

# ヒューマンコンピュータインタラクション特論 第11回

情報理工学部門 複合情報工学分野

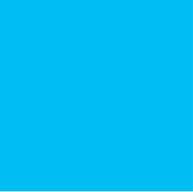
**Human-Computer Interaction (HCI) 研究室**

**小野哲雄 (8-12室)**

**[tono@ist.hokudai.ac.jp](mailto:tono@ist.hokudai.ac.jp)**

# 授業計画 (予定)

1. 10/1(火) 4限 (小野) HCI入門
2. 10/4(金) 3限 (小野) HCI概論  
10/8(火) 4限 : **[休講]**
3. 10/11(金) 3限 (小野) HCI概論(2) # 論文発表の説明
4. 10/15(火) 4限 (小野) HCIの構成原理と構成方法(1) # 登録開始
5. 10/25(金) 3限 (小野) HCIの構成原理と構成方法(2)
6. 10/31(木) 4限 <- 火曜授業 (小野) **[論文発表1(1)]**
7. 11/1(金) 3限 (小野) **[論文発表1(2)]**
8. 11/5(火) 4限 (小野) ヒューマンロボットインタラクション(HRI) # 説明
9. 11/8(金) 3限 (小野) ヒューマンエージェントインタラクション(HAI) # 登録
10. 11/12(火) 4限 (小野) HRI/HAI & Predicting Human Decision-Making
11. 11/15(金) 3限 (小野) Predicting Human Decision-Making
12. 11/19(火) 4限 (坂本) HCIにおける実験と評価(1)
13. 11/22(金) 3限 (坂本) HCIにおける実験と評価(2)
14. 11/26(火) 4限 (小野) **[論文発表2(1)]**
15. 11/29(金) 3限 (小野) **[論文発表2(2)]**



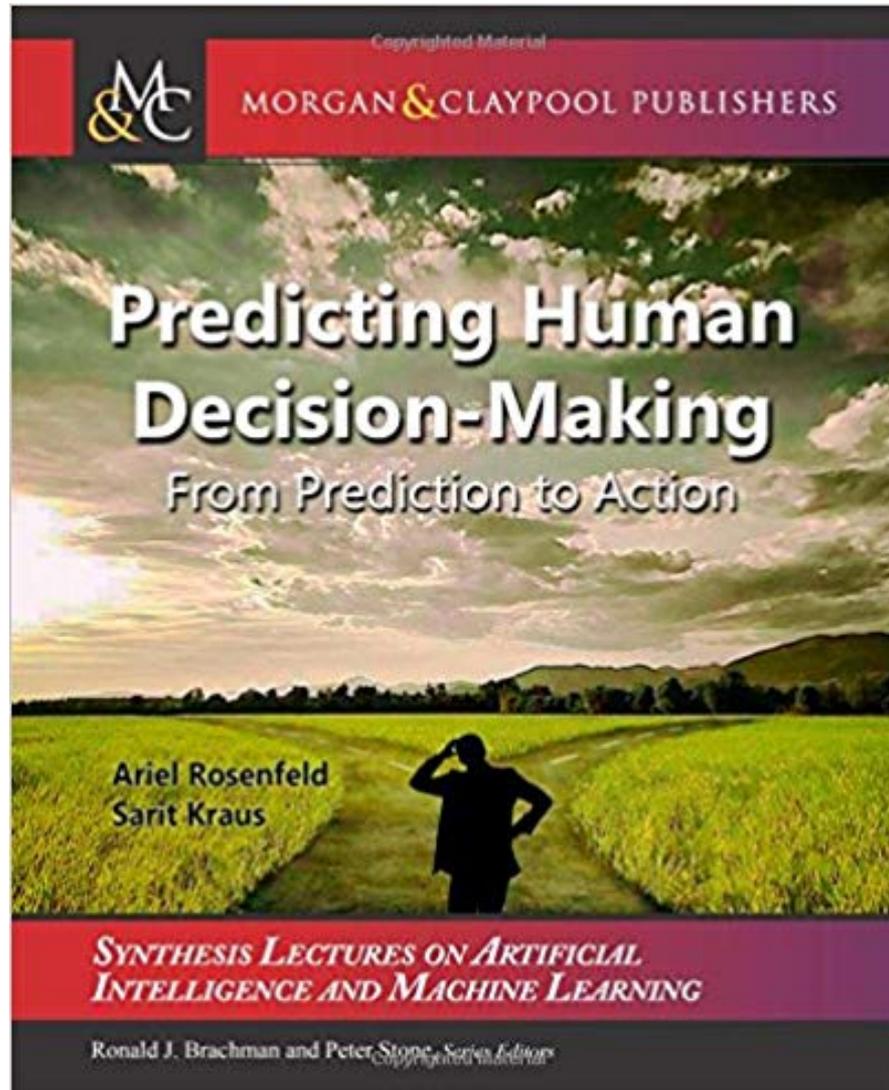
# ***Predicting Human Decision-Making***

