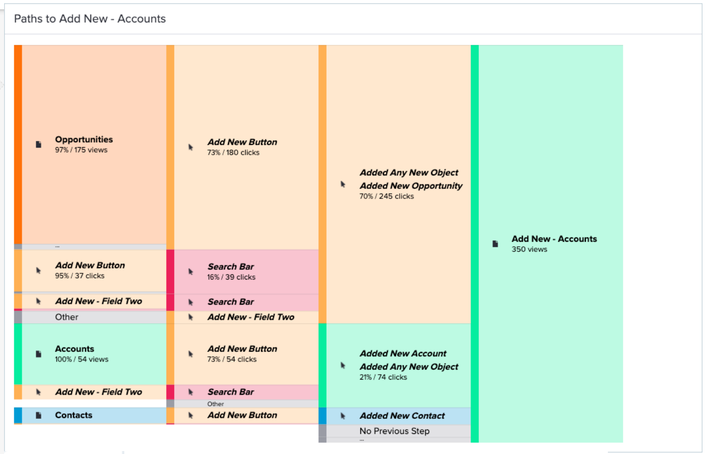
What a funnel sees

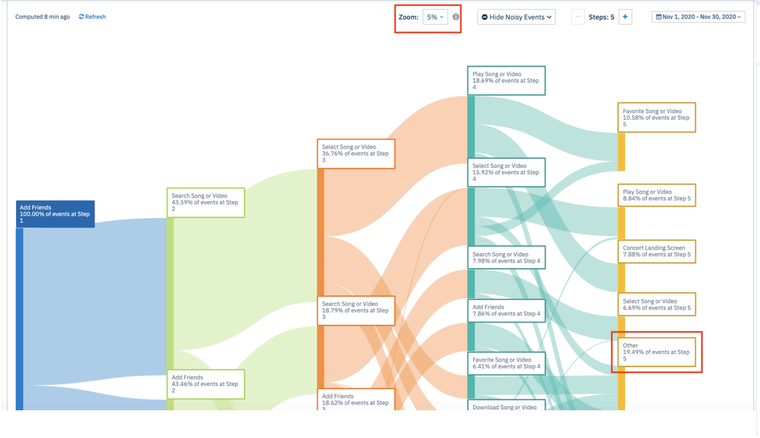
What the user did

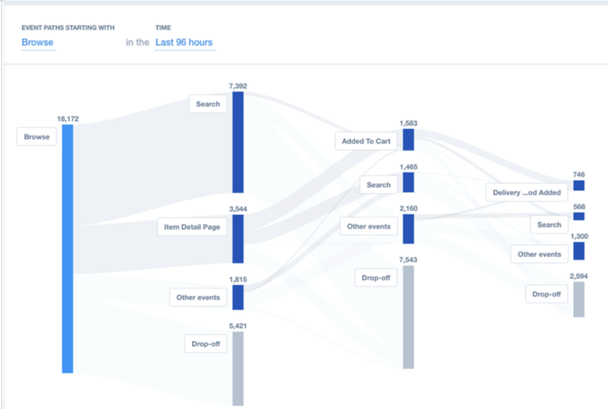


Previous tools for measuring nonlinear sequences count *consecutive* actions









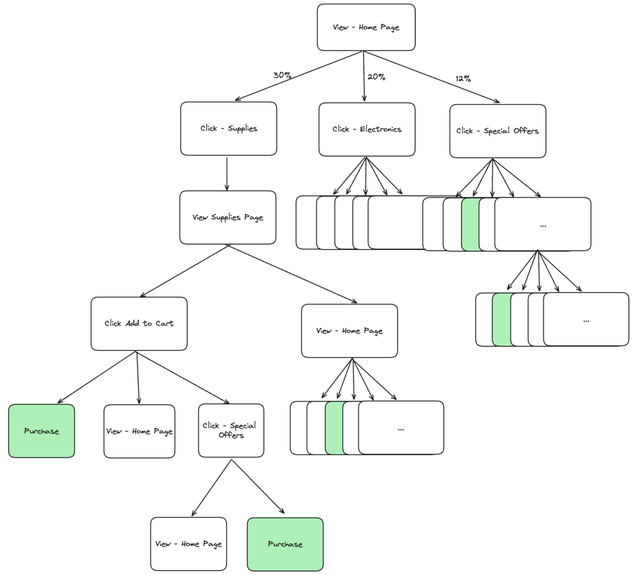

```
Land on /home
Click "Special Offers"
View /special-offers
Click "Paperclips"
View /paper-clips
Click "Add to Cart"
Click "Check Out"
View /checkout
Click "Save for Later"
Drop Off
```

```
Land on /home
Click "Special Offers"
View /special-offers
Click Back
Click "Search"
Enter "Headphones"
Click - Sort by Price
Click - Sort by Category
Drop Off
```

```
Land on /promo
Click "50% Off Special Offers"
View /special-offers
Click Back
Click - Sign In
Click "Search"
Enter "Keyboard"
Click - Sort by Price
Click - Filter by Price
Click - Mechanical Keyboard
View /products/mechanical-keyboard
Click - "Buy (One Click)"
Purchase
```

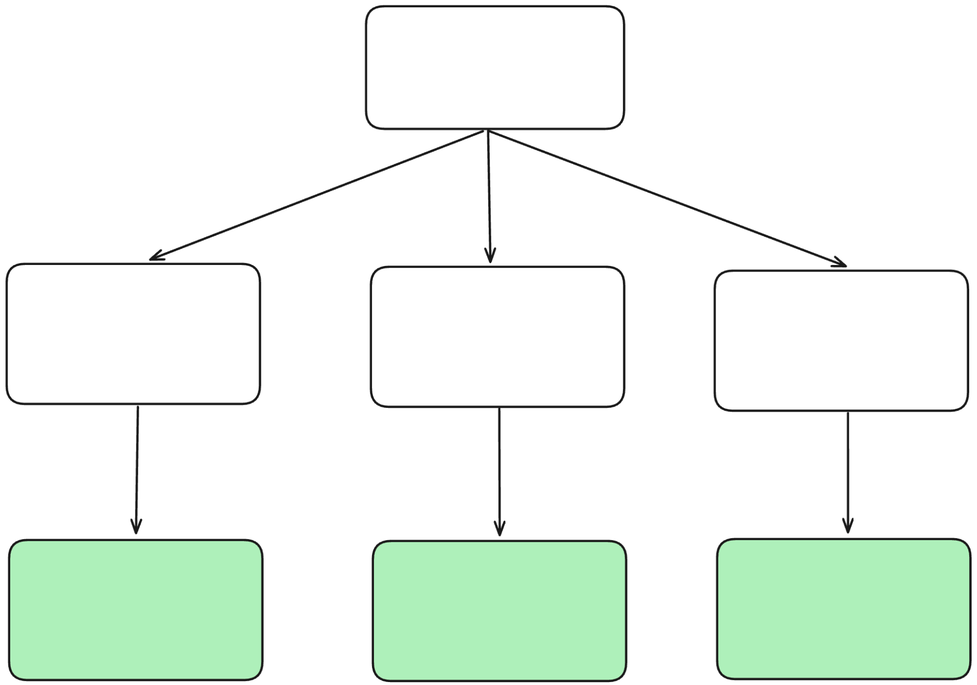
```
Land on /home
Click "Browse"
View /categories
View /recent
Click Back
Drop Off
```

```
View /home
Click "Supplies"
View /supplies
Click "Paper"
View /products/paper
Click "Add to Cart"
View /supplies
Click "Scissors"
View /products/scissors
Click Back
Click "Premium Scissors"
View /products/scissors
Click "Add to Cart"
View - Checkout
Click - Enter Name
Click - Enter Credit Card
Validation Error
Click - Enter Credit Card
Click - Check Terms & Conditions
Purchase
```



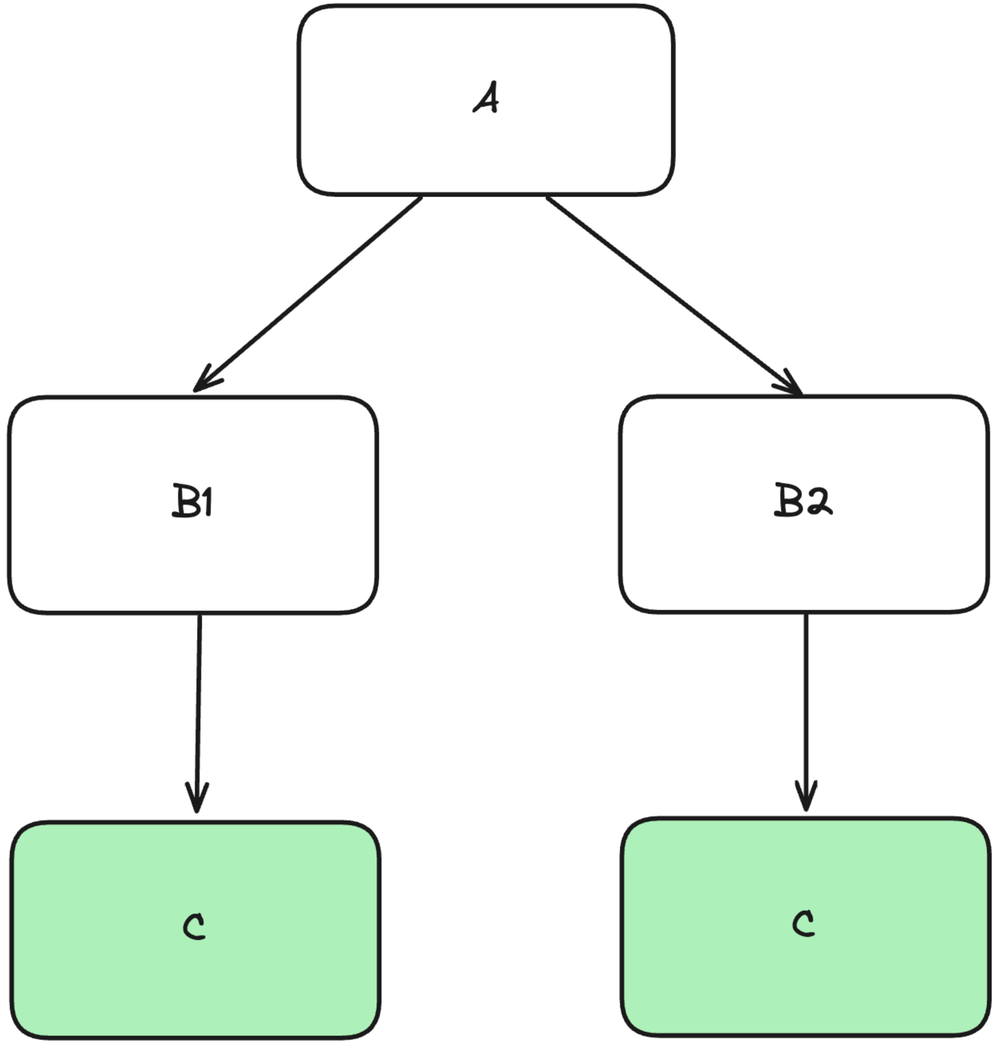
This leads to an explosion in state space: very difficult to draw conclusions

Journeygrams



Journeygrams are a calculus that uses a directed rooted tree to quantify sequences of events

We **specify** the tree based on the causal hypotheses we're considering



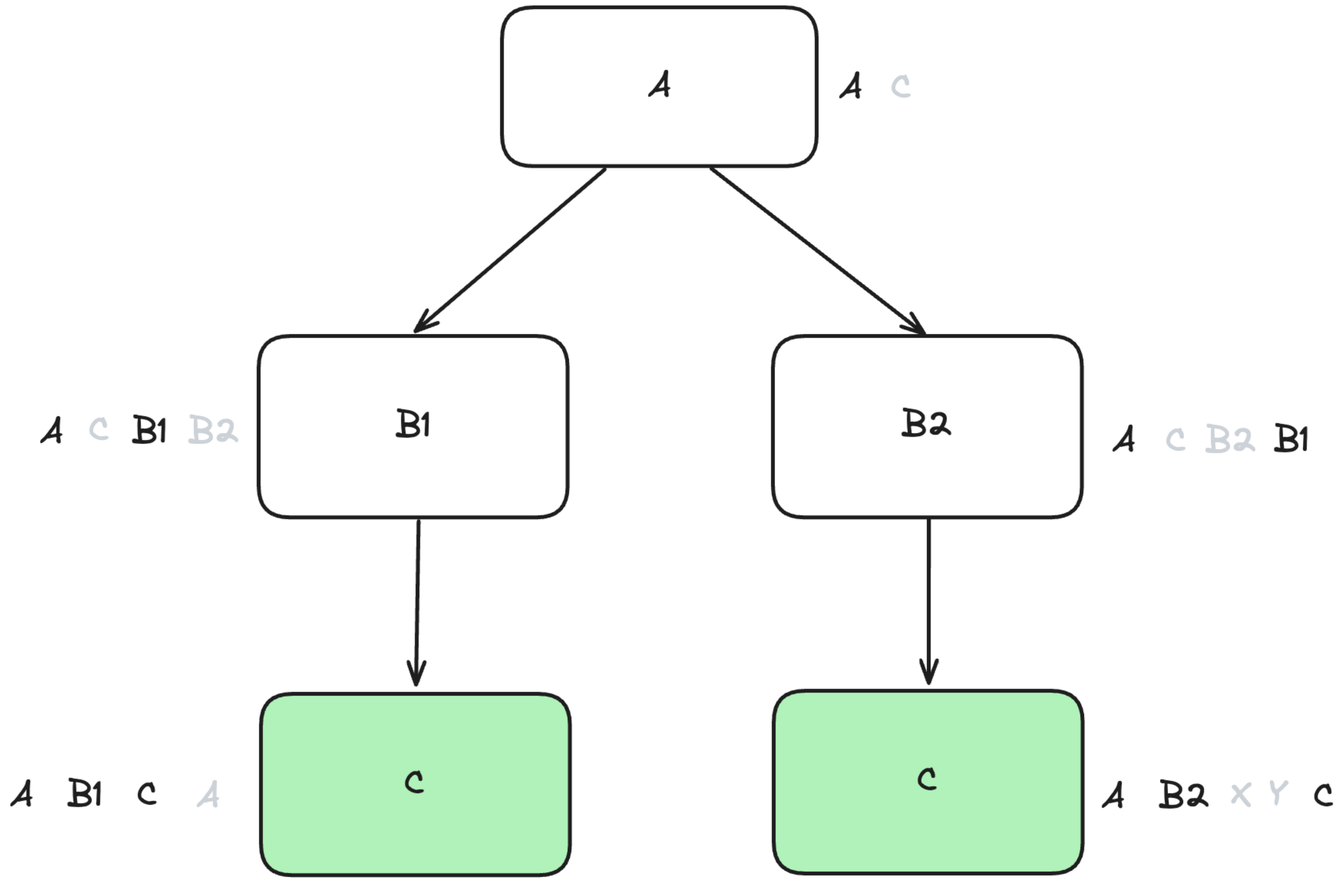
How Journeygrams Work

Sequences **walk** the tree, transitioning to the next step when they see an event

