Inferring the Long-Term Causal Effects of Long-Term Treatments from Short-Term Experiments

Allen Tran, Aurélien Bibaut, Nathan Kallus
ICML 2024

Long-term Treatments, Short-term Tests

Long term outcomes are of primary importance

Tests are often short (ethical concerns, business constraints, etc) even when we care about a “long-term treatment”*:*

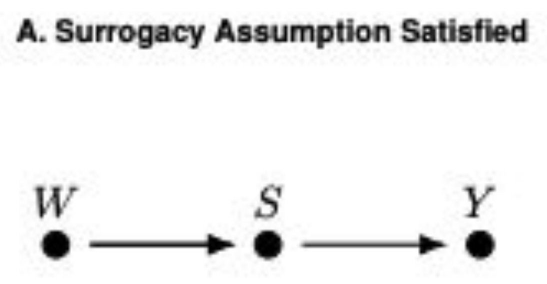
- *continuous exposure to a novel intervention that extends beyond the length of the experiment*

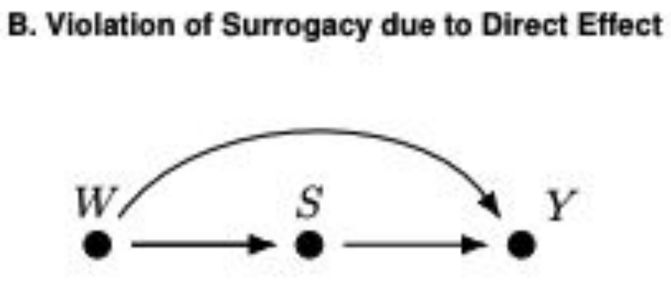
How can we measure the causal effect on long-term outcomes from a long-term treatment when experiments are short?

Treatment Duration and Surrogacy Assumptions

Large literature using surrogates (short-term proxies for long term outcomes)

Surrogate assumption can only hold for **short term** treatments





But many treatments of interest are **long-term**

What's in the paper?

Method to estimate long-term effects of long-term treatment from short experiment:

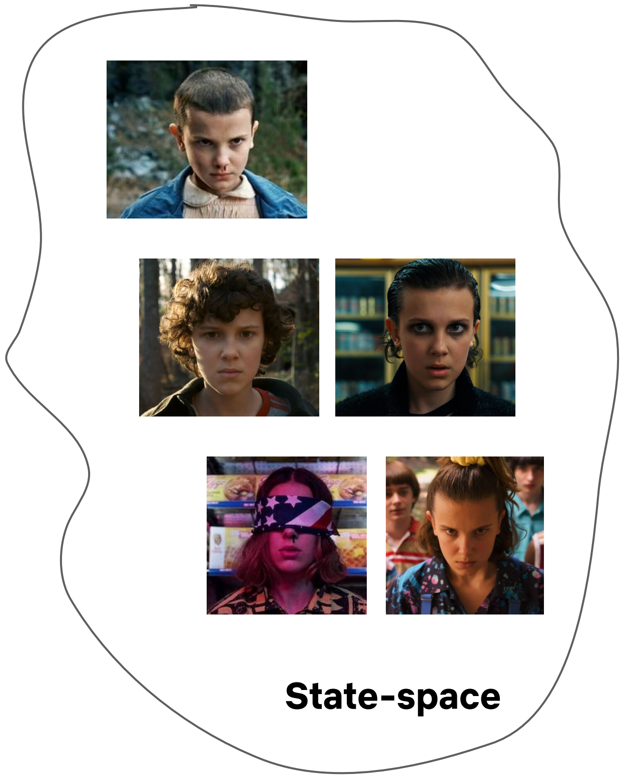
- no surrogacy assumptions
- no need for an observational dataset

