

How to “Improve” Prediction Using Behavior Modification



Galit Shmueli

Institute of Service Science



國立清華大學
NATIONAL TSING HUA UNIVERSITY

**Joint work with
Ali Tafti @ UIC**

Editor-in-Chief, *INFORMS Journal on Data Science*

Keynote @ The 23rd *International Conference on Electronic Business* (ICEB), Taiwan, Oct 21, 2023



How to “improve” prediction using behavior modification ☆

Galit Shmueli ^a , Ali Tafti ^b

Discussion

Ethics and interventions: A commentary on how to “improve” prediction using behavior modification

International Journal of Forecasting, Available online 18 November 2022, ...

Foster Provost

Discussion

Forecasting, causality and feedback

International Journal of Forecasting, Available online 26 October 2022, ...

Rob J. Hyndman

Discussion

How to improve prediction using behavior modification?

International Journal of Forecasting, Available online 17 October 2022, ...

Bart Baesens

Discussion

Rejoinder: How to “improve” prediction using behavior modification

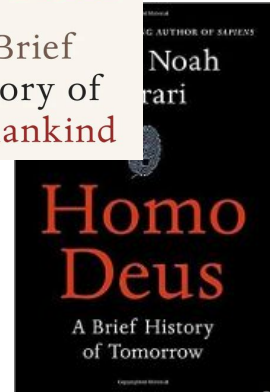
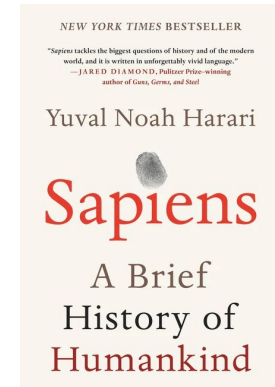
International Journal of Forecasting, Available online 2 January 2023, ...

Galit Shmueli, Ali Tafti

Yuval Noah Harari reveals the real dangers ahead

The TED Interview

“The ability to hack human-beings means the ability to understand humans better than they understand themselves. Which means being able to **predict their choices...** to **manipulate their emotions,** to **make decisions** for them.”



Today's Talk

Background

1. Digital platforms **predict user behavior**
2. Digital platforms **modify user behavior**

This work: how behavior modification can “improve” prediction

- Scenarios
- New statistical notation
- Analyzing the Expected Prediction Error
- Alarming discoveries

1.

Digital platforms

predict

user behavior

Behavioral Big Data (BBD)

Big Data

- ❖ many records
- ❖ multiple sources
- ❖ rich, detailed info
- ❖ micro-level



Behavioral Data

people's measurable
actions, reactions,
interactions,
self-reported opinions,
thoughts, feelings,...

Traditional companies collect Behavioral Big Data

and use it to **predict** our future behavior



Finance



Retail



Insurance

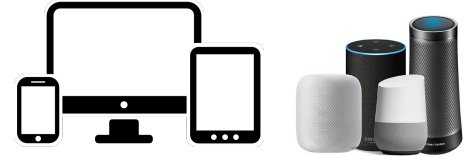


Why?

- personalize
- improve
- product/service
- detect fraud
- advertising
- marketing
- logistics/operations

Digital platforms collect Behavioral Big Data

and use it to **predict** our future behavior



Why?

personalize
improve product/service
detect fraud
advertising
marketing
logistics/operations


Digital platforms **predict** different user behaviors:

- purchase probability
- churn probability
- time-on-site / time-on-app
- emotional state
- voting
- life event


Reach predictive audiences in Google Ads

Suggested audiences

Additional audience suggestions for you to consider

GENERAL
TEMPLATES
 PREDICTIVE NEW



Analytics builds predictive audiences based on behavior like buying or churning. [Learn more](#)




Likely 7-day purchasers

Target users who are likely to make a purchase in the next 7 days

ELIGIBILITY STATUS




Not eligible to use





Likely first-time 7-day purchasers

Target users who are likely to make their first purchase in the next 7 days

ELIGIBILITY STATUS




Ready to use





Likely 7-day churning purchasers

Target purchasing users who are likely to not visit your property in the next 7 days

ELIGIBILITY STATUS




Ready to use




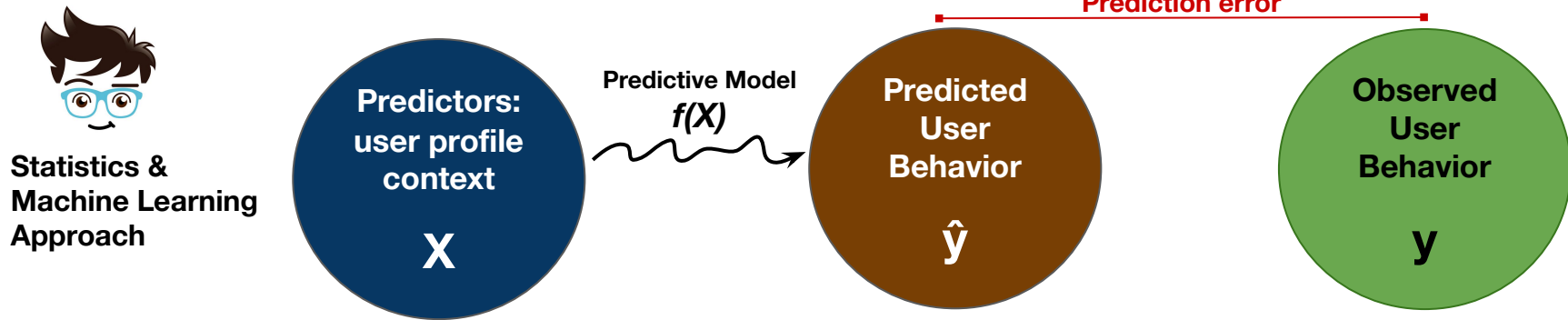
Likely 7-day churning users

Target active users who are likely to not visit your property in the next 7 days

ELIGIBILITY STATUS


Ready to use


Prediction Efforts (“Offline Learning”)



- Improve $f(x)$
- Better data (X, y)
- Better computing infrastructure

What companies do:

- Hire data scientists
- Purchase & collect better data
- Develop infrastructure

2.