When there's a fork, it follows the one that happens **first**

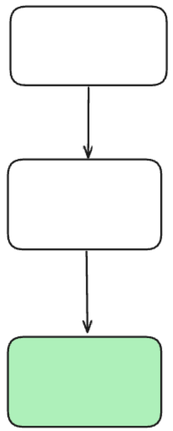
They **can't skip** steps

It **ignores** events other than the ones it is waiting for

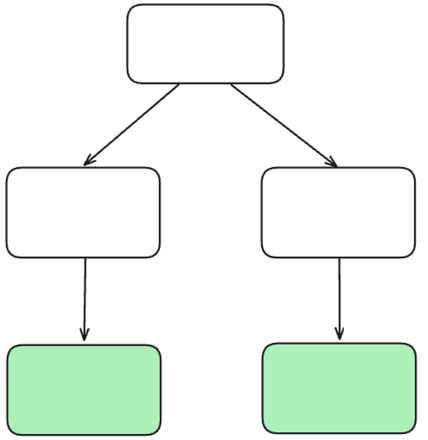


Funnels are a **special case** of journeygram

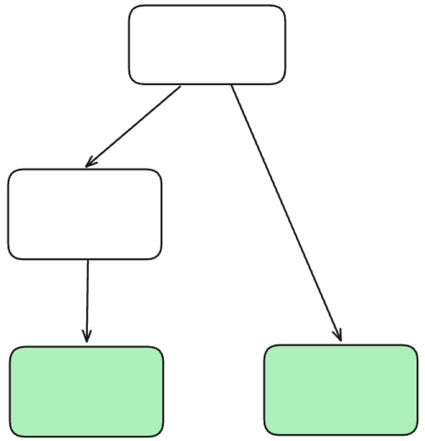
Other shapes let us extend the kinds of causal conclusions we intuitively draw from funnels