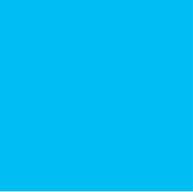
# *Outline of Today's Lecture:*

## *Predicting Human Decision-Making*

- HCI, HAI/HRI におけるPHDMの重要性
  - *Should have been: Human-Interacting Agents based on the Predicting Human Decision-Making*
- 意思決定におけるモデルとは？
  - *Expert-Driven Paradigm: Theory*
  - *Data-Driven Paradigm: Practice*
- 知覚・行動・他者の思考の推論(意識)
  - *Predictive Coding*
  - *Trust and Decision-Making*

# “Predicting Human Decision-Making: From Prediction to Action”





***Should have been:  
Human-Interacting Agents based on  
the Predicting Human Decision-Making***

# *Why do agents care about human decisions?*

- *Would you consider an autonomous car that is incapable of predicting what people (pedestrians, drivers, or its own passengers) are about to do next intelligent or even safe? Most probably not...*
- Endowing agents with Intelligence:
  - Mimicking
  - Reacting to
  - Reasoning with
  - Improving upon

*Human decision-making*

# Human Decision-Making (1)

- Complex Cognitive process
- A lot of theoretical questions
- A lot of practical implications



- Should we try and Understand or Predict?

# Understanding DM

“...much have been learned about how and why we make decisions... yet, we are still discovering the tip of the iceberg...”

- Rorie AE, Newsome WT. A general mechanism for decision-making in the human brain? *TRENDS in Cognitive Sciences*. 9(2):41-43 (2009).
- Gold JI, Shadlen MN. The neural basis of decision making. *Annual Reviews in Neuroscience*. (30):535-574 (2007).
- Kiani R, Shadlen MN. Representation of confidence associated with a decision by neurons in the parietal cortex. *Science*. 324(5928):759-764 (2009).
- J. P. Gallivan, D. A. McLean, K. F. Valyear, C. E. Pettypiece, J. C. Culham. Decoding Action Intentions from Preparatory Brain Activity in Human Parieto-Frontal Networks. *Journal of Neuroscience*, 2011;

# Predicting DM

- Easier to validate
- Observable - No intrusive investigation
- Great theoretical and practical benefit
  - To the Agents community:
    - Enhancing Human Interaction with Software\Robots...
    - Training People
    - Replacing People
    - Supporting People
    - Learning from People
    - ...



# What drives human DM?

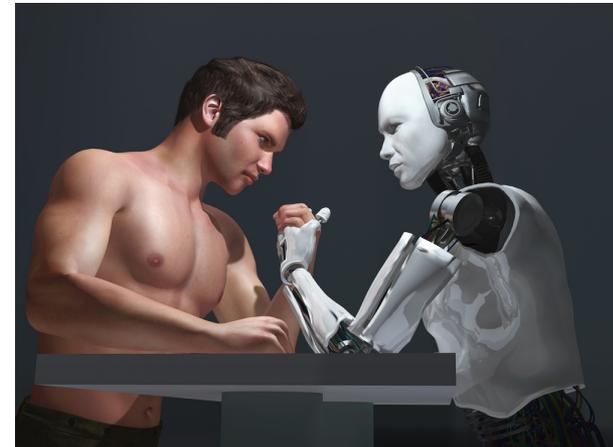
- Among others...
  - **Past experience** (Juliusson, Karlsson, & Gärling, 2005, Sagi, & Friedland, 2007)
  - **MANY Cognitive biases** (e.g., Marsh, & Hanlon, 2007; Nestler. & von Collani, 2008; Stanovich & West, 2008; West et al., 2008, Epley, & Gilovich, 2006).
  - **Individual differences such as age, cognitive abilities, gender...** (de Bruin, Parker, & Fischhoff, 2007; Finucane, Mertz, Slovic, & Schmidt, 2005, Reed, Mikels, & Simon, 2008)
  - **Decision Complexity** (Goldstein & Gigerenzer, 2002; Hilbig & Pohl, 2008)
  - **Social aspects** (Acevedo and Krueger , 2004).