Digital platforms
manipulate
user behavior

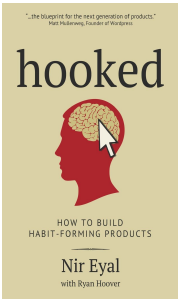
Internet platforms use
behavior modification techniques & BBD
to **change** users' behaviors:

- purchase probability
- churn probability
- time-on-site / time-on-app
- emotional state
- voting
- life event



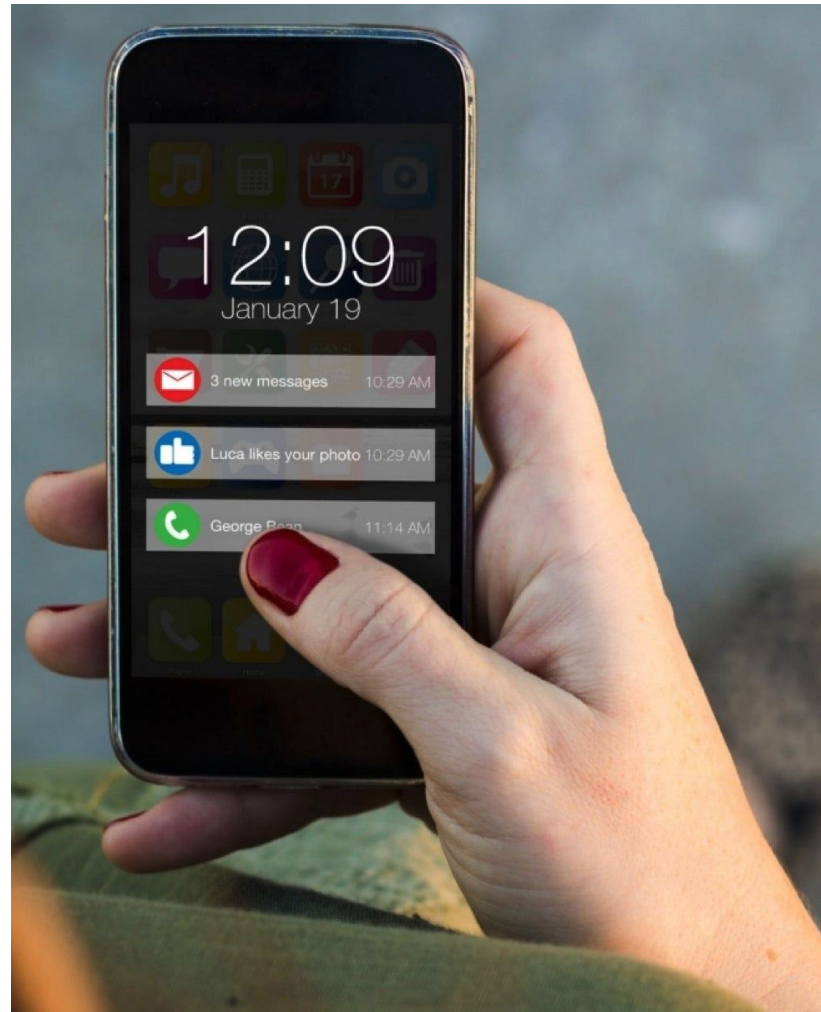
Behavior modification

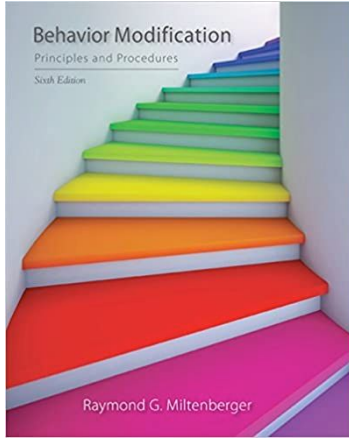
“intervention designed to enhance certainty by altering causal processes that regulate behavior”
= **Modify behavior in a predictable way**



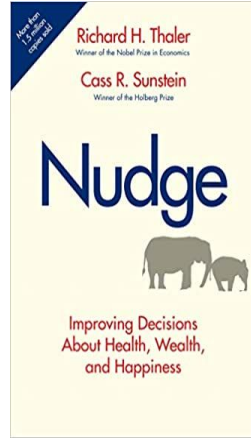
“Companies must learn not only what compels users to click, but also what makes them tick.”

Habits are defined as behaviors done with little or no conscious thought.”