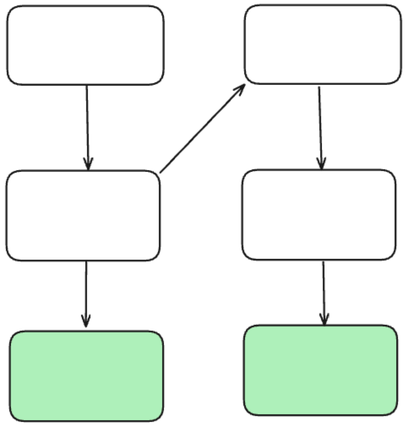
Funnel



Fork

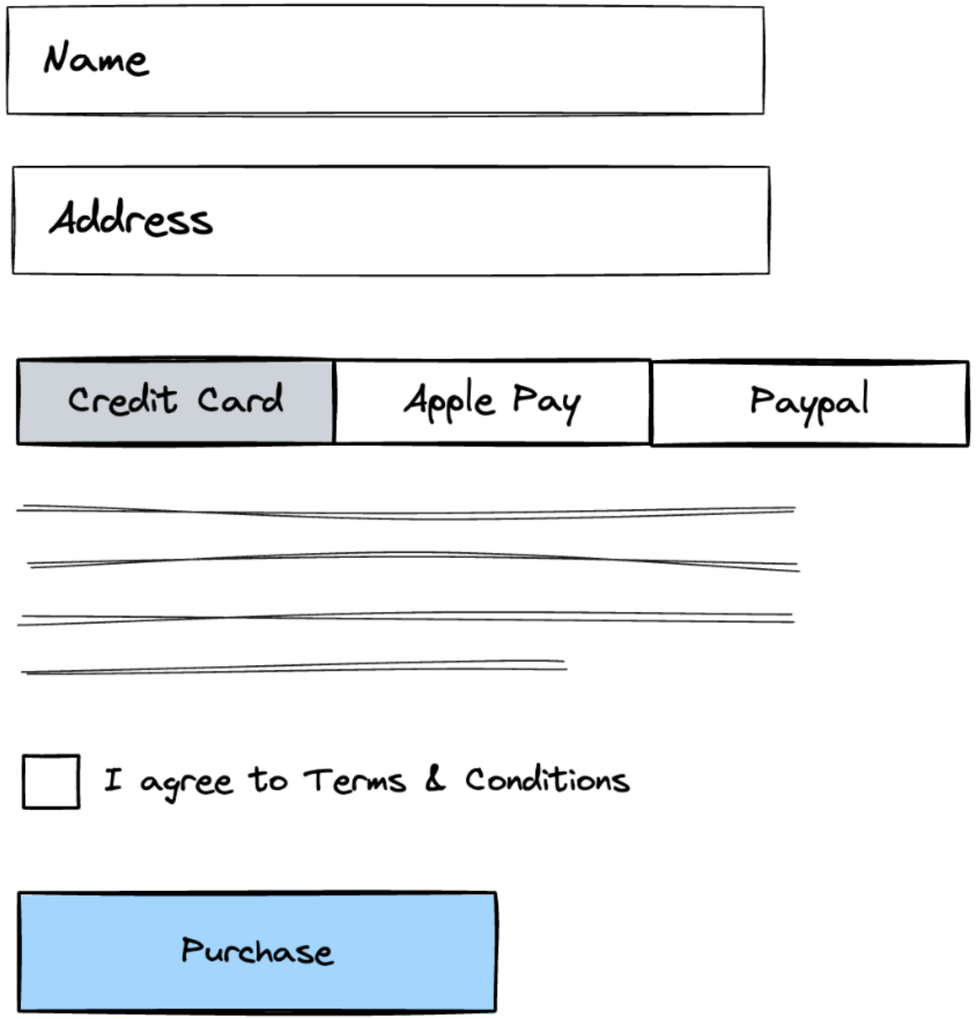


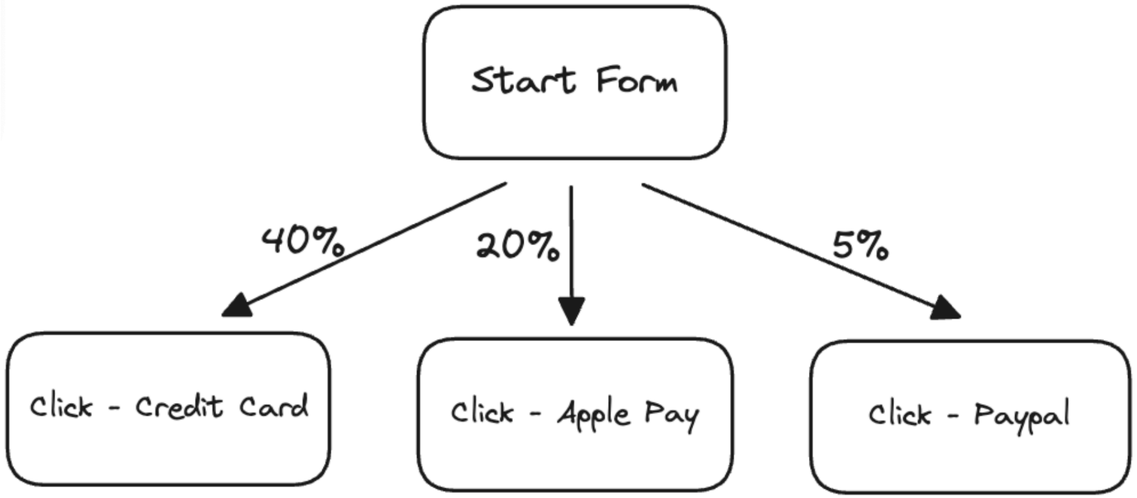
Skippable Steps

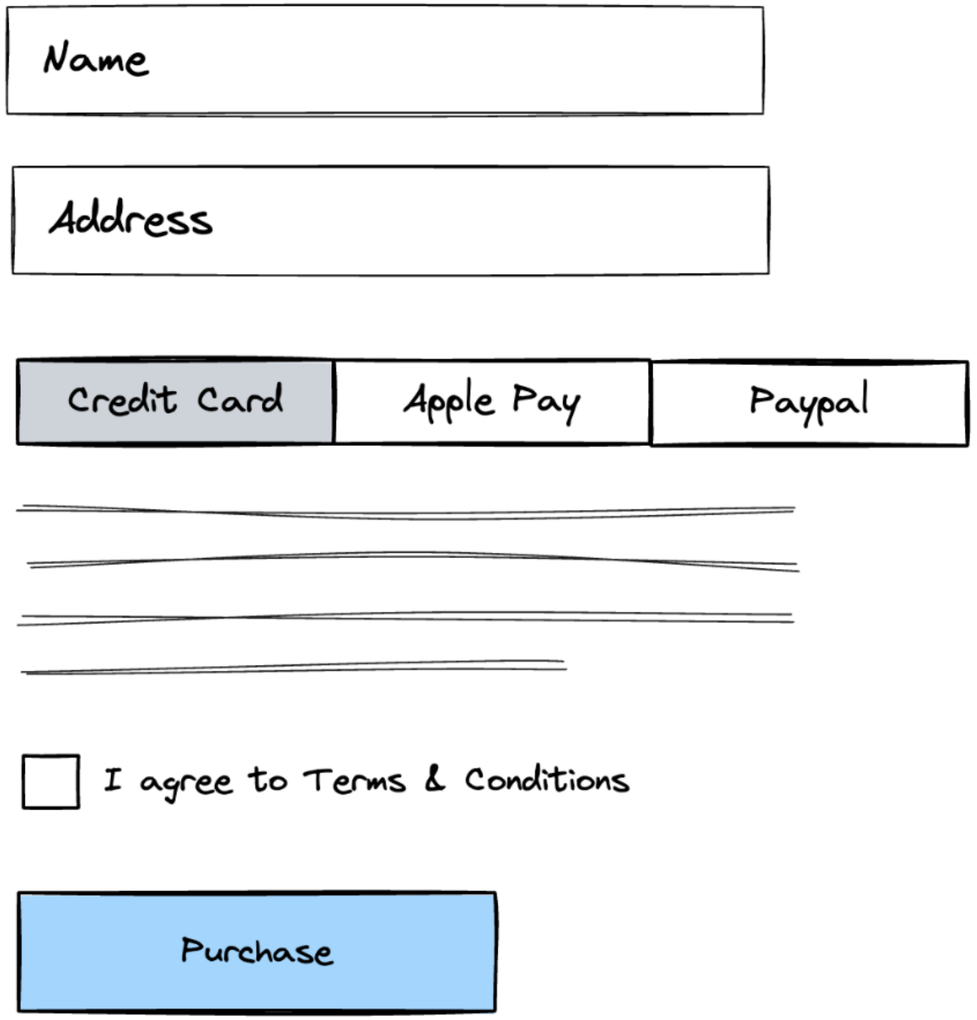


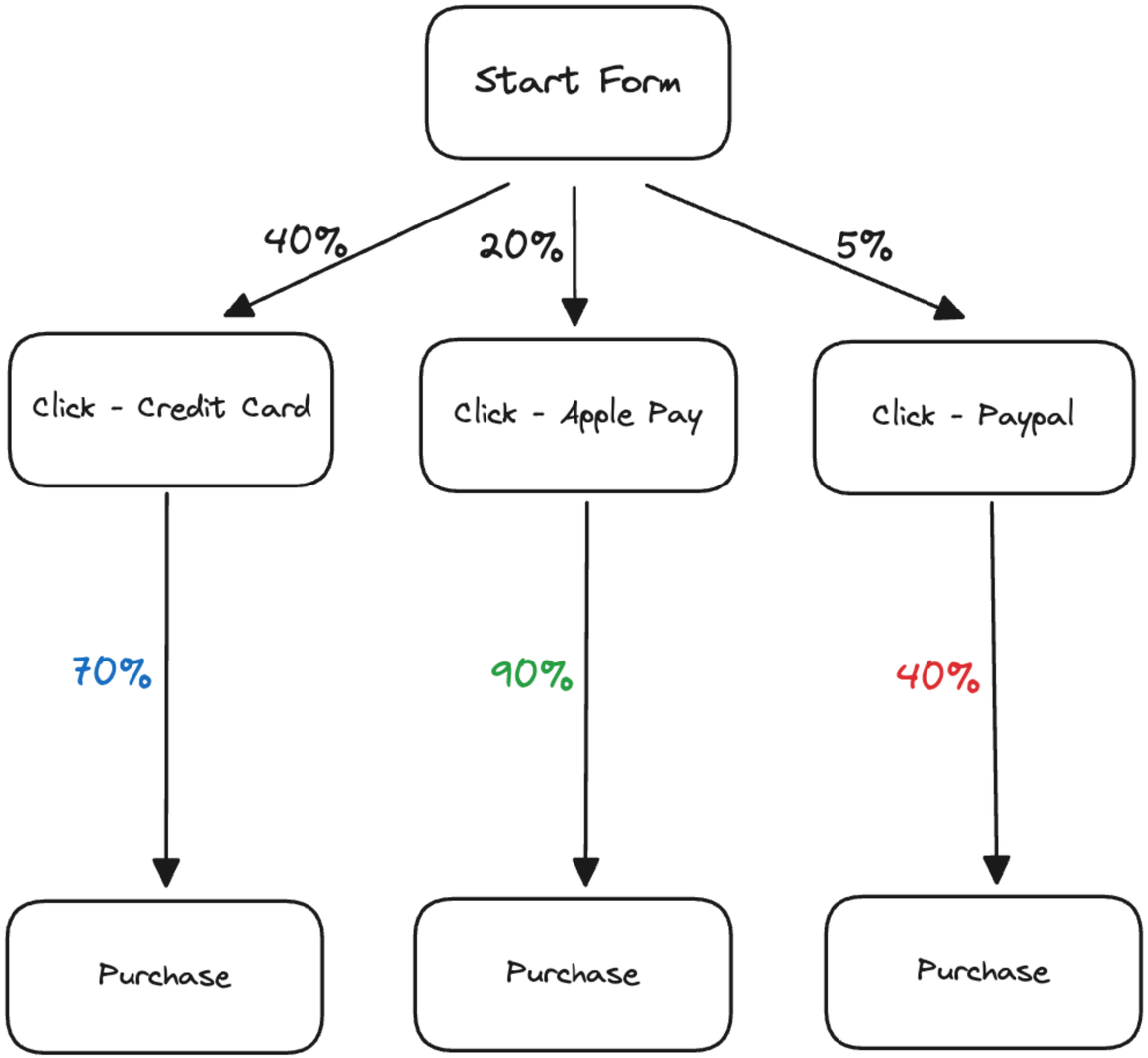
Backtracking

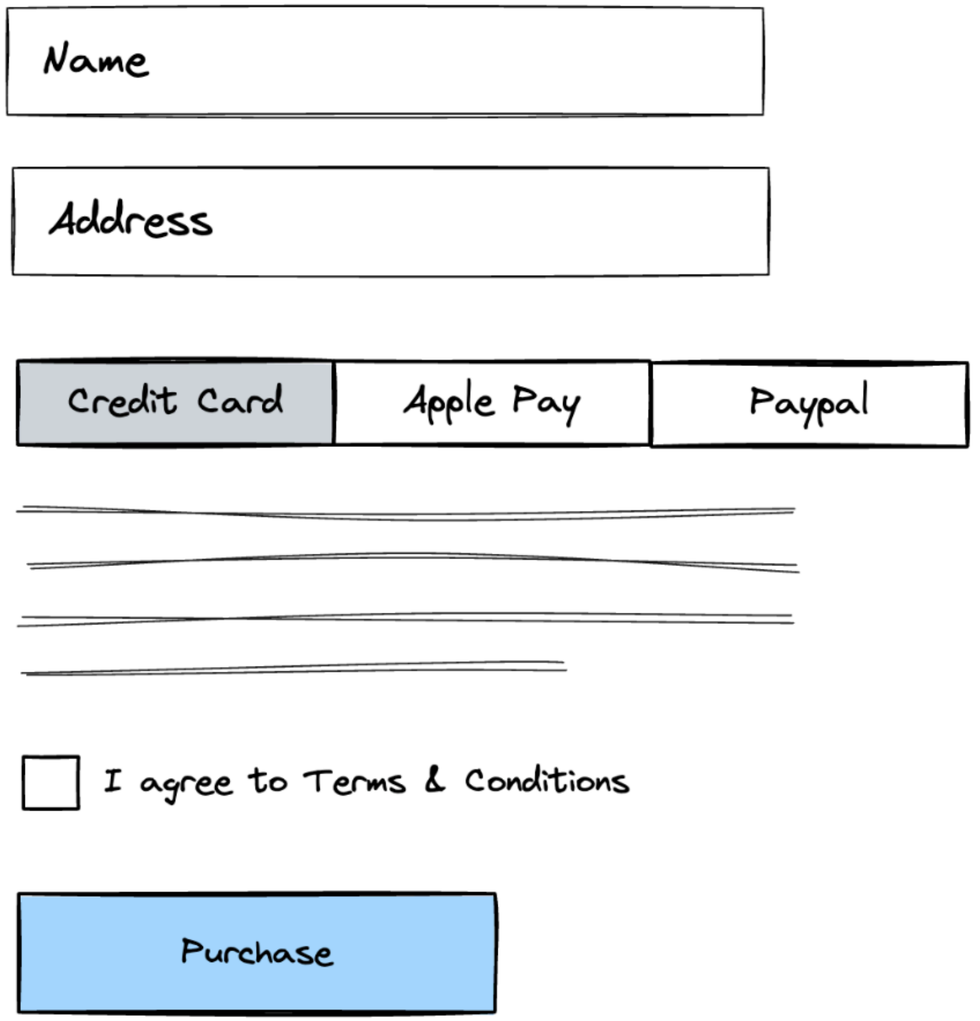
If clicks happen at the same point in a journey, we *can* compare them



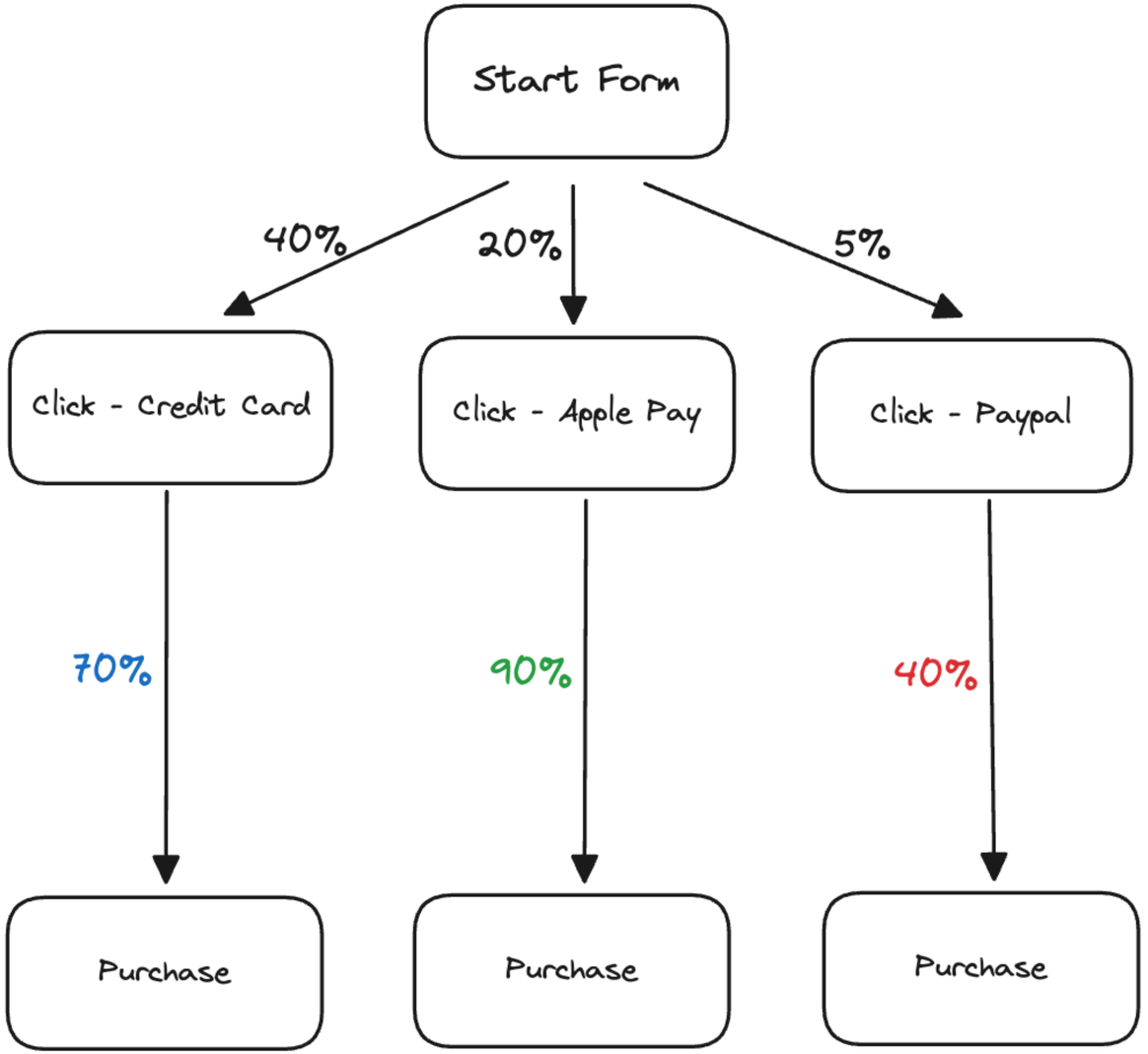

Start Form
40%
20%
5%
Click - Credit Card
Click - Apple Pay
Click - Paypal


Name
Address
Credit Card
Apple Pay
Paypal
I agree to Terms & Conditions
Purchase


Start Form
40%
20%
5%
Click - Credit Card
Click - Apple Pay
Click - Paypal
70%
90%
40%
Purchase
Purchase
Purchase

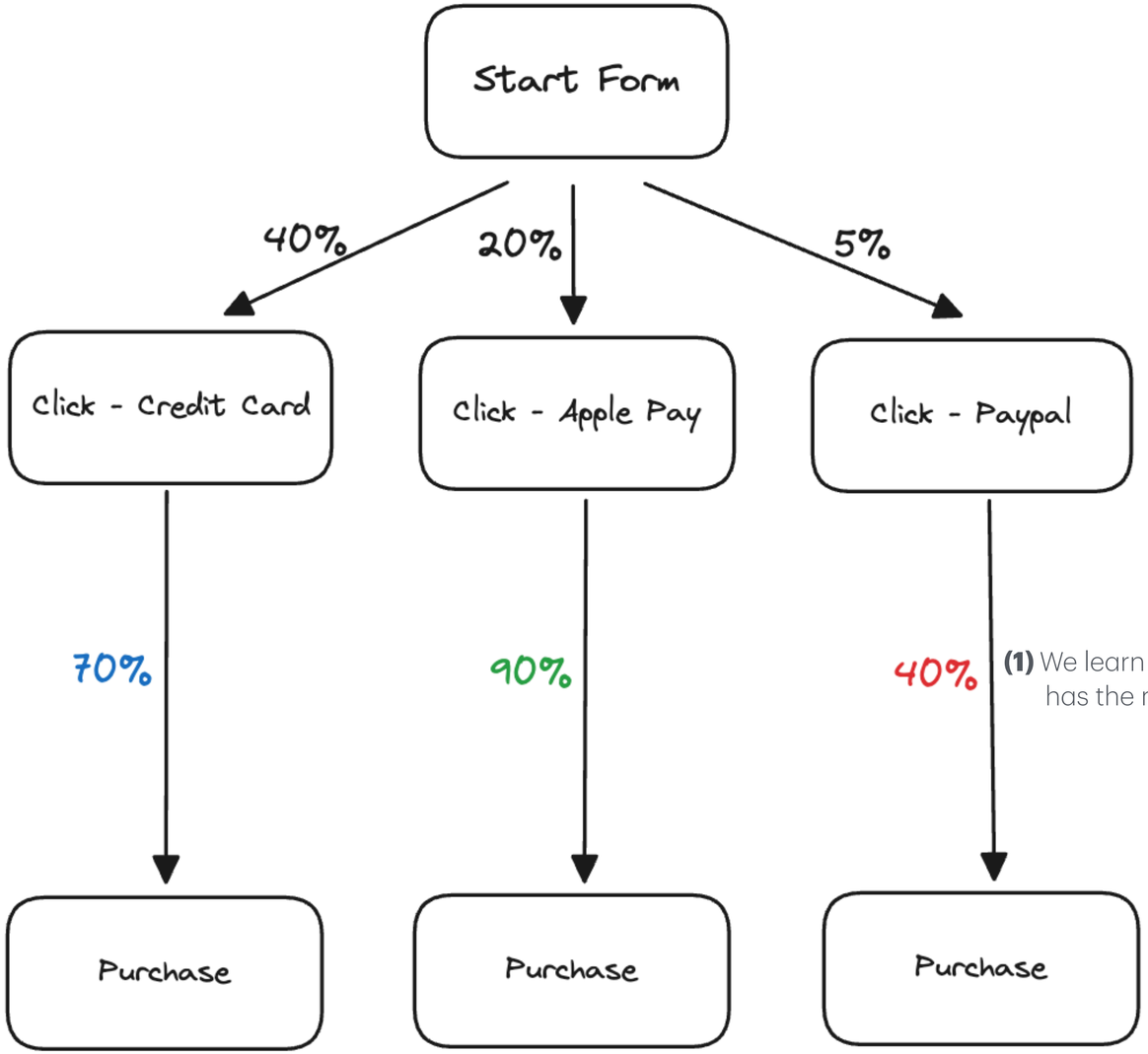

Name
Address
Credit Card
Apple Pay
Paypal
I agree to Terms & Conditions
Purchase

In causal inference terms, we're *matching*



How do we use a journey to reason about causal effects?

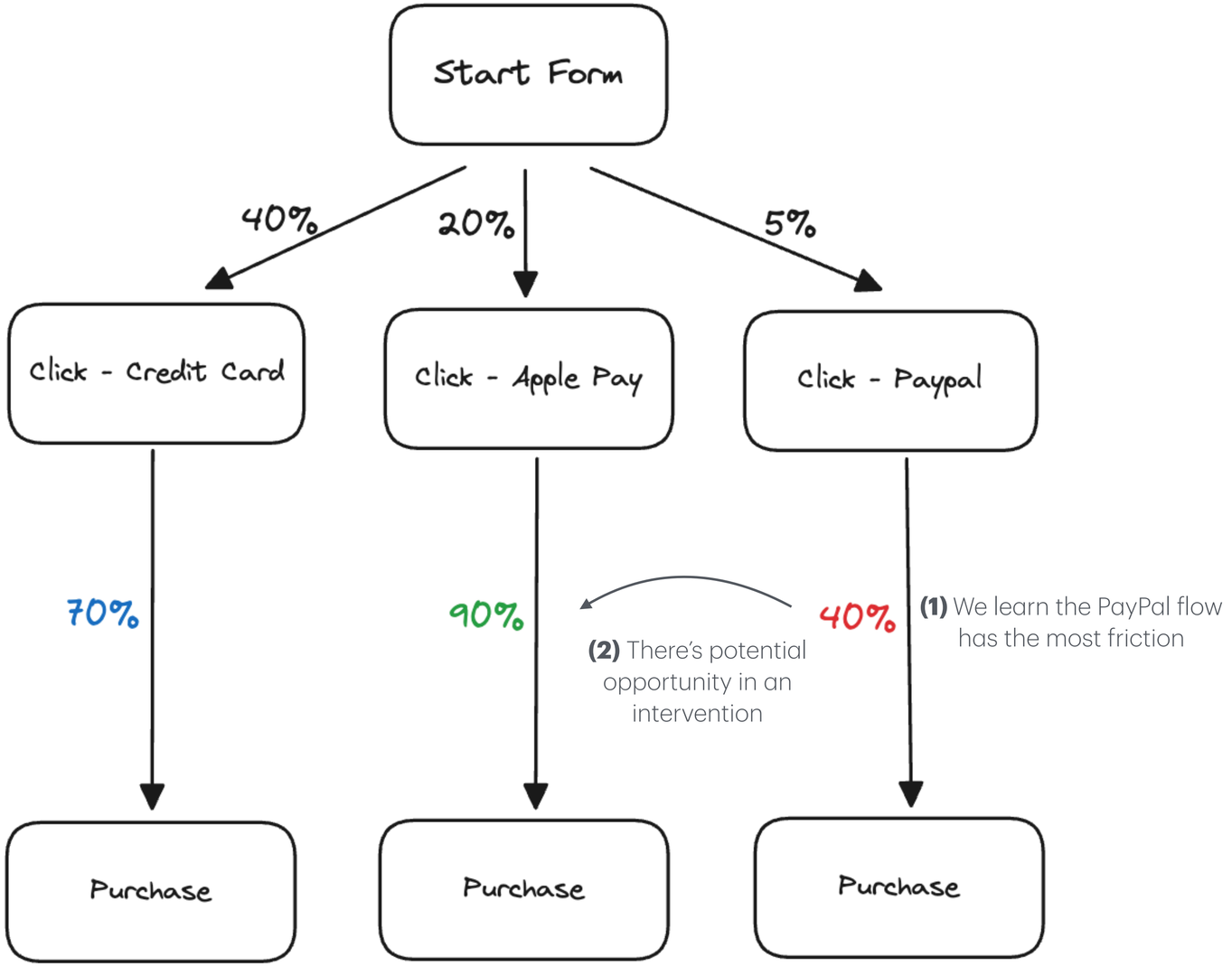
In causal inference terms, we're *matching*