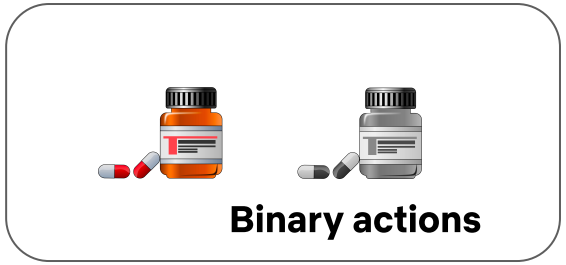
Identification proofs + assumptions (express estimand as the difference in Q functions via offline RL)

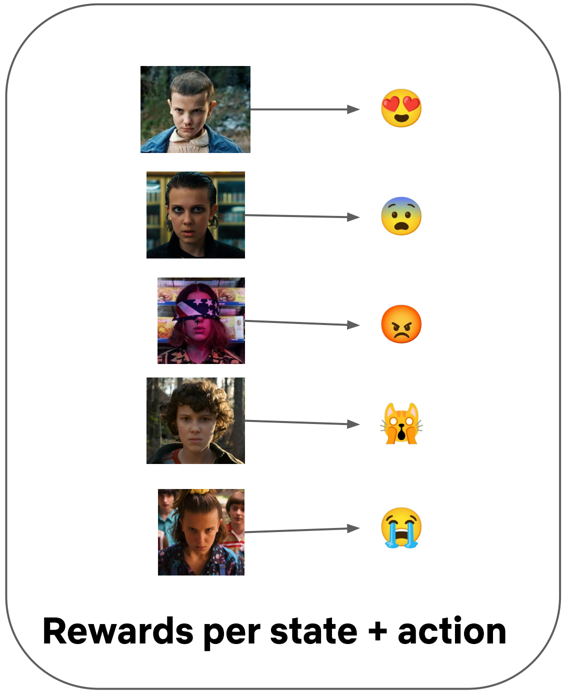
Estimation (borrow double ML, doubly-robust, asymptotically efficient estimator from Kallus & Uehara (2022))

Simulation Details + Results (and code!)