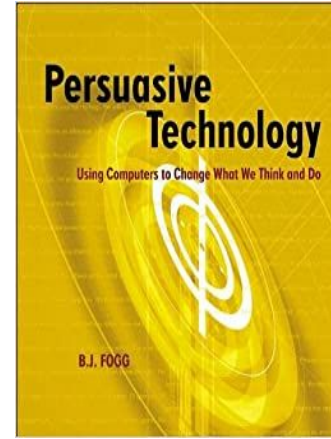




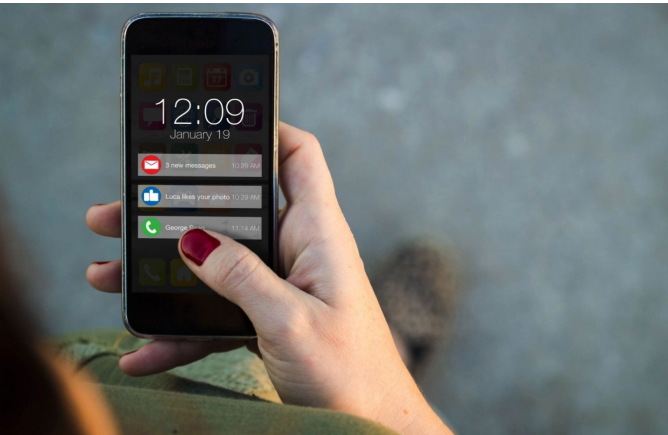
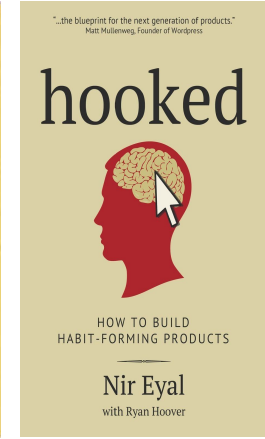
Psychology



Behav. Economics



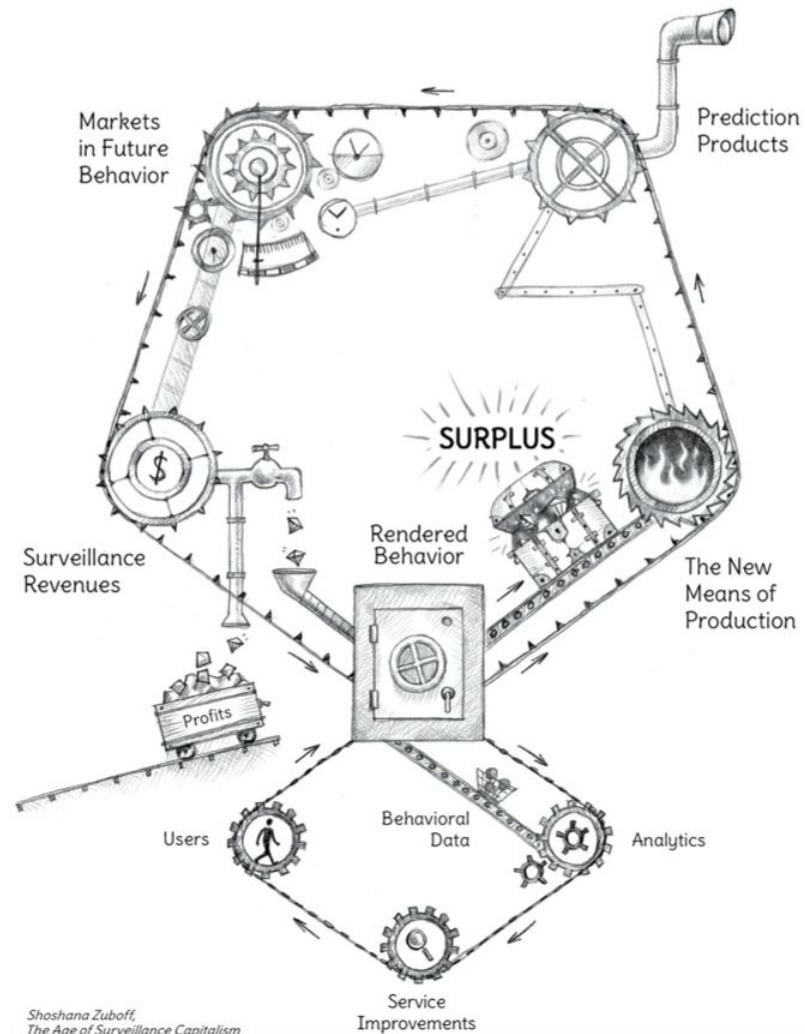
Human-Computer Interaction



Behaviorist Techniques (adapted to digital)

- Tuning, nudging - subliminal cues in context
- Herding, tunneling - manipulate/engineer context
- Conditioning - reinforcement schedules