(1) We learn the PayPal flow has the most friction

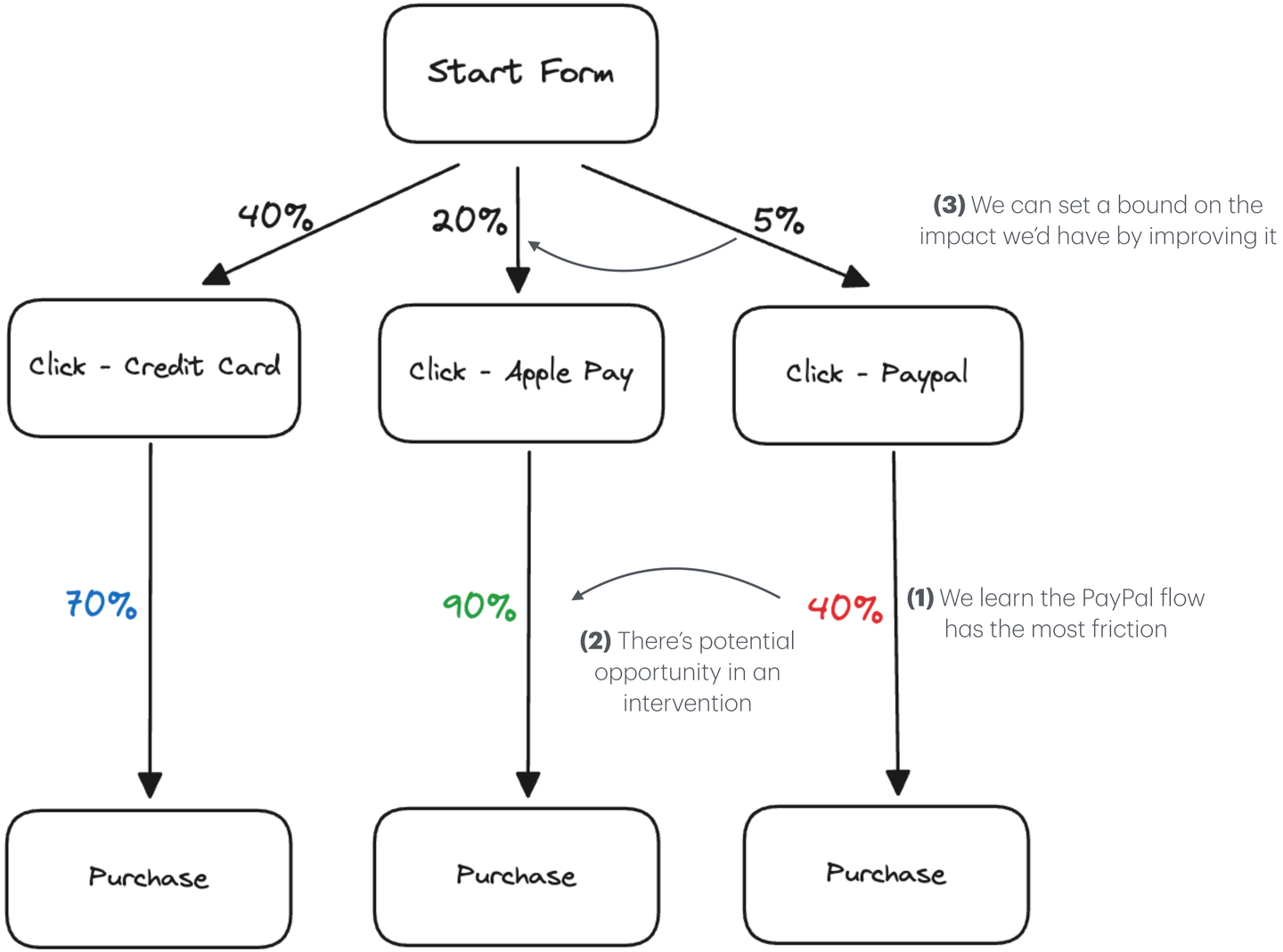
How do we use a journey to reason about causal effects?

In causal inference terms, we're *matching*

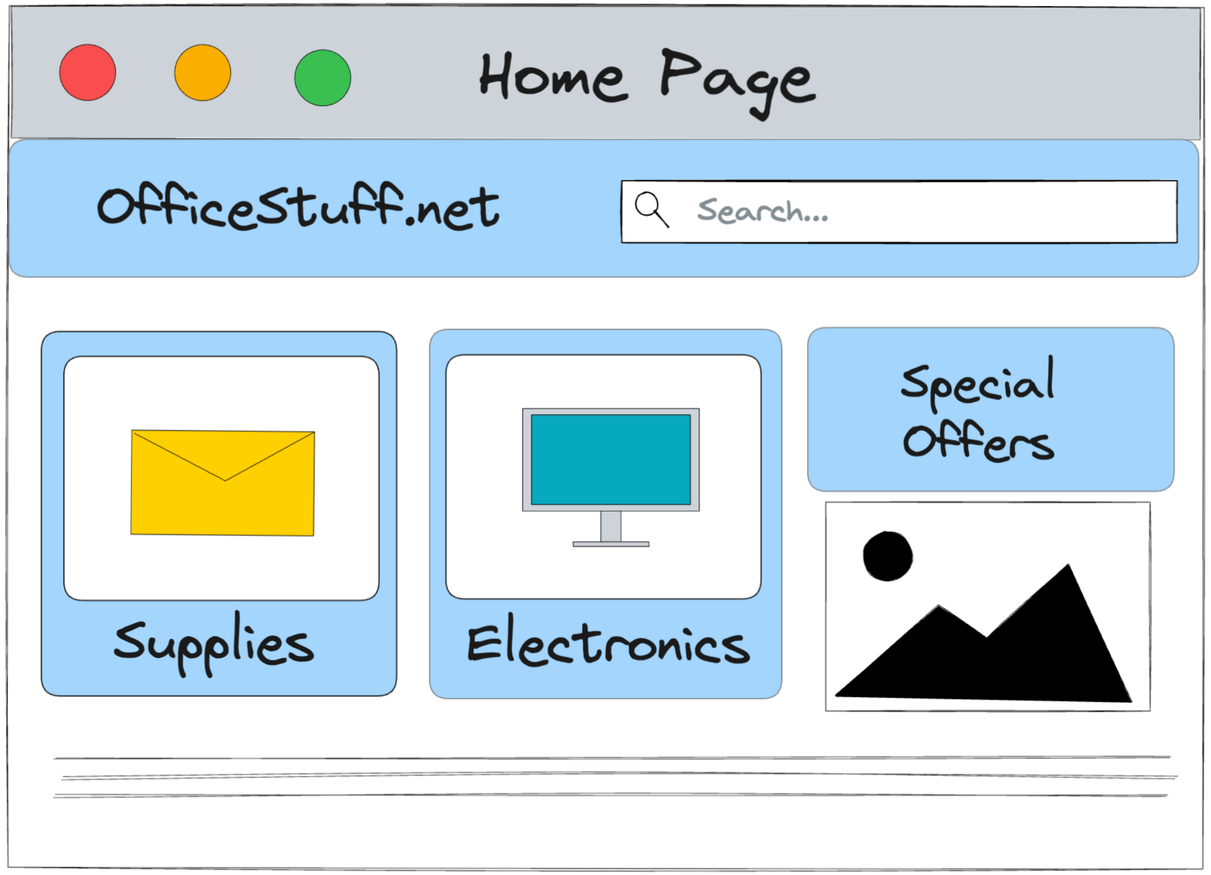


How do we use a journey to reason about causal effects?

In causal inference terms, we're *matching*

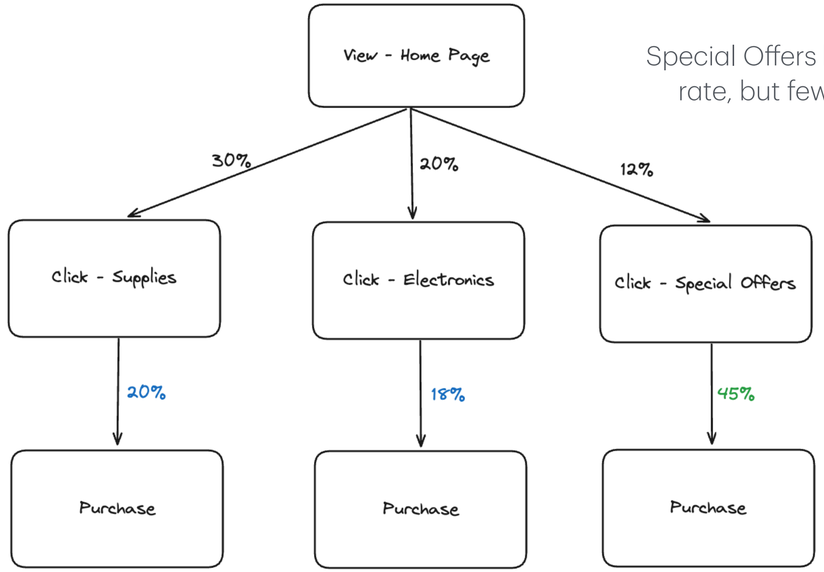


How do we use a journey to reason about causal effects?