Environment is a Markov Decision Process



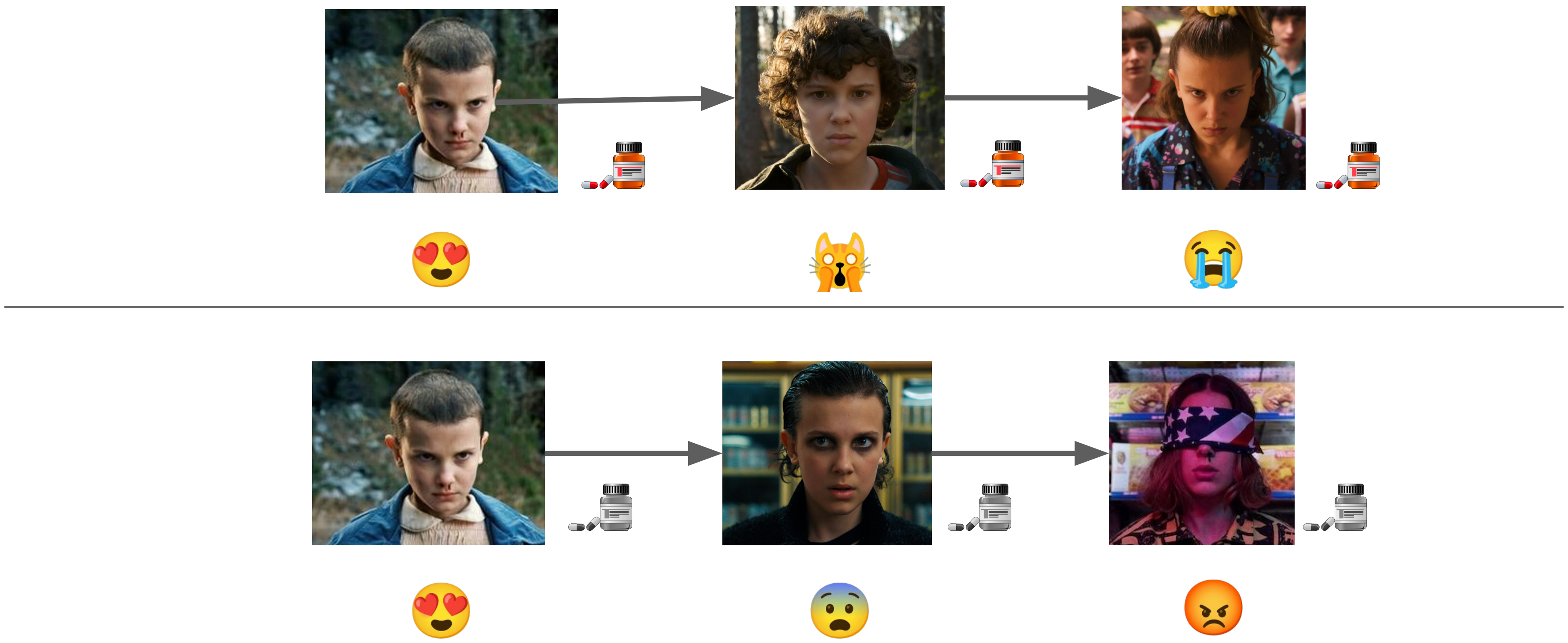




$$Y(\pi^T) \equiv (1-\gamma) \sum_{t=0}^{\infty} \gamma^t Y_t(\mathbb{1}_{t<T})$$

Long-term cumulative "potential outcomes"

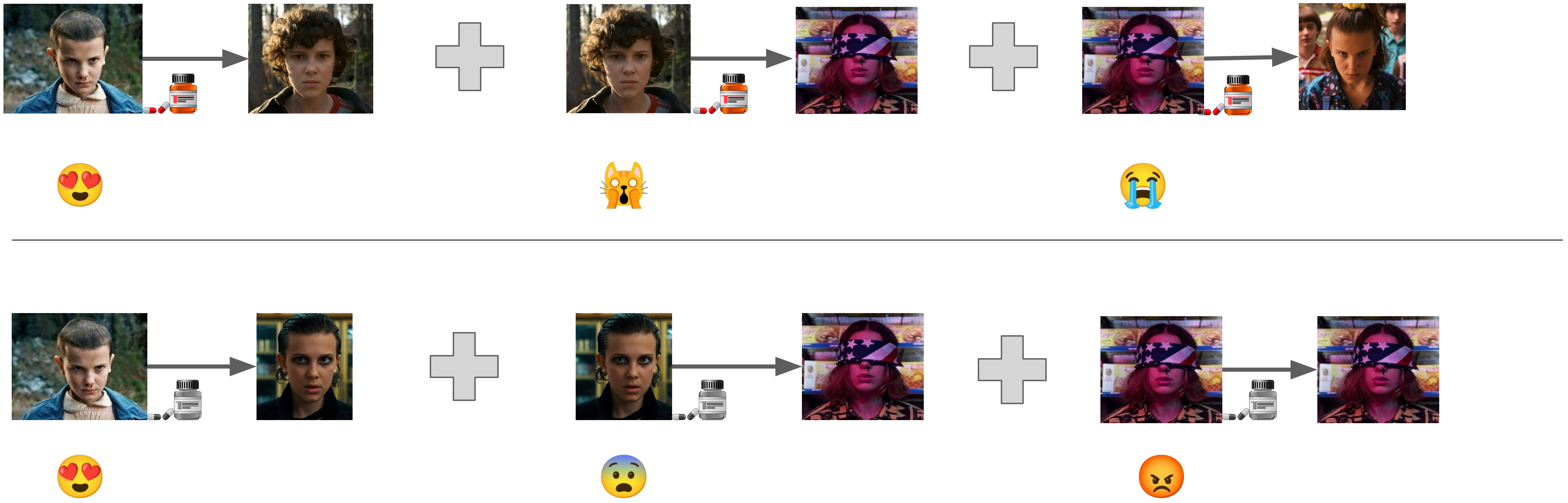
Ideally, run a long-term RCT



Long-term ATE = 😍 + γ * 🙀 + γ^2 * 😭 - (😍 + γ * 😨 + γ^2 * 😠)