**How behavior
modification can
“improve” prediction**



Better predictions → more valuable **prediction product**

“Traditional” statistical approach

- Improve $f(X)$
- Better data (X, y)
- Better computing infrastructure

Hire data scientists

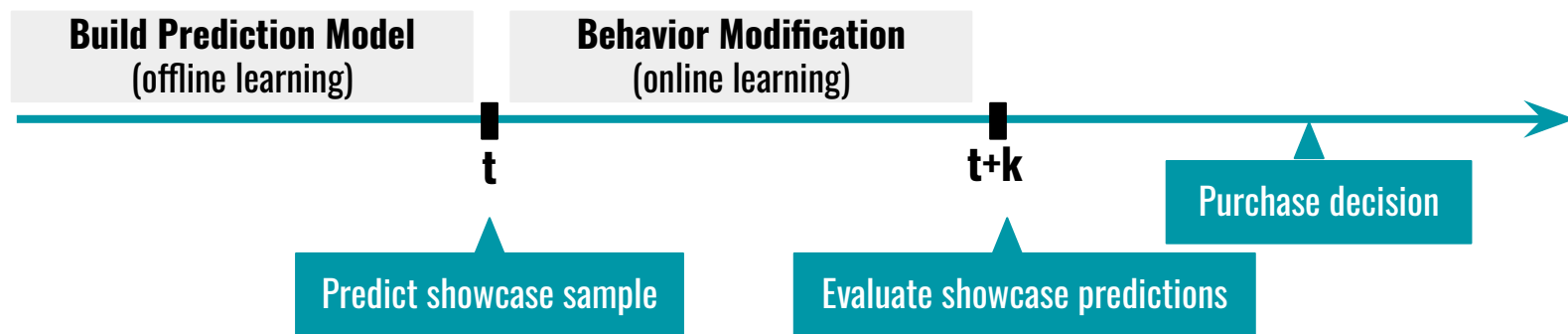
Purchase & collect better data



Internet platform's potential new approach

\$ Push users' behaviors
towards their predictions

‘improve’ prediction using predict-then-modify strategy



Predict: At time t , predict a user's future behavior at time $t+k$ ($k>0$) using model learnt offline: $\hat{y}_{t|t+k}$

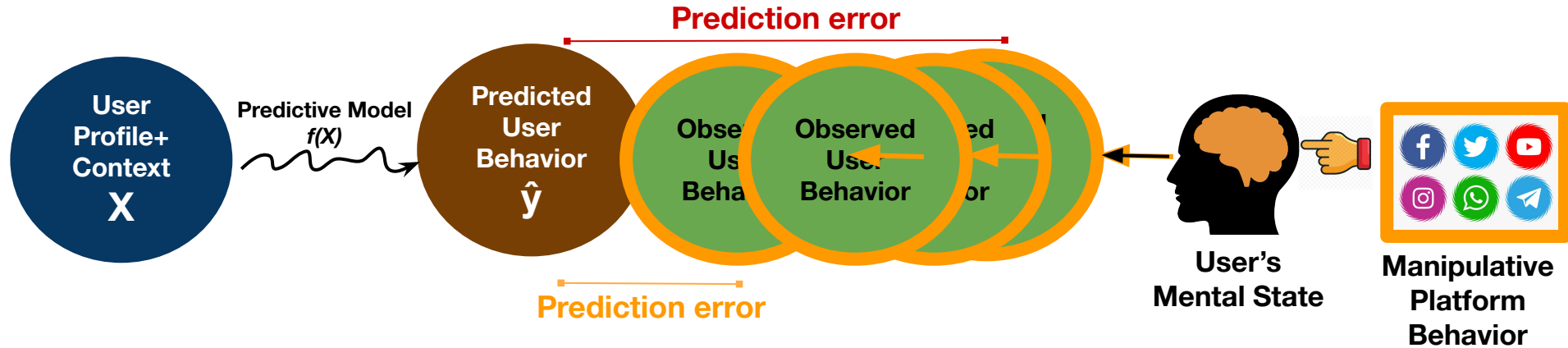
Modify: During $(t, t+k)$ push user's behavior towards its predicted value $\hat{y}_{t|t+k}$ using continual behavior monitoring (online learning)

How to “improve” prediction

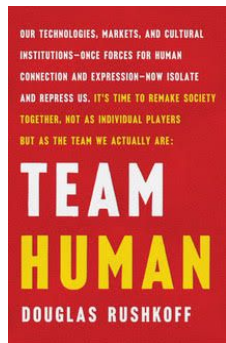


Statistical / Machine Learning
Approach

Behavior Modification
Approach

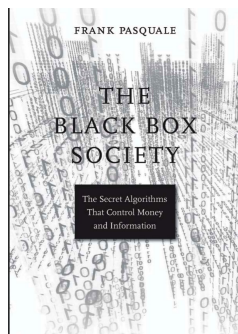


Why would a platform do this?



Media Theorist Douglas Rushkoff:

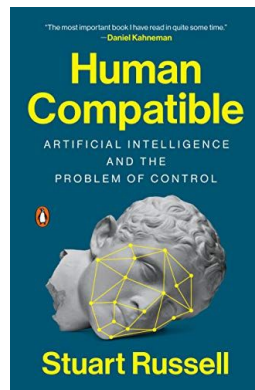
The better [the platform] does at making Mary conform to her algorithmically determined destiny, the more **the platform** can boast both its **predictive accuracy** and its **ability to induce behavior change**.



Law Prof. Frank Pasquale (UMD):

Facebook's behavioral prediction work is "eerie" [he worries] how the company could turn algorithmic predictions into "self-fulfilling prophecies," since "**once they've made this prediction**, they have a **financial interest in making it true**.

theintercept.com/2018/04/13/facebook-advertising-data-artificial-intelligence-ai/



AI Expert Stuart Russell (CS @Berkeley):

Content selection algorithms on social media... are designed to maximize click-through. The solution is simply to present items that the user likes to click on, right? Wrong. The solution is to **change the user's preferences** so that they become more **predictable**.

Bottom Line

- ✓ Platform has financial incentive & technical capability

Example:
Food Delivery
Drivers' Risk Scores



Imaginary Scenario



- Insurance wants to avoid high risk customers
- Internet platform wants to **sell predicted risk scores**
- Insurance company wants to buy scores on ongoing basis, but wants to know expected prediction accuracy
- Platform data science team sets *app use rate while driving* as behavior of interest. Agree on 7-day ahead prediction
- Platform can modify app use rate via **driver engagement** (notifications volume, type, delivery mode)

Imaginary Scenario - cont.