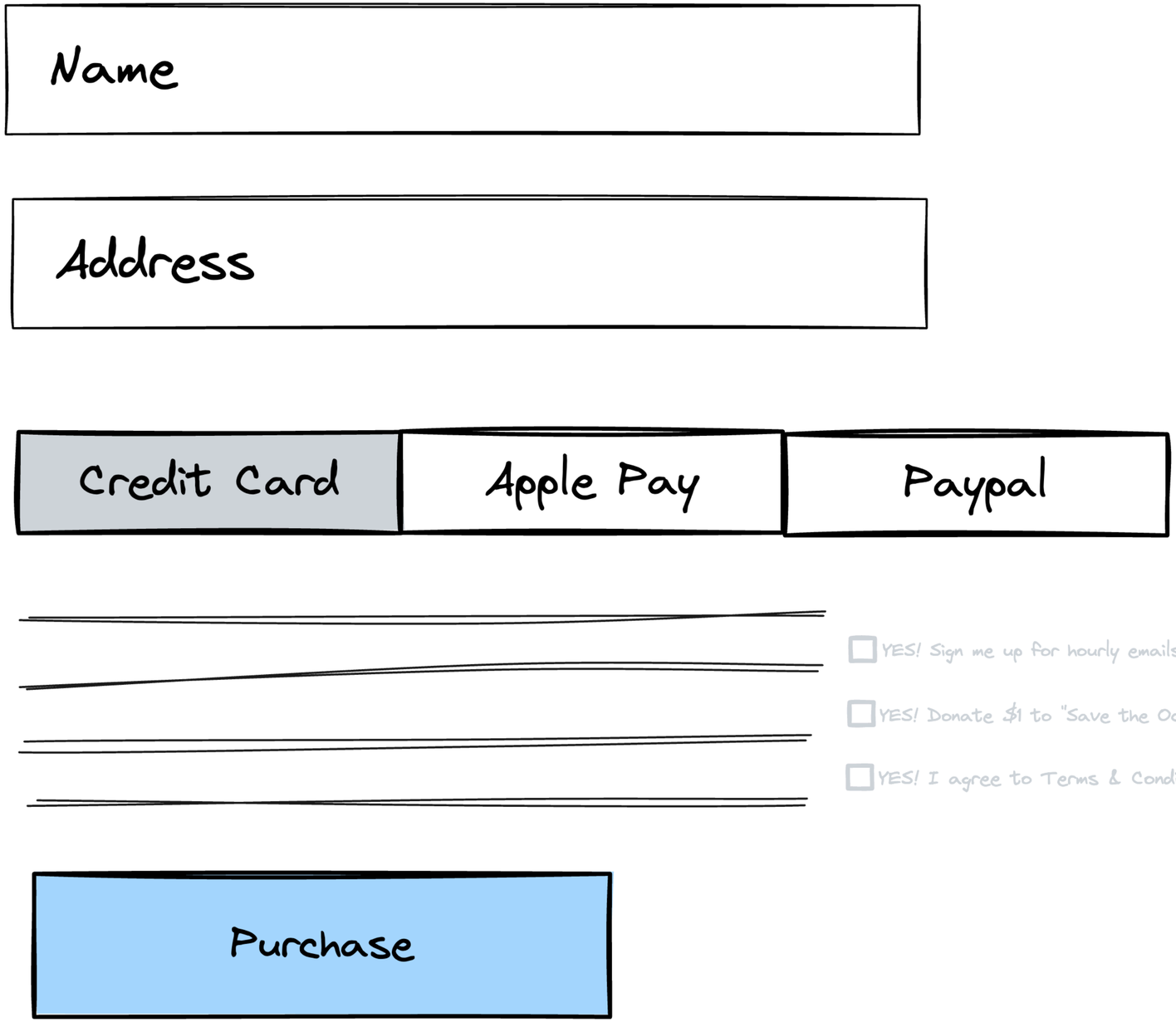
Opportunity

Special Offers has the highest success rate, but fewer users are exposed

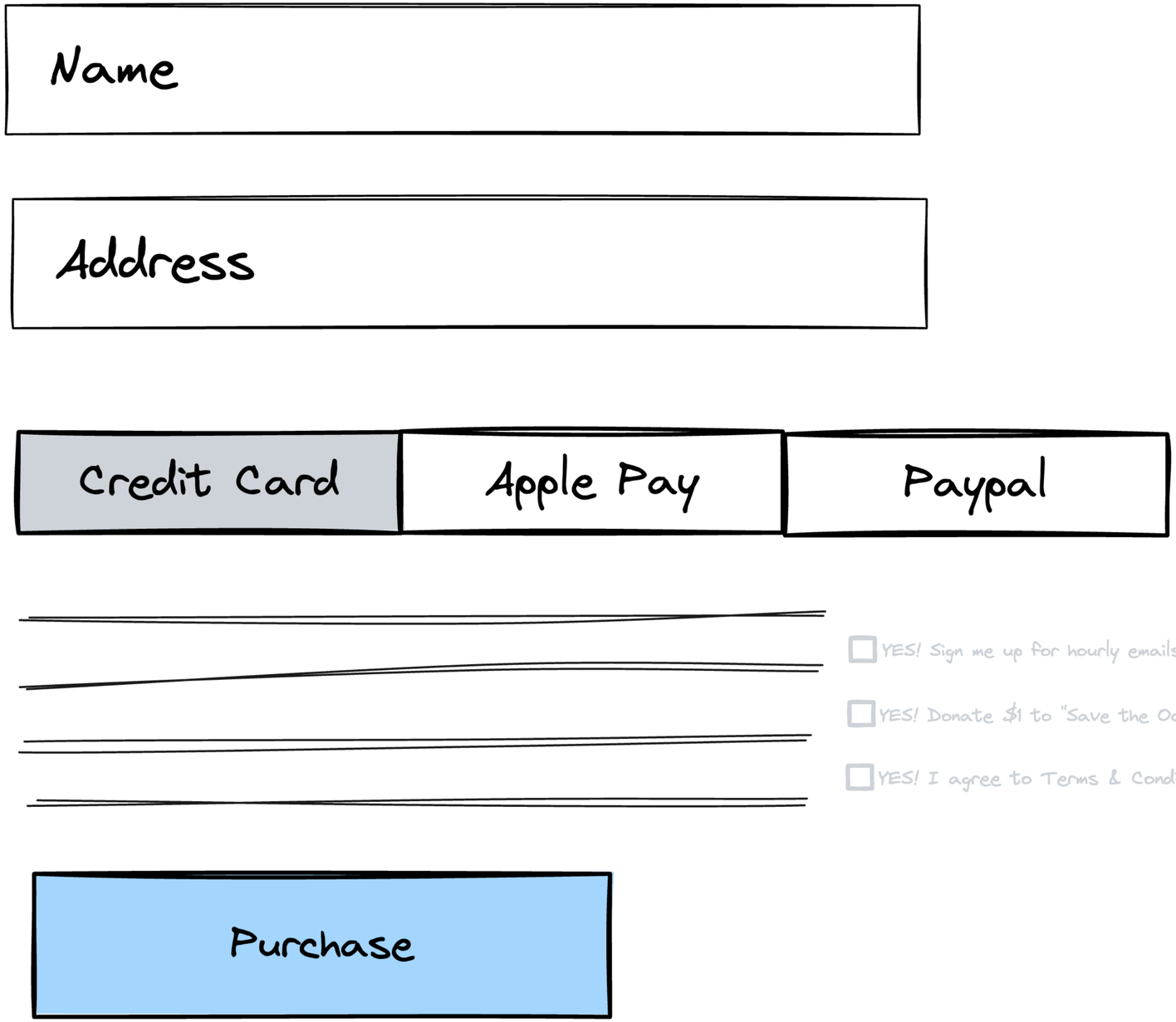


Forks can measure larger goals, not just the UX within one page

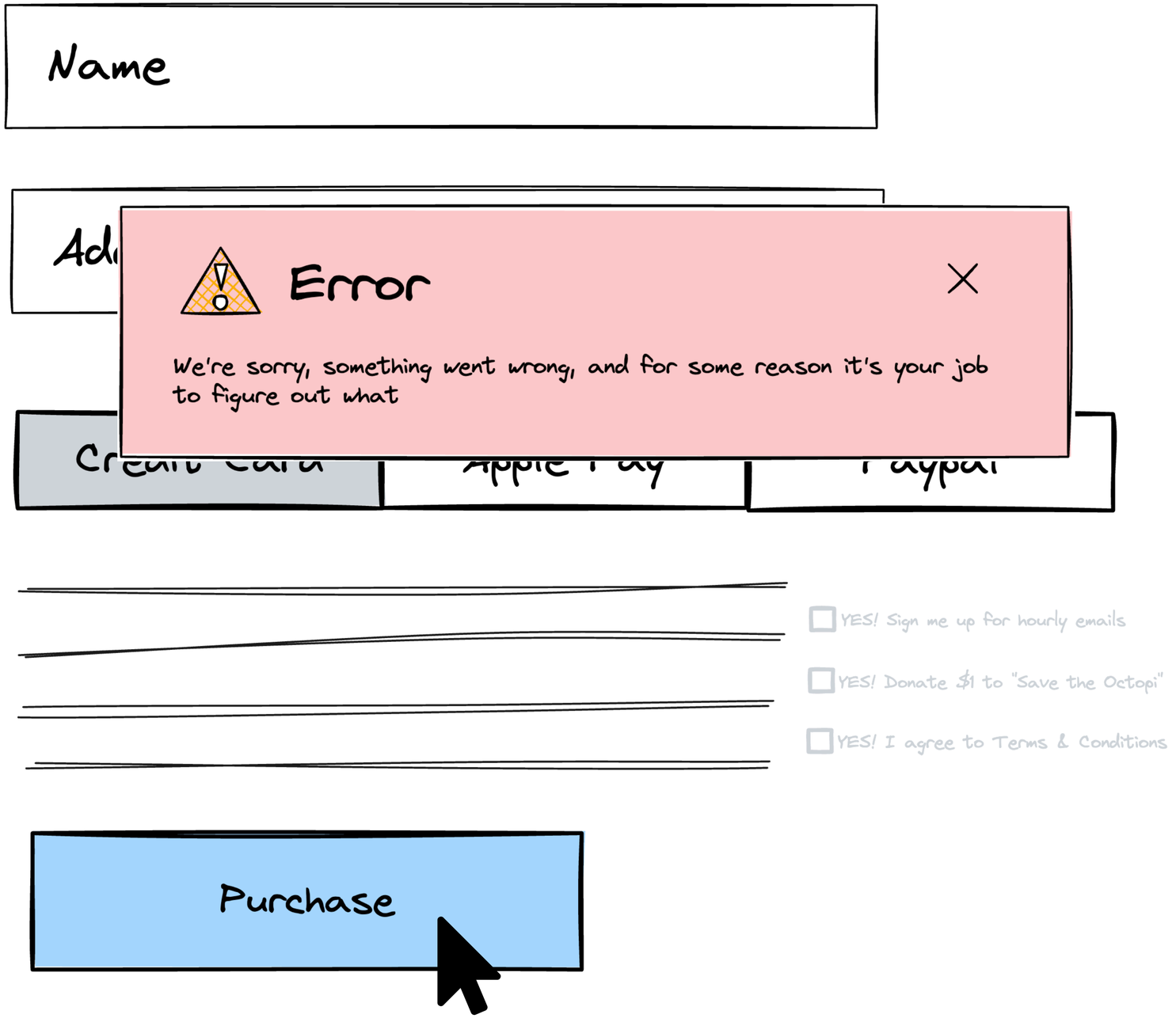
Journeys: Skippable Steps



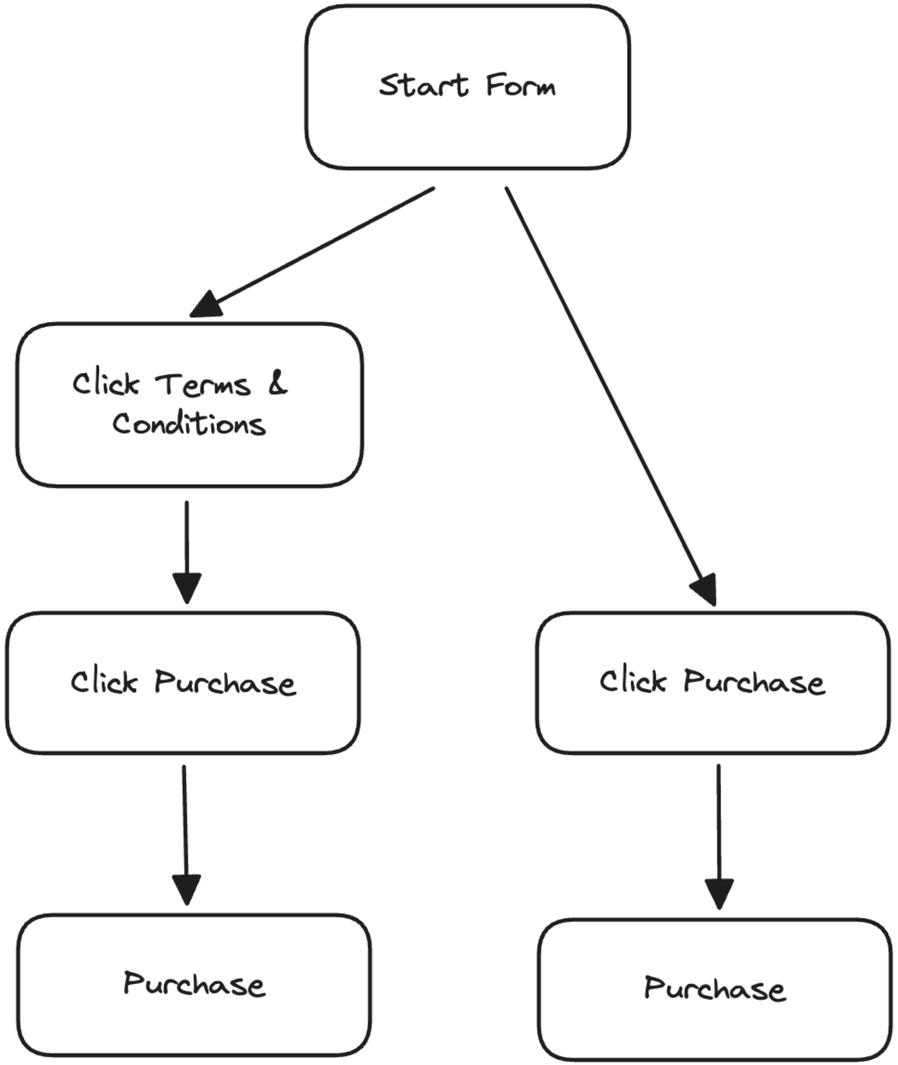
Bad UI can cause users to miss important steps

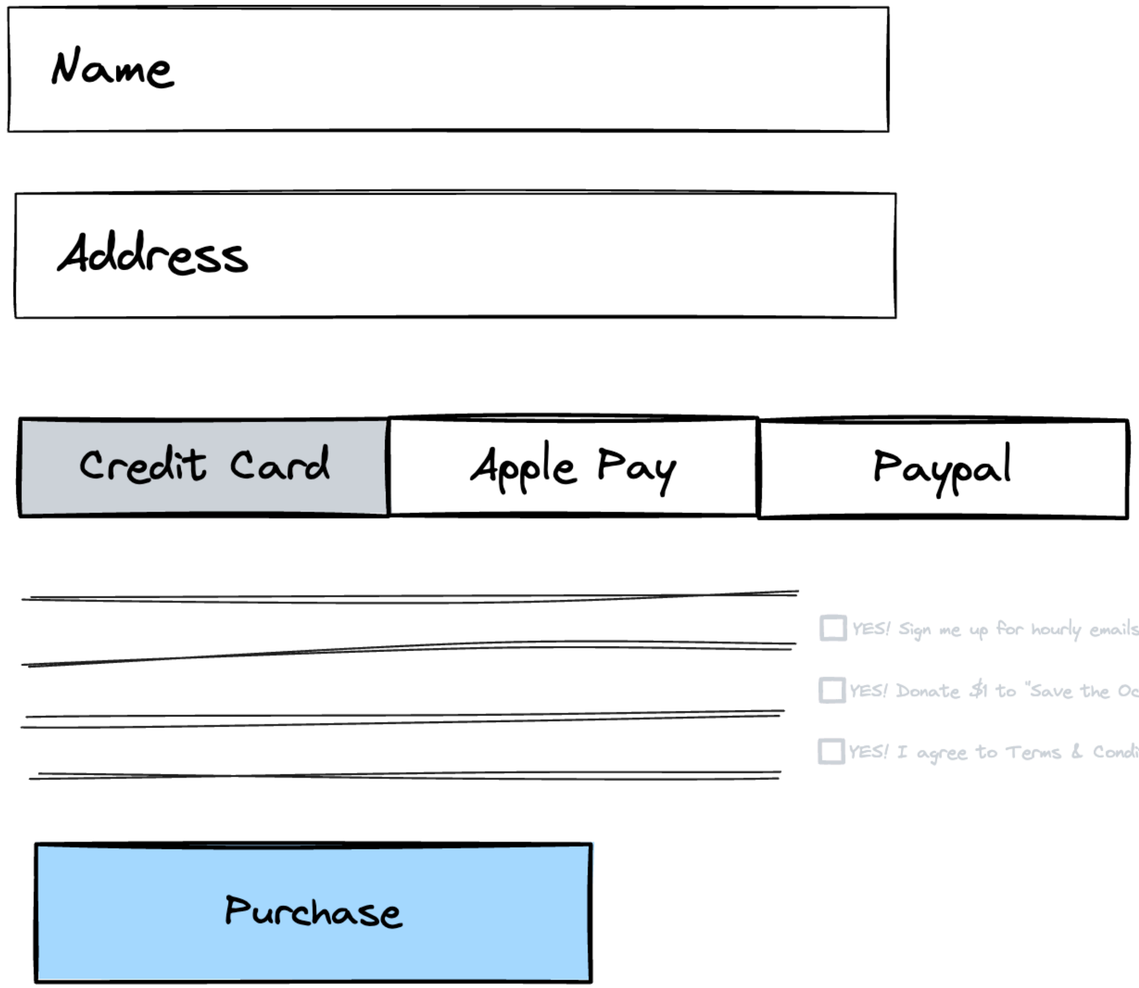


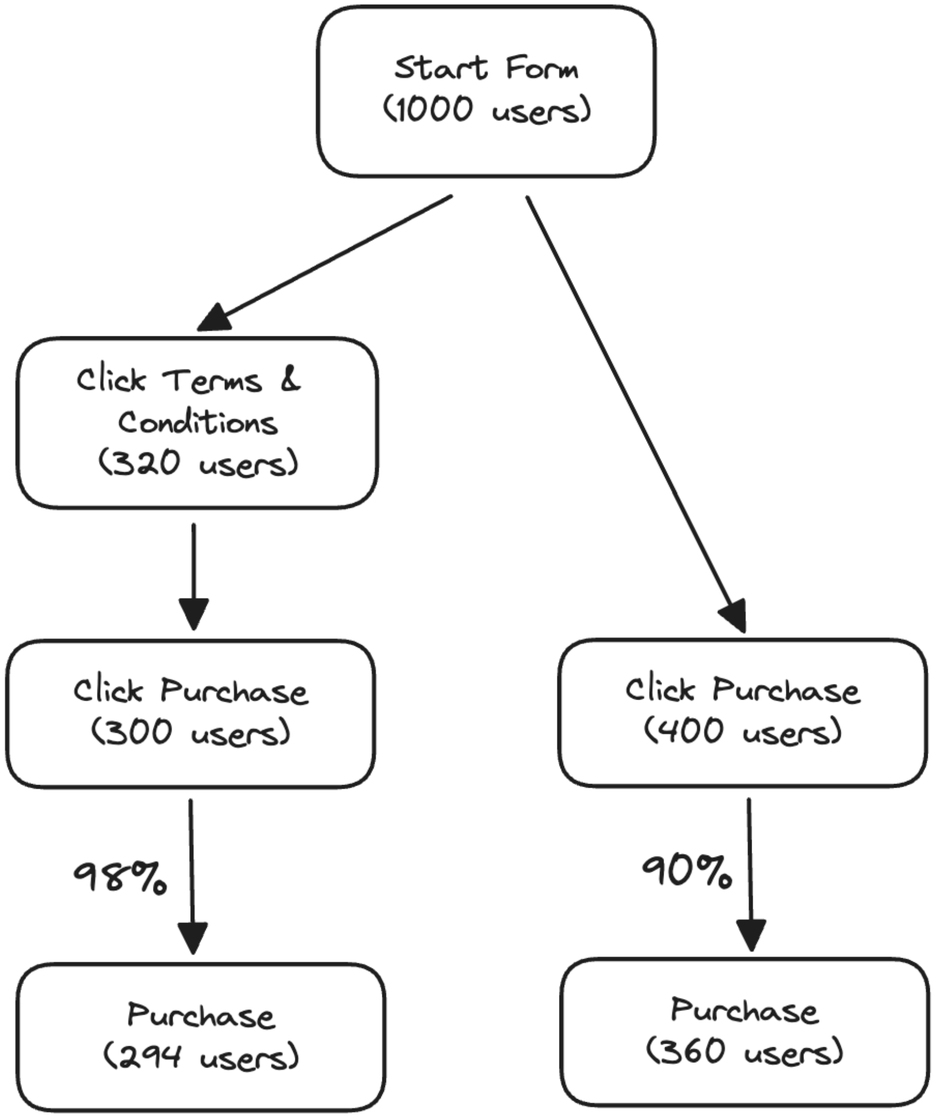
Bad UI can cause users to miss important steps