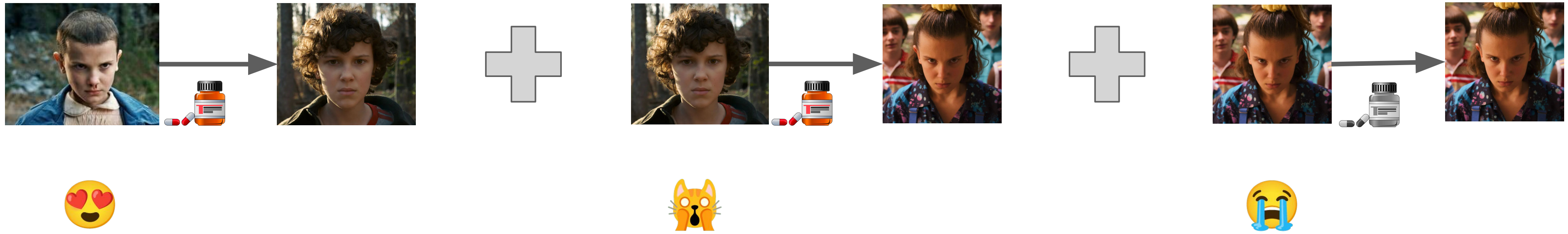
Idea: run a short-term test on “everyone”

Mimic a long-term RCT from short-term RCT



This is the difference between the cumulative rewards from two Markov chains

Inference of *any*-duration treatment regimes



e.g Two periods of treatment instead of 3 (T_0 = 1, T_1 = 1, T_2 = 0)

Generalize to any-duration treatment regime from a single experiment!

“Mixing” actions requires going beyond Markov chains -> Q functions and policies

Beyond intuition: identification with tools from reinforcement learning

Basic idea is to fit separate Markov chains on treatment and control is impractical

- many states are required for the Markov property to hold -> curse of dimensionality
- states are continuous -> can't form a transition matrix
- desire to evaluate different duration treatments

Solution: use ML-based function approximators of the Q function

Requires: asymptotic efficient estimators from offline reinforcement learning

Estimand is the difference in long-term potential outcomes

Estimand is difference in long-term potential outcomes

$$\varphi^T = \mathbb{E}\left[Y(\pi^T) - Y(\pi^0)\right] \qquad (2.1)$$

where π^T is the policy of treatment for T periods then nothing thereafter

Assumption 1 (Additive rewards).

$$Y(\pi^T) \equiv (1-\gamma)\sum_{t=0}^{\infty}\gamma^t Y_t(\mathbb{1}_{t<T}) \qquad (2.2)$$

Identification: long-term ATE = difference in Q functions

Lemma 1 (Stationary T-Duration Treatments). *For a non-stationary policy π^T that sets $a = 1$ for T periods and $a = 0$ thereafter, (i) there exists an equivalent stationary stochastic policy $\bar{\pi}^T$ that yields the same cumulative discounted reward and (ii) the average of that stationary stochastic policy across states is $1 - \gamma^T$.*

We use that equivalent stationary policy to define a stationary Q function.

$$q^T(s, a) \equiv \mathbb{E}_y [y|s, a] + \gamma \mathbb{E}_{s' \sim p(\cdot|s,a), a' \sim \bar{\pi}^T(\cdot|s')} \left[q^T(s', a') \right]$$

Theorem 1 (Identification by Stationary-policy Q). *Suppose Assumptions 1-5 hold. Then the expected average treatment effect of a T-duration treatment policy is equal to expectation over the difference of Q functions, associated with the equivalent stationary policy, $\bar{\pi}^T$ and the control policy.*

$$\varphi^T = (1 - \gamma) \mathbb{E}_{s \sim p_0(\cdot), a \sim \bar{\pi}^T(\cdot|s)} \left[q^T(s, a) - q^0(s, 0) \right]$$