Platform showcases its predictive power:

1. **Train:** get pre-trained prediction model $\hat{y}_{t+7} = f(x_t)$
2. **Predict:** Select showcase sample of drivers and generate predictions $\hat{y}_{i,t+7}$ ($i=1,\dots,n$)
3. **Modify:** during $(t, t+7)$ use info \hat{y}_{t+7}, x_t to push drivers' behaviors towards \hat{y}_{t+7}
4. **Evaluate:** at time $t+7$, compare app use rates of showcase sample to the predictions (e.g. MSE)



Result:
Platform turns
high-risk scores
into high-risk
realities

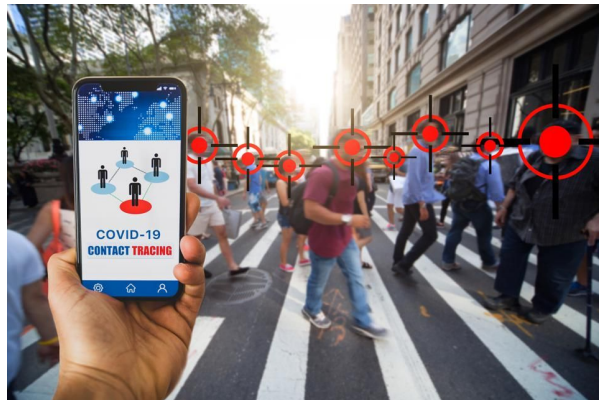
Imaginary Scenario #2

1. Political consulting firm wants to reach out to people likely to vote for candidate A
2. Internet platform wants to sell **predicted *vote for A* scores** of their users
3. Platform showcases its predictive power:
 - generates ***vote for A*** user scores
 - pushes users' behaviors towards their scores



Platform turns voting predictions into voting realities

Imaginary Scenario #3

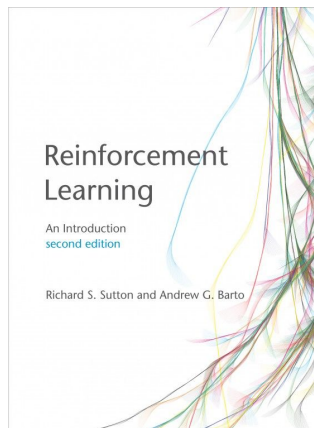


1. Government wants to identify citizens with high risk of becoming severely ill with covid-19 (e.g., for vaccination priority)
2. Platform wants to sell *covid-19 susceptibility scores*
3. Platform showcases its predictive power:
 - generates user risk scores
 - pushes users' behaviors towards their scores (encourage/discourage crowded locations)

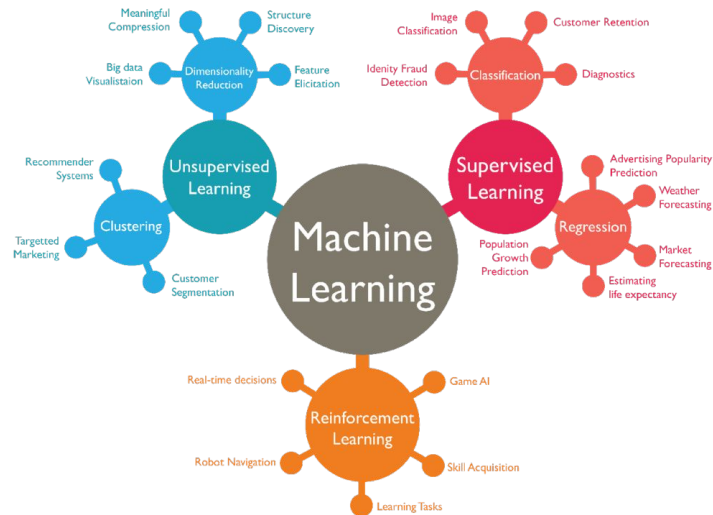
Platform turns high-risk scores into high-risk realities

How can a platform modify behavior so well?

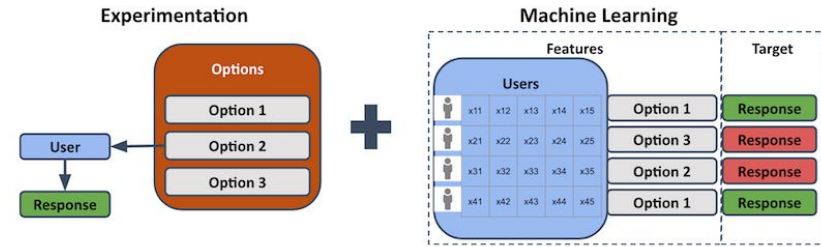
Platforms now use **Reinforcement Learning** that applies sequential, personalized, online **interventions**, to achieve a prespecified objective function via delayed reward