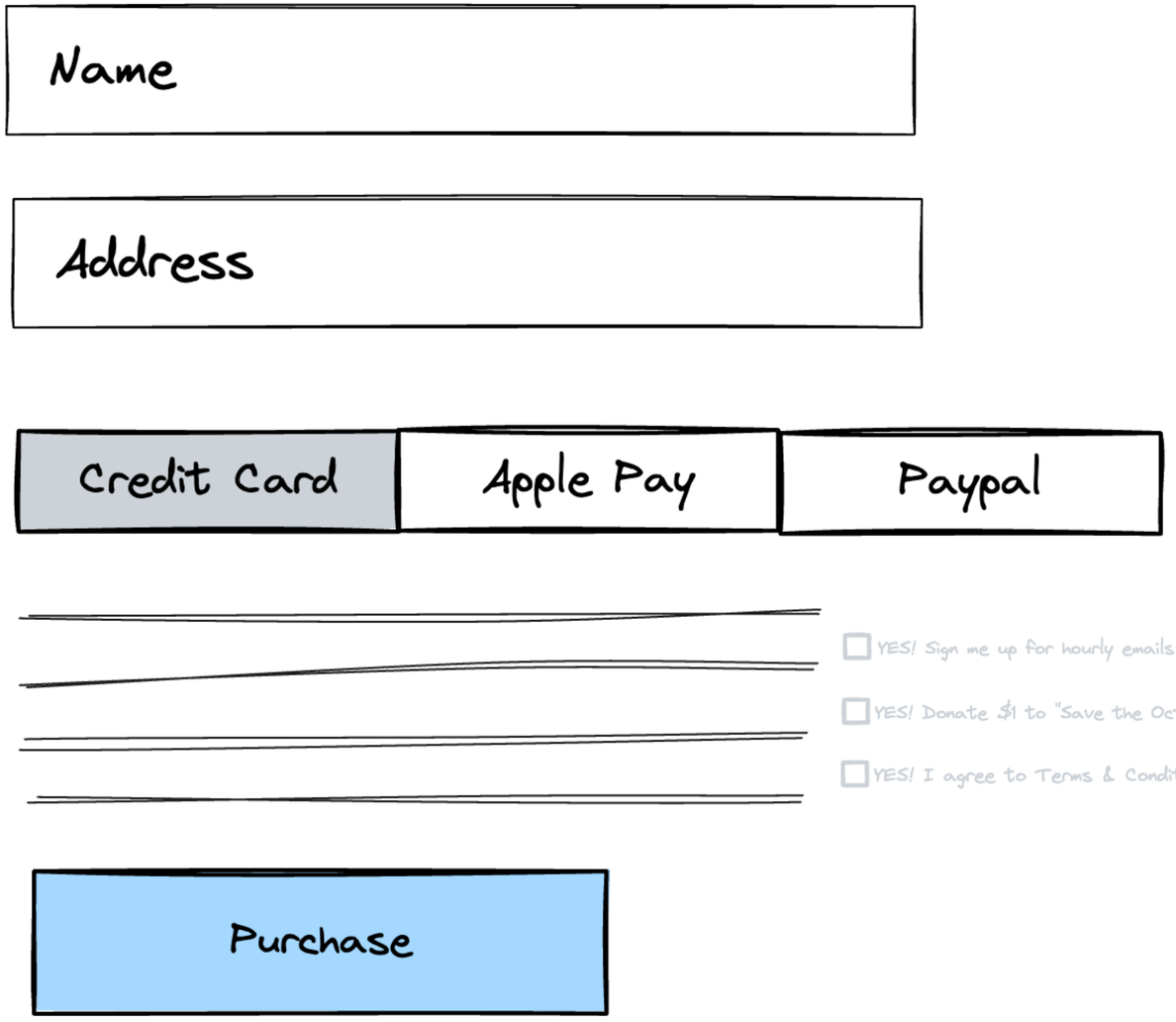


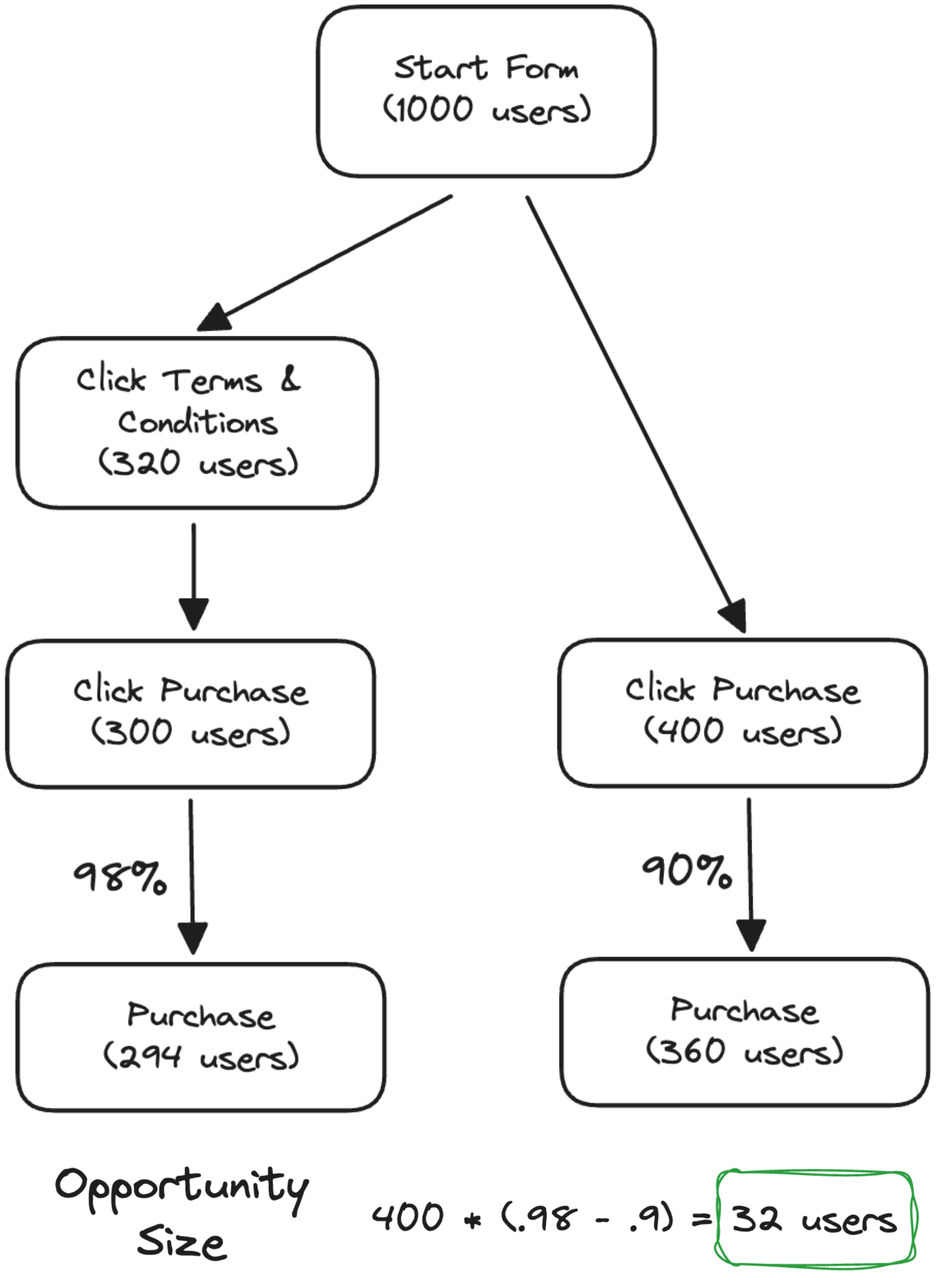
Bad UI can cause users to miss important steps

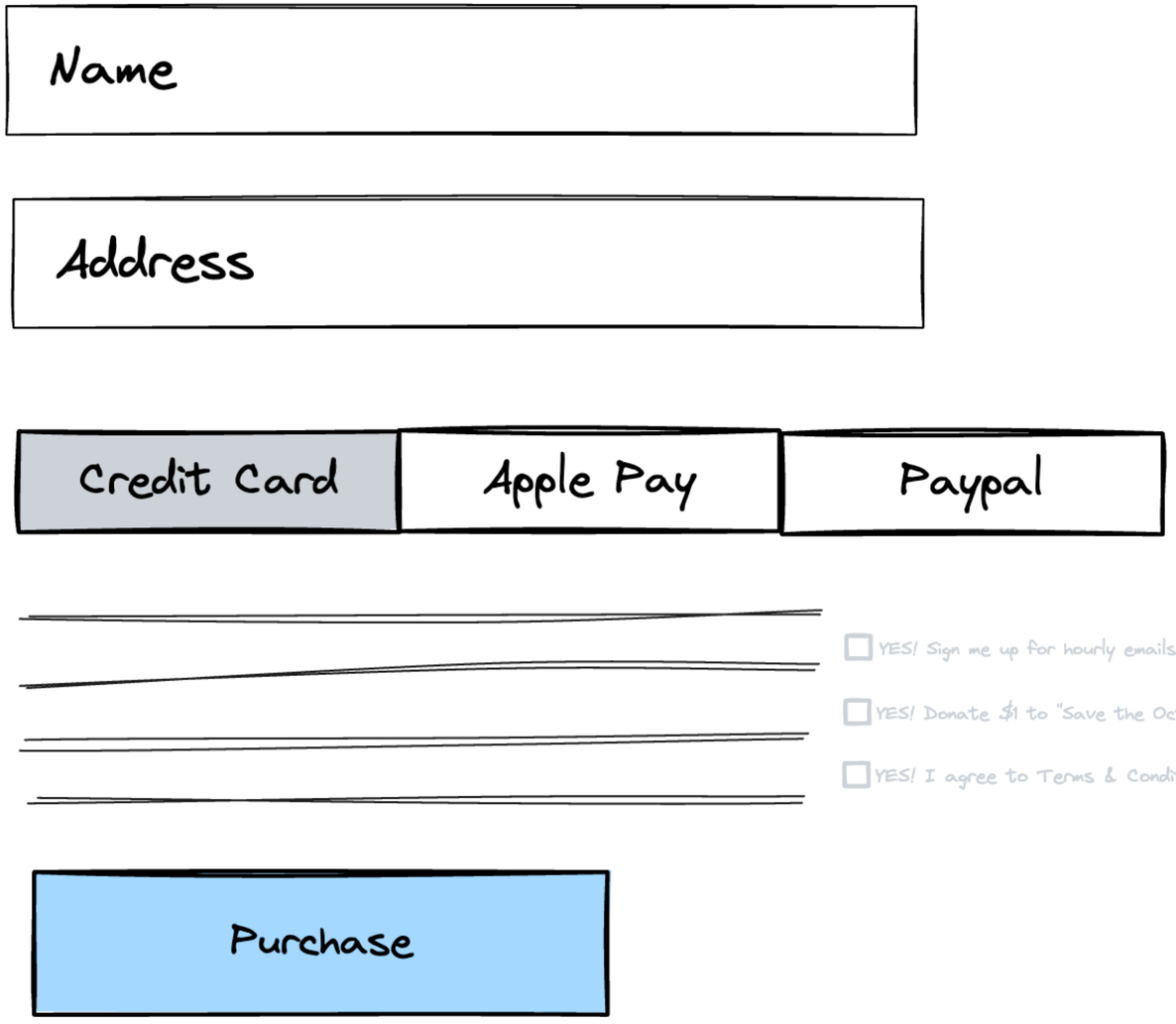

Start Form
Click Terms & Conditions
Click Purchase
Purchase
Click Purchase
Purchase

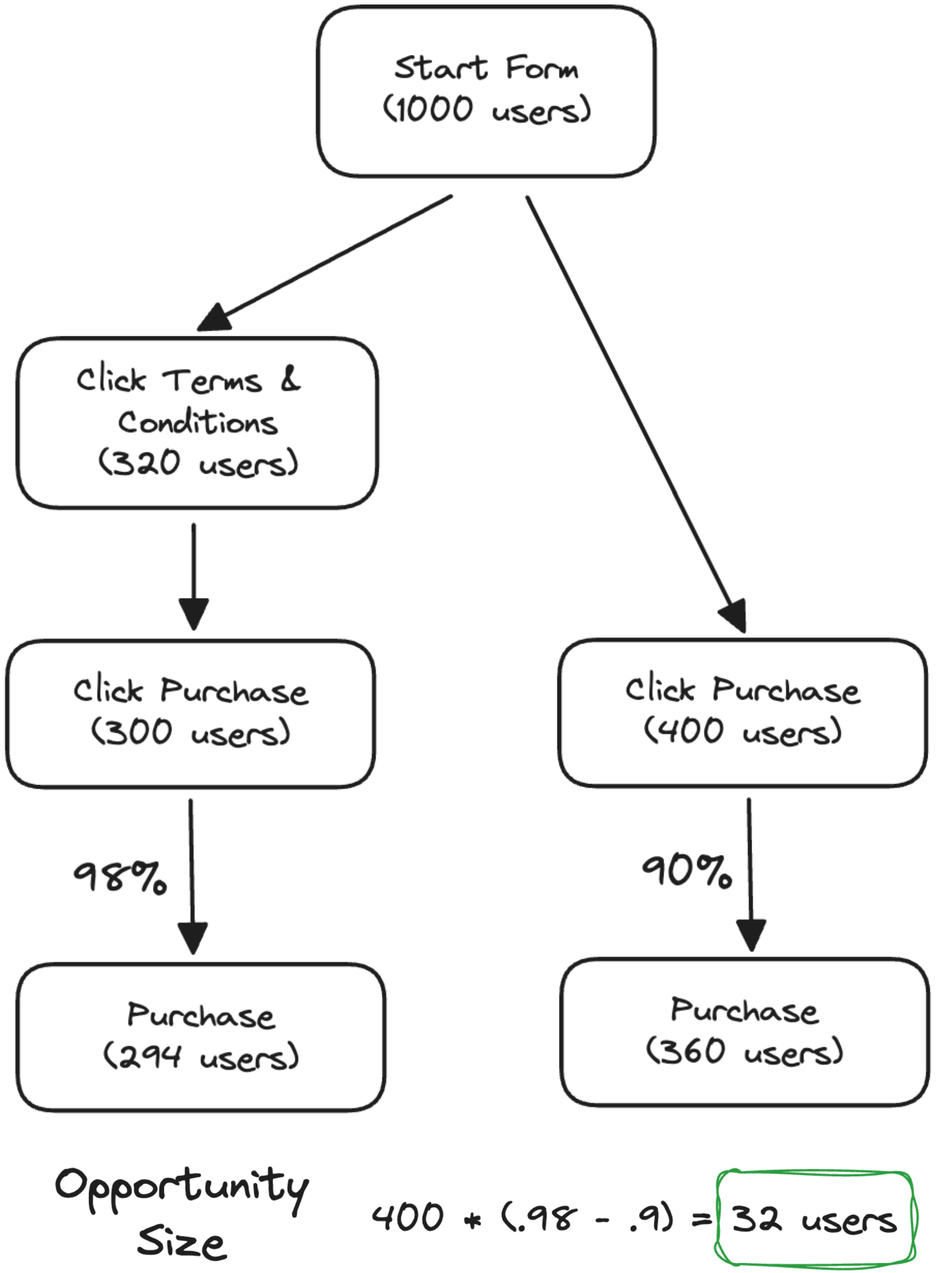

Name
Address
Credit Card
Apple Pay
Paypal
YES! Sign me up for hourly emails
YES! Donate $1 to "Save the Octopi"
YES! I agree to Terms & Conditions
Purchase