Preview of Experiments

Simulate a long-run experiment where treatment is applied for T periods

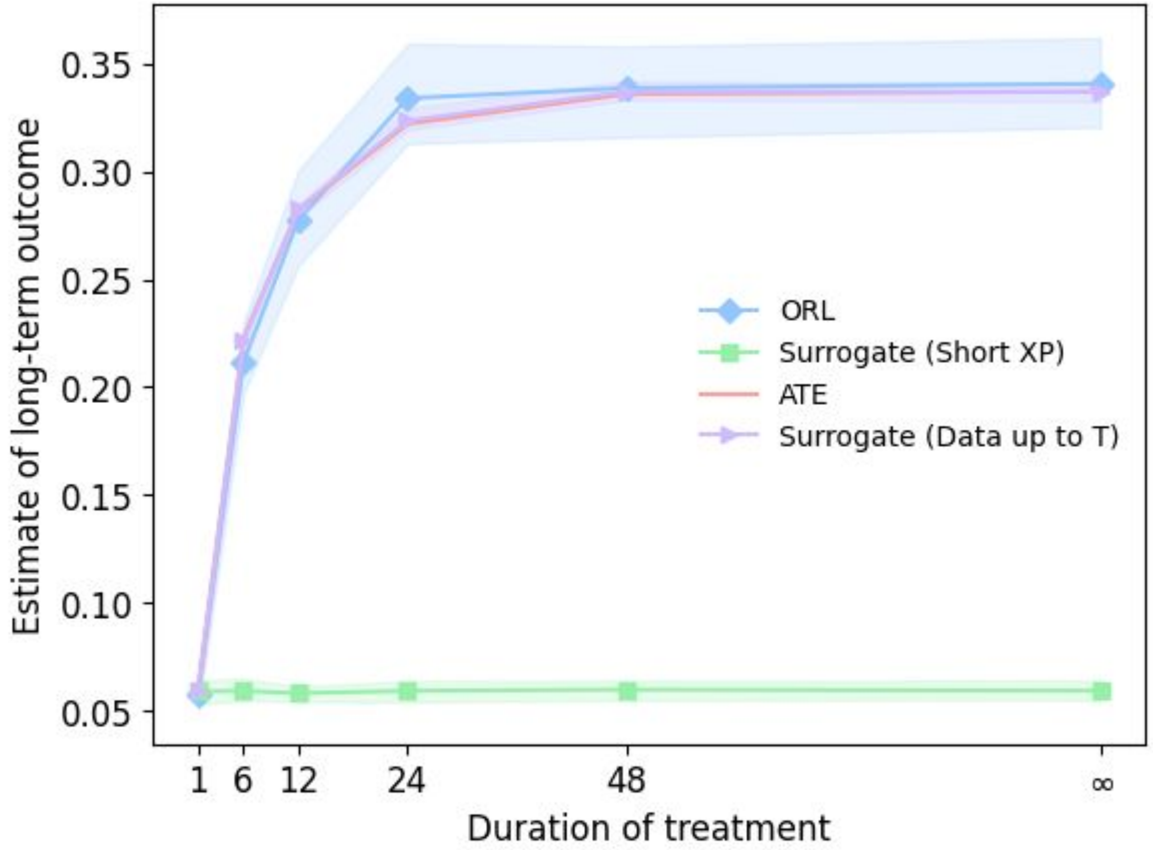
Calculate the true **long-term ATE** always over ∞ horizon (red)

Short experiment = experiment runs for 2 periods

- surrogate method (green)
  - also gets observational dataset under control
  - under short-term experiment, only $T=1$ estimate is correct
  - $T > 1$ unbiased only if you experiment runs T periods

- our method (blue) matches the ATE for all T from short experiment

Experimental Results - Toy MDP



1 continuous state Markov Chain (drift diffusion process)

Treatment affects (i) state transition and (ii) reward mapping

Experimental Results - Sepsis Simulator