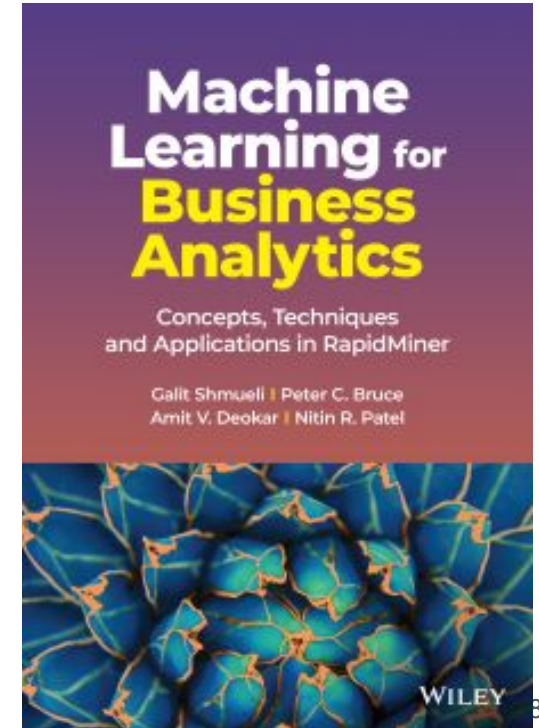
Reinforcement learning (RL) is a general-purpose family of methods for addressing sequential decision making problems characterized by non-deterministic dynamics, delayed decision-outcome pairings, and a lack of ground truth regarding optimal decisions.



How would a platform do this?



- RL uses computational algorithms that learn by interacting with users, via *dynamic treatment assignment*
- **Like supervised learning**, goal is to learn predictive model $y=f(X)$
- **Unlike supervised learning** that uses pre-existing data (offline learning), **RL starts without data (online learning)**; assigns treatment X to get most useful y for learning $y=f(X)$ -- prespecify *objective function*
- **Unlike A/B testing**, RL **treatment space can be huge**

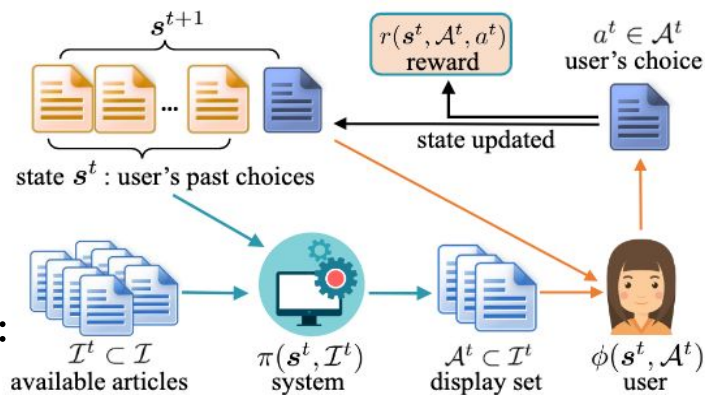


RL is useful for online personalization

Multiple contexts (location, time, device...)

Multiple **behavior modification** dimensions, each with many options (multiple factors with many levels):

- Content (ad, news item, friend,...)
- Design (website morphing)
- Reinforcement type (reward schedule & type: time pressure, social pressure)



N

/the social dilemma



*RL can “**improve**” prediction accuracy of a prediction product*

- Product team explains importance of **prediction accuracy** to data science team
- Data science team sets RL objective function to **minimize prediction errors**, where predictions are based on prior offline learning

Can happen **maliciously** or **myopically**



If a platform uses this “improve” prediction strategy...

Questions arise:

1. Can behavior modification **mask poor predictions**?
2. Can customers **infer** the real predictive performance?
3. Can customer's routine A/B testing **detect** this scheme?
4. What is the role of **personalization**?

To answer these Qs:

1. Represent new scenario of prediction + manipulation
2. Derive the *Expected Prediction Error* under manipulation

$$EPE = E[(\textit{manipulated } y - \hat{y})^2]$$

The challenge: technical vocabulary

Notation for predictive relationships

≠

Notation for causal relationships



Econometric approach: structural equations model?

Equation 1: $y|x = f(x) + \varepsilon$

Estimate/train model and generate predictions \hat{y}

Equation 2: ???

“Push” outcome y towards \hat{y}

Goal is not estimating a causal effect

Need notation for predictive+manipulation operations

Predictive notation (no manipulation)

$$y|x = f(x) + \varepsilon$$

$$E[\varepsilon]=0, \text{Var}(\varepsilon)=\sigma^2$$

$$\hat{y}|x = \hat{f}(x)$$

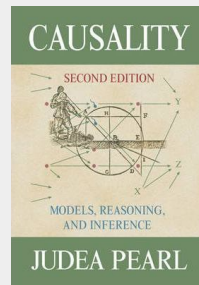
Manipulation notation (given prediction)

$$\tilde{y} \stackrel{\text{def}}{=} g(y| \text{do}(B), x) + \tilde{\varepsilon}$$