Start Form
(1000 users)
Click Terms &
Conditions
(320 users)
Click Purchase
(300 users)
98%
Purchase
(294 users)
Click Purchase
(400 users)
90%
Purchase
(360 users)


Name
Address
Credit Card
Apple Pay
Paypal
YES! Sign me up for hourly emails
YES! Donate $1 to "Save the Octopi"
YES! I agree to Terms & Conditions
Purchase


Start Form
(1000 users)
Click Terms & Conditions
(320 users)
Click Purchase
(300 users)
Click Purchase
(400 users)
98%
90%
Purchase
(294 users)
Purchase
(360 users)
Opportunity Size
400 * (.98 - .9) = 32 users


Name
Address
Credit Card
Apple Pay
Paypal
YES! Sign me up for hourly emails
YES! Donate $1 to "Save the Octopi"
YES! I agree to Terms & Conditions
Purchase

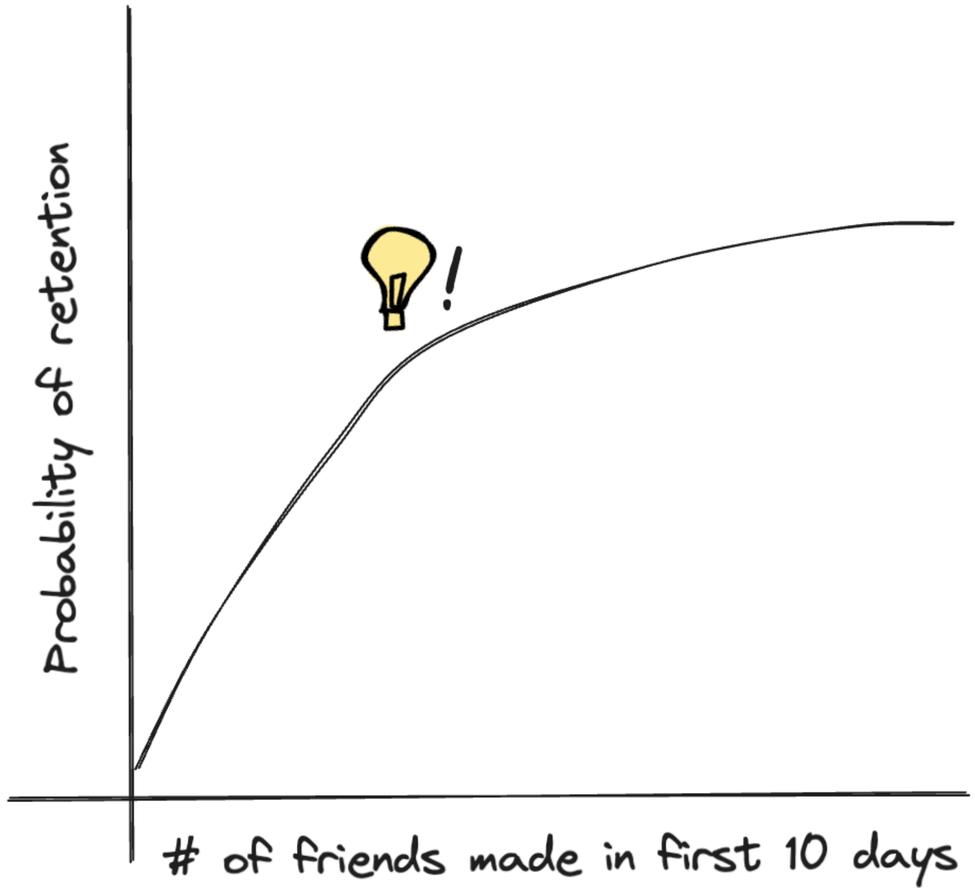


This information is *lost* in a bag-of-clicks approach

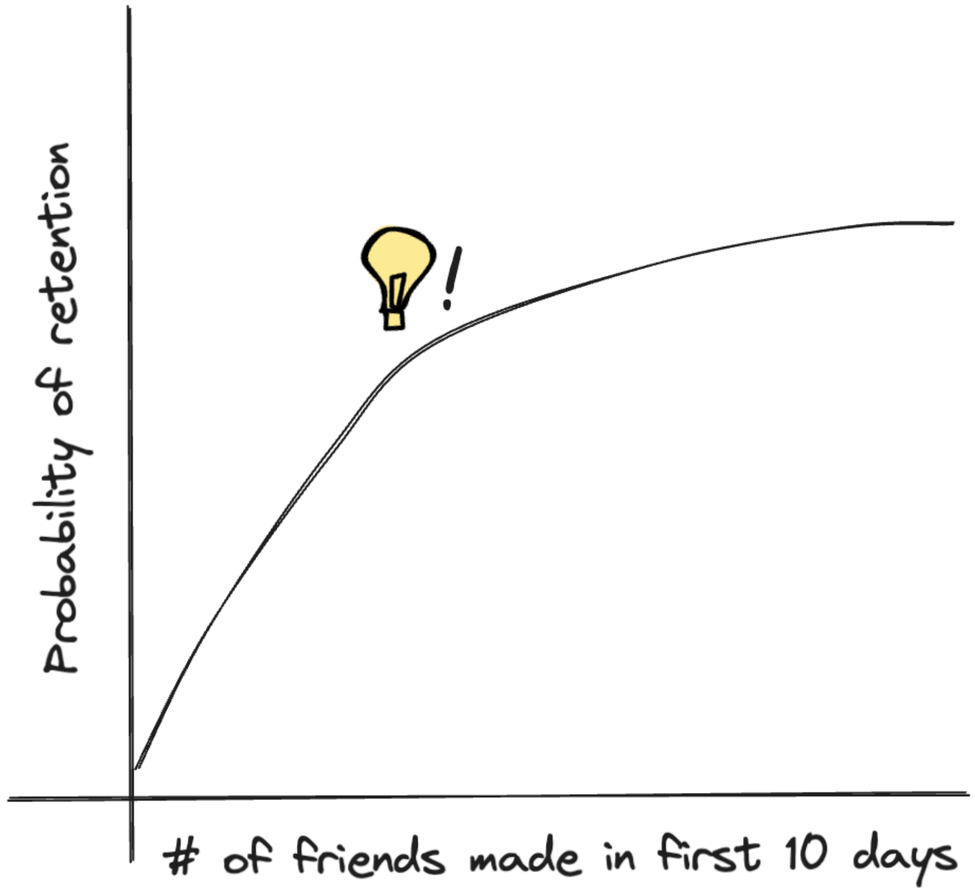
user_id	viewed_home	clicked_search	clicked_read_reviews	clicked_learn_more	...yada yada...	purchased
187180	0	1	1	0	...	1
662863	0	0	0	0	...	1
684172	1	1	1	0	...	0
261982	1	1	0	0	...	0
579325	1	1	1	1	...	0
829755	1	1	0	1	...	0
454574	1	1	0	0	...	0
216105	1	1	1	1	...	1
181215	0	1	0	0	...	0
996510	1	1	0	0	...	0
379967	1	1	0	0	...	0
511056	1	1	0	1	...	0
948165	1	1	1	0	...	0
575719	1	1	0	1	...	0
356383	1	1	1	0	...	0
582609	1	1	1	0	...	0
988781	1	0	0	1	...	0
857058	1	1	0	0	...	0
498438	1	1	0	0	...	0
470140	1	0	1	0	...	0
714748	1	0	0	1	...	0
624947	1	1	1	1	...	0
524455	1	0	0	1	...	1

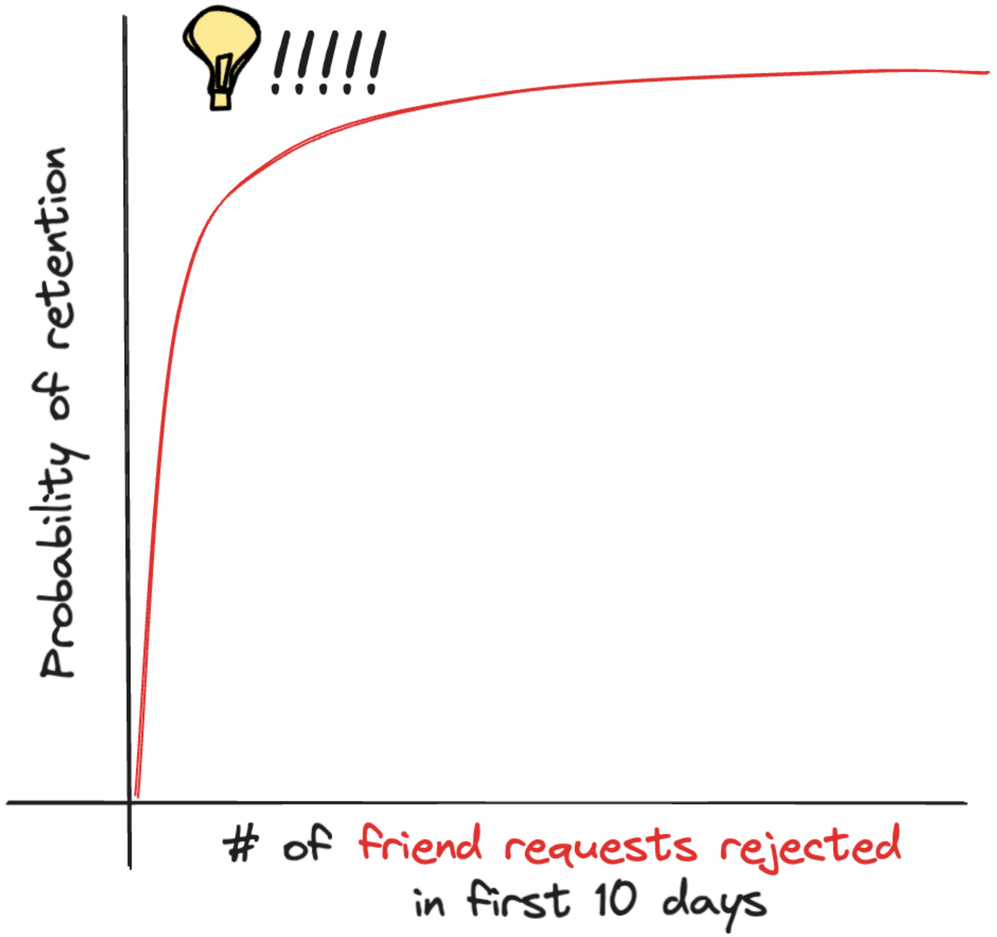
Retention

Facebook's "aha" moment

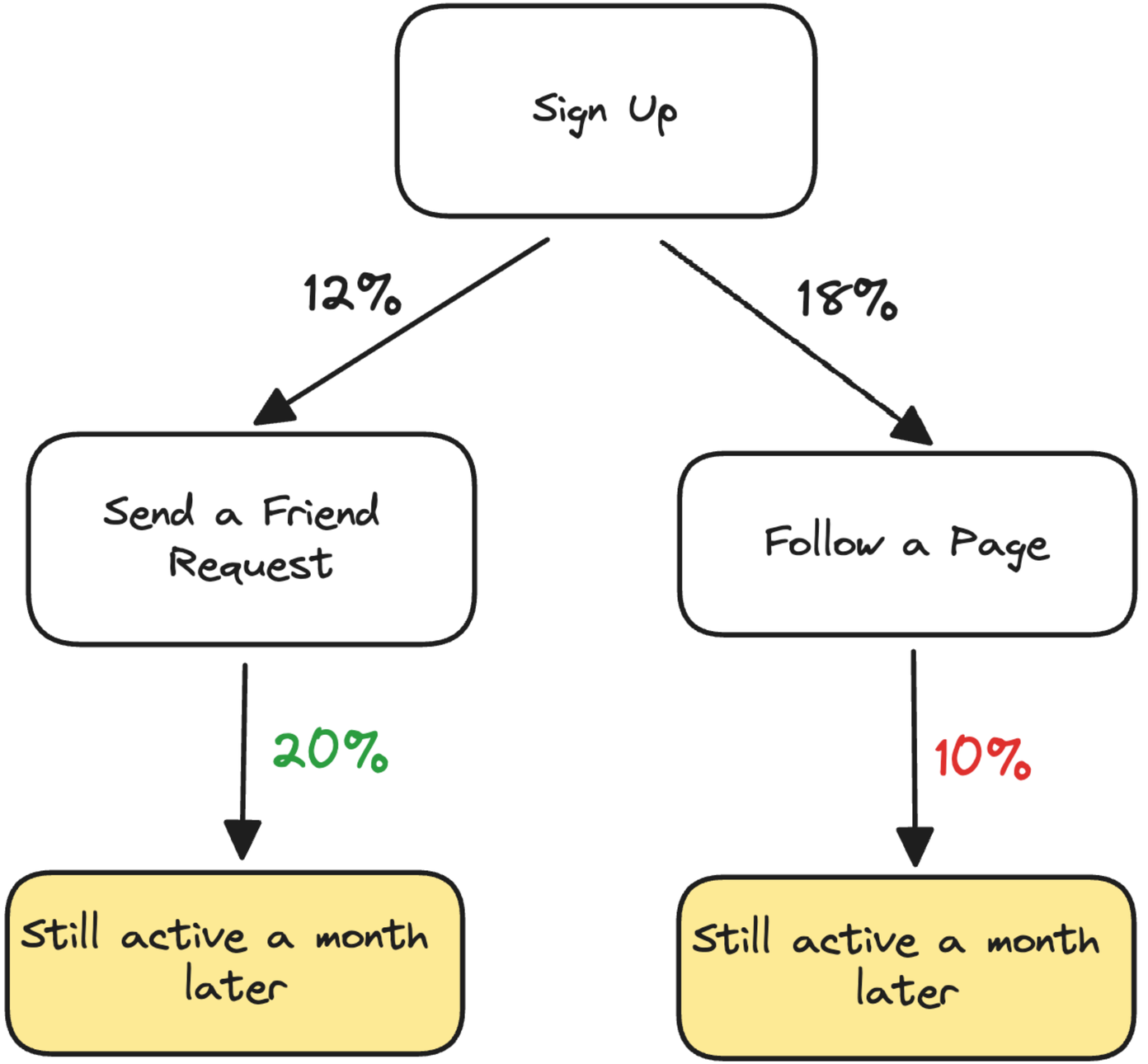


Facebook's "aha" moment?