$$B = h(\hat{y}, x) + v$$

$$E[\tilde{\varepsilon}]=0, \text{Var}(\tilde{\varepsilon})=\tilde{\sigma}^2$$

Pearl's $\text{do}(\cdot)$ operator

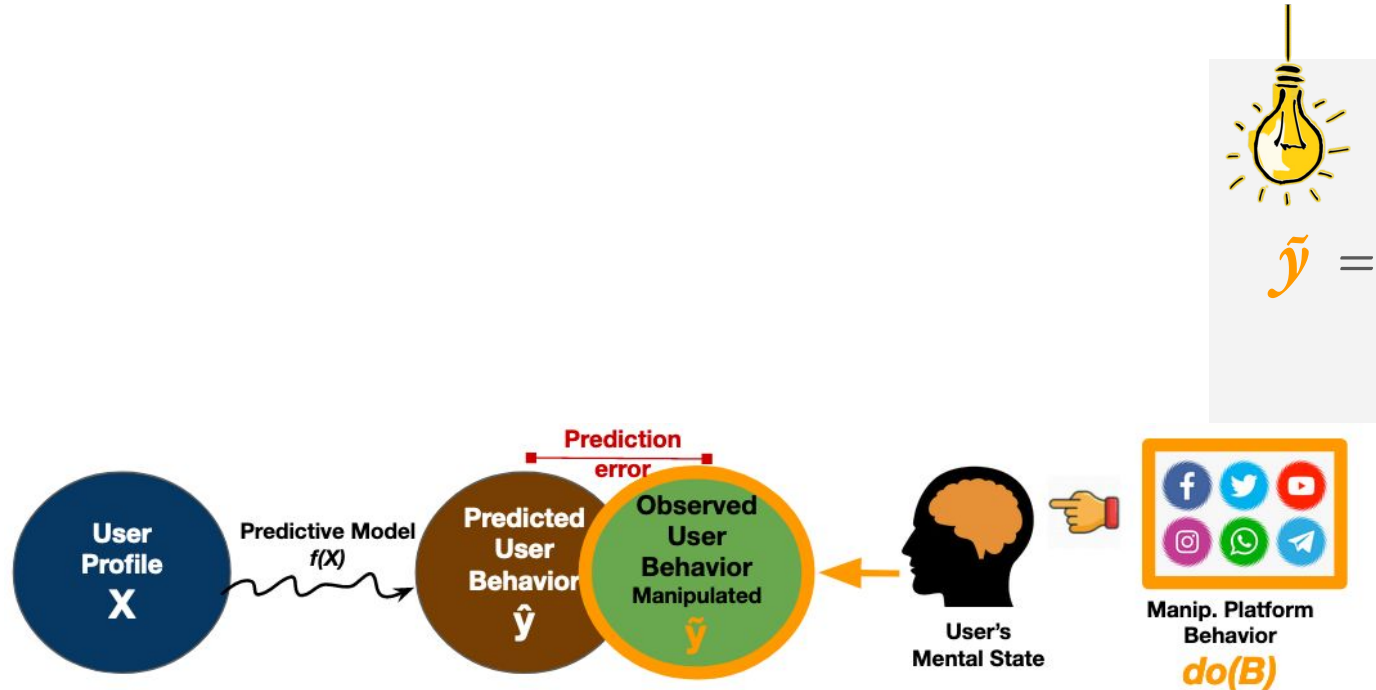


Combine!

$$\tilde{y} = g(\text{do}(B), x) + \tilde{\varepsilon}$$

$$\tilde{y} - \hat{y}$$

Need notation for predictive+manipulation operations



Combine!

$$\tilde{y} = g(do(B), x) + \tilde{\epsilon}$$

$$\tilde{y} - \hat{y}$$

No manipulation

$$\begin{aligned} EPE(X) &= E[(y - \hat{y})^2] \\ &= \sigma^2 + bias^2 + variance \end{aligned}$$

With manipulation (new derivation)

$$\begin{aligned} \widetilde{EPE}(X) &= E[(y - \hat{y})^2] \\ &= \tilde{\sigma}^2 + [CATE(X) + bias]^2 + variance \end{aligned}$$

Behavior modification can
affect error *variability*
Ideally: more homogeneous

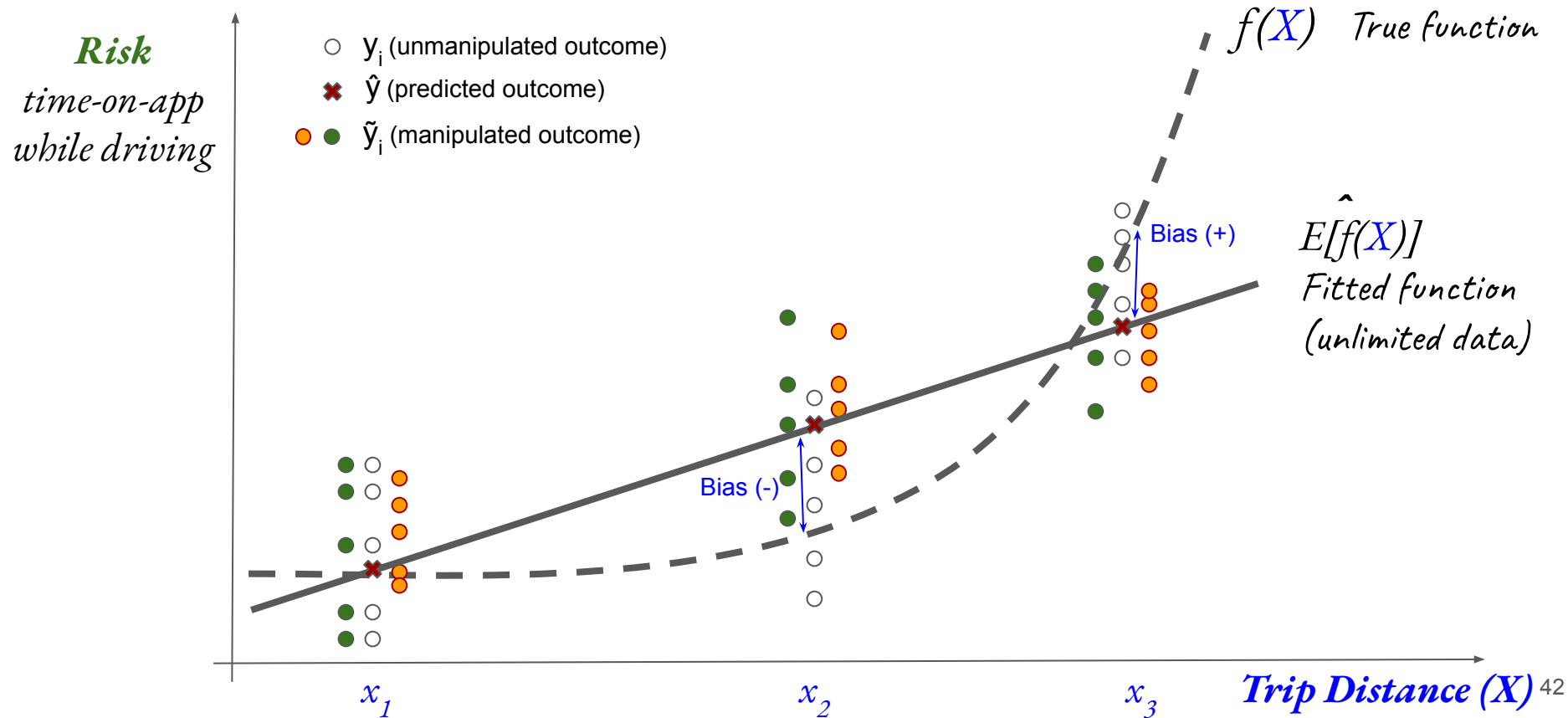
(Conditional) average behavior
modification *magnitude* (CATE) can
counter predictive algorithm's bias
Ideally: CATE = -Bias