What I'd rather see- find two actions that represent the same moment in a journey, and *compare* them



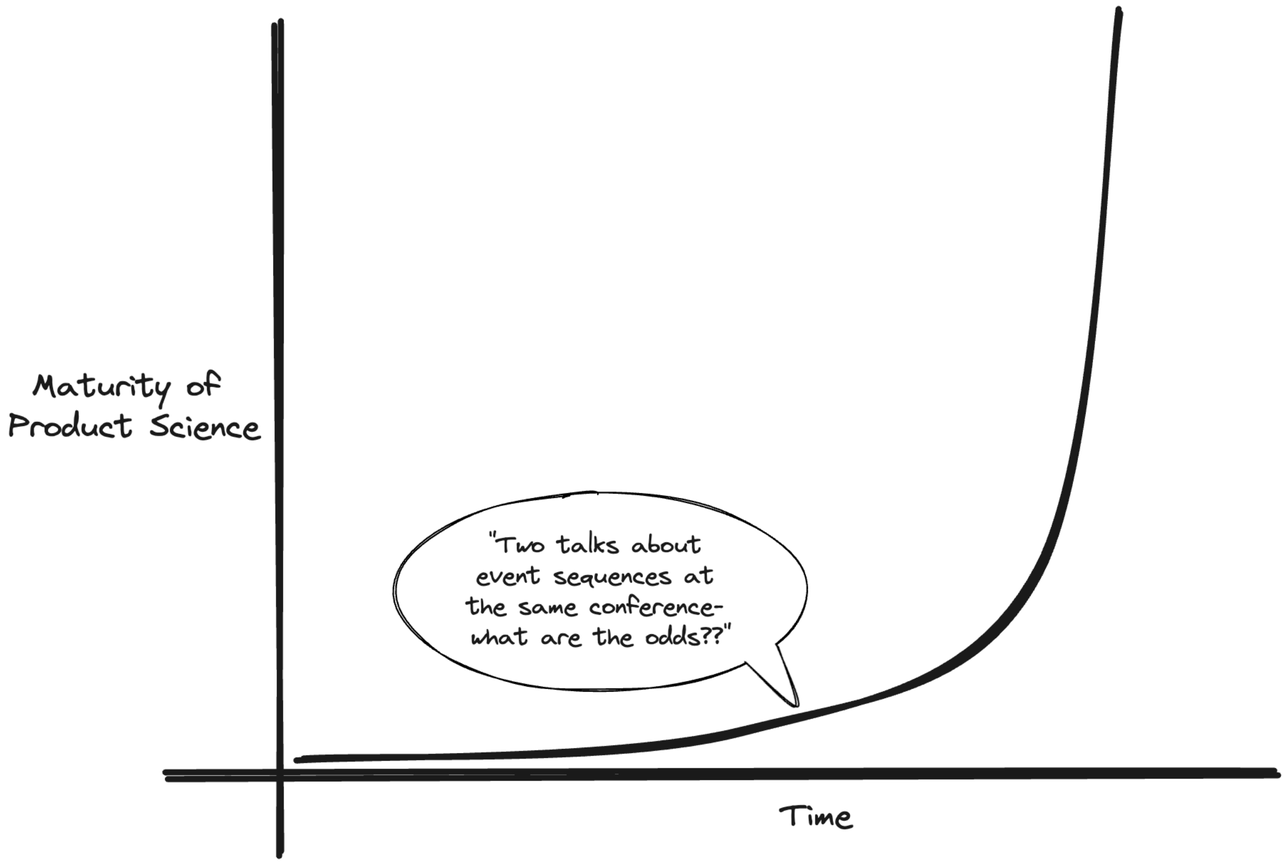
Conclusion

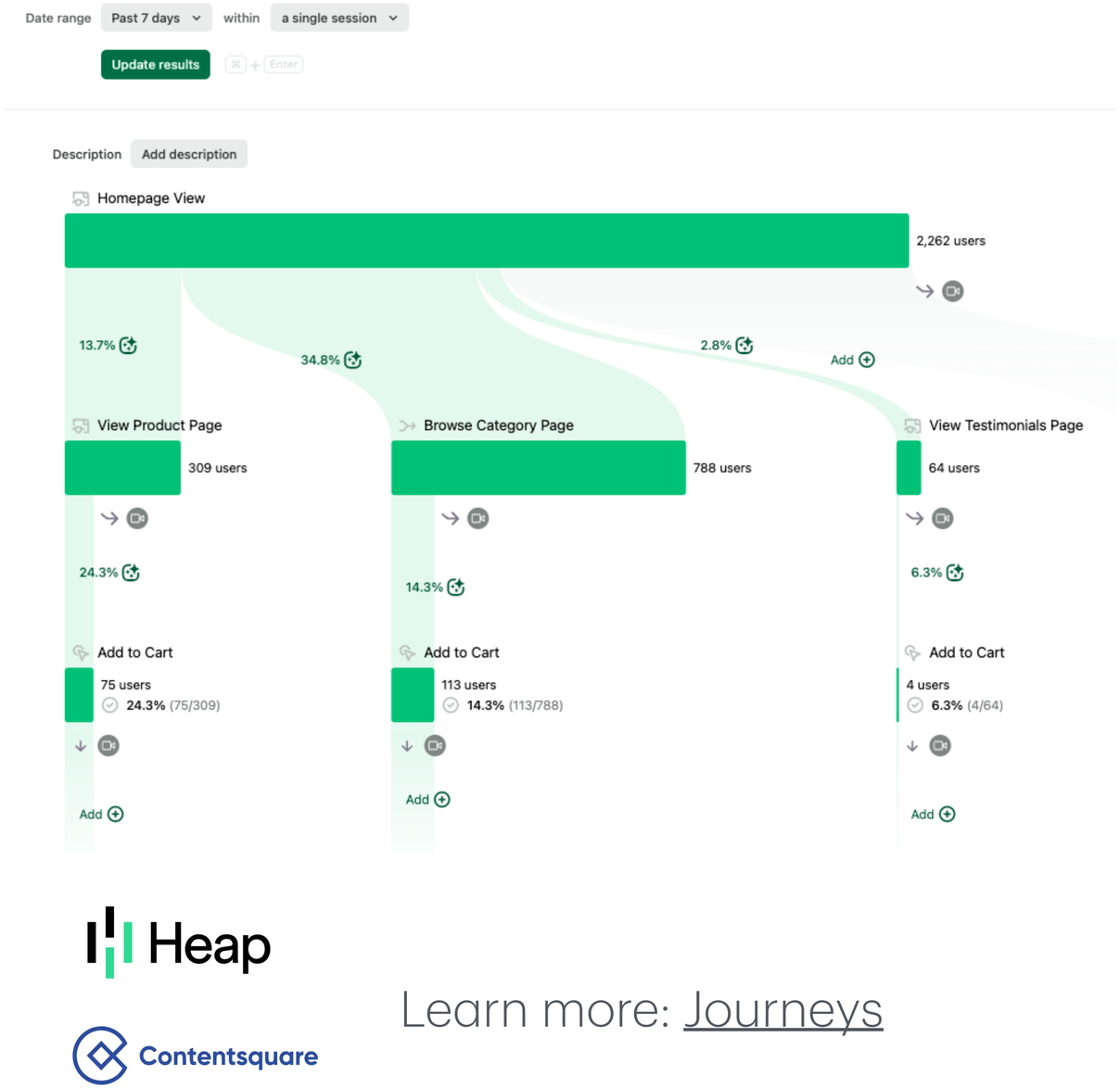
“After growing wildly for years, the field of computing appears to be reaching its infancy.”

John Pierce, Bell Labs

"After growing wildly for years, the field of computing appears to be reaching its infancy."

John Pierce, Bell Labs

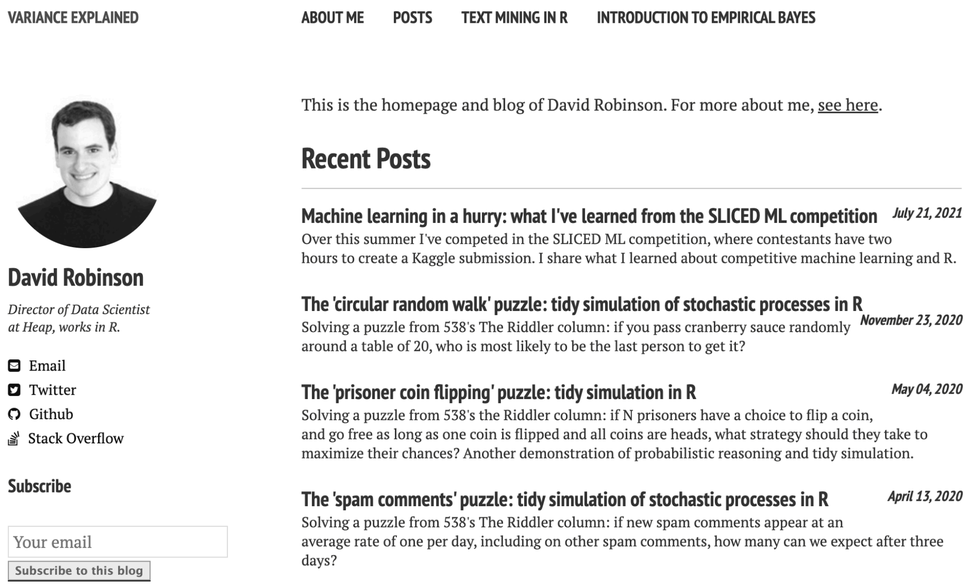




Heap

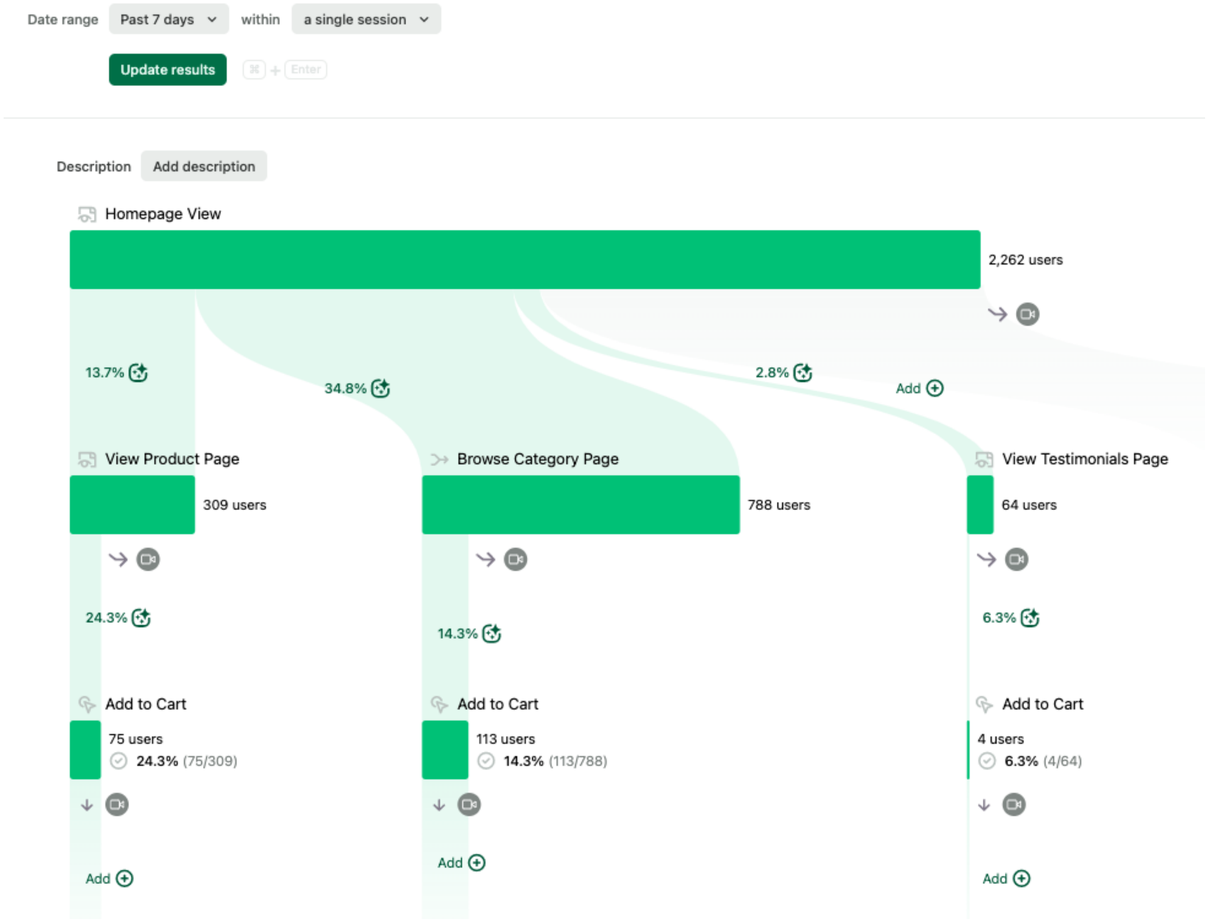
Contentsquare

Learn more: Journeys



varianceexplained.org

- Nicole Del Giudice
- Jared Lander
- Teams at Heap & Contentsquare

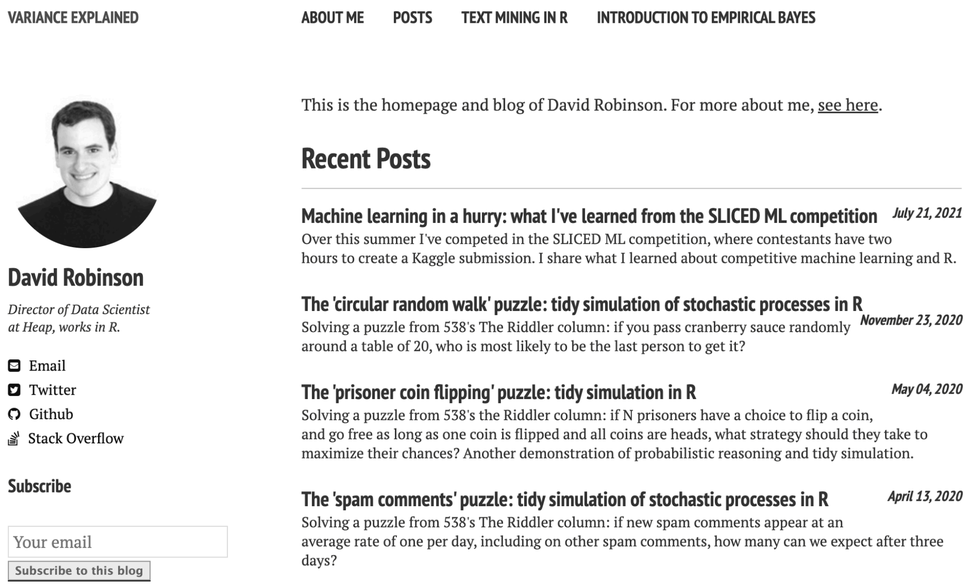


Heap

Contentsquare

Learn more: Journeys

Thank you!



varianceexplained.org

- Nicole Del Giudice
- Jared Lander
- Teams at Heap & Contentsquare