Affected by sample
size
Ideally: large sample

Example:
Food Delivery
Drivers' Risk Scores



Effect of manipulation: Predict drivers' risk + manipulate app engagement





Alarming Discoveries

Answering the 4 Questions

Q1: Can behavior modification *mask poor predictions*?

Yes. It can counter algorithm bias and reduce error variability

$$\begin{aligned}\widetilde{EPE}[X] &= E[(\tilde{y} - \hat{y})^2] \\ &= \tilde{\sigma}^2 + [CATE(X) + bias]^2 + variance\end{aligned}$$

Q2: Can customers *infer* the real predictive performance?

Not likely. Difference is a function of $CATE$, $Bias$, $\tilde{\sigma}^2$, σ^2

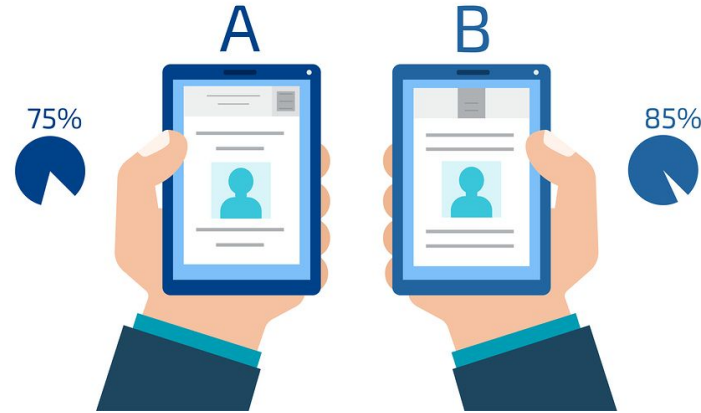
No manipulation

$$\begin{aligned} EPE(X) &= E[(y - \hat{y})^2] \\ &= \sigma^2 + bias^2 + variance \end{aligned}$$

With manipulation (new derivation)

$$\begin{aligned} \widetilde{EPE}(X) &= E[(y - \hat{y})^2] \\ &= \tilde{\sigma}^2 + [CATE(X) + bias]^2 + variance \end{aligned}$$

Q3: Can customer's *routine A/B testing* detect this scheme?



No. Random user assignment in A/B test masks the effect!
(spreads $do(B)$ equally across both groups)

Q4: What is the role of *personalization*?

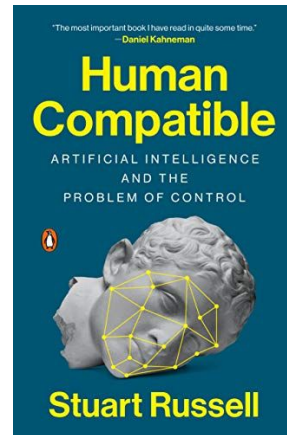
- 1) **Personalized prediction** $\hat{f}(x_i)$ -- *offline learning*

Better prediction → less modification

- 2) **Personalized behavior modification** $do(B_i)$ -- *online learning*

Better modification → better prediction product

“The solution is to **change** the user’s preferences so that they become more predictable.”



Moral Concerns

- ✓ **Behavior modification** (*behavioral change, habit formation*) originally intended for improving human life and society → persuasive technology

This predict-then-modify strategy

- ✗ Makes users conform to their algorithmically-determined destiny
- ✗ **Without users' knowledge or consent**
- ✗ Optimizes platform's commercial interest at odds with **users' well-being & agency**

A statistical approach can uncover the implications



To Conclude

- ✓ Internet platforms now combine prediction & behavior modification
- ✓ This approach is absent from statistics/ML literature
- ✓ Need technical vocabulary to analyze implications — Pearl's *do(.)*
- ✓ Vocabulary can be used to describe other *predict+manipulate* (uplift)
- ✓ Imaginary scenarios might not be imaginary
- ✓ Serious implications for platform customers and users