# Role of the prediction agent

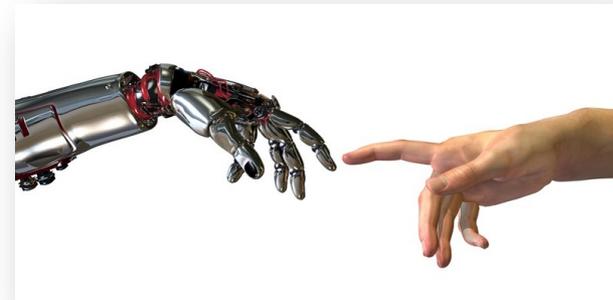
– Cooperative



– Adversarial



– Partially Conflicting



# Cooperative prediction agents

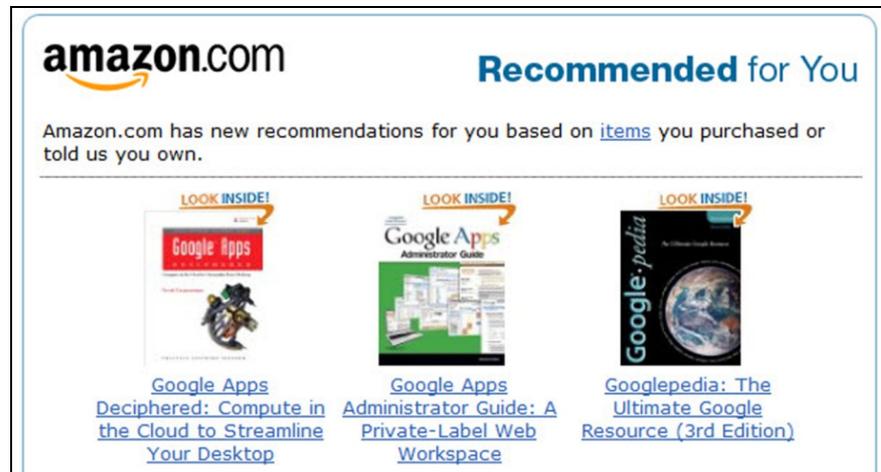
- The agent only cares about the human benefit, like “Teammates”
  - e.g., Rehabilitation, Automated Assistant
- Example: Robot Assistant
  - Needs to anticipate what the person’s next action will be and act accordingly
    - bring the needed pieces\tools...
    - Movie – Robotic Assistant (Cornell)

# Adversarial prediction agents

- The agent's success means worse outcomes for the person
  - e.g., game playing.
- Example: Poker
  - Carmel and Markovitch (1993) were the first to try to automatically learn a model of the opponent's move selection progress
  - Opponent modeling is a strong buzz-word in games
  - 2017: Carnegie Mellon University's Libratus just won a 20-day tournament against the world champs of Texas Hold'em (Noam Brown and Tuomas Sandholm)

# Partially Conflicting Interests

- The agent is interested in A, which partially conflicts with the human interest
  - e.g, negotiation
- Example: Economical Recommender Systems
  - “Buy this book, you will like it!”
  - (AND it will make us more money than the other options...)



The screenshot shows the Amazon.com website's 'Recommended for You' section. At the top left is the 'amazon.com' logo. To the right, it says 'Recommended for You'. Below this, a message states: 'Amazon.com has new recommendations for you based on [items](#) you purchased or told us you own.' The main content area displays three book covers, each with a 'LOOK INSIDE!' button. The first book is 'Google Apps Deciphered: Compute in the Cloud to Streamline Your Desktop'. The second is 'Google Apps Administrator Guide: A Private-Label Web Workspace'. The third is 'Googlepedia: The Ultimate Google Resource (3rd Edition)'.

**amazon.com** **Recommended for You**

Amazon.com has new recommendations for you based on [items](#) you purchased or told us you own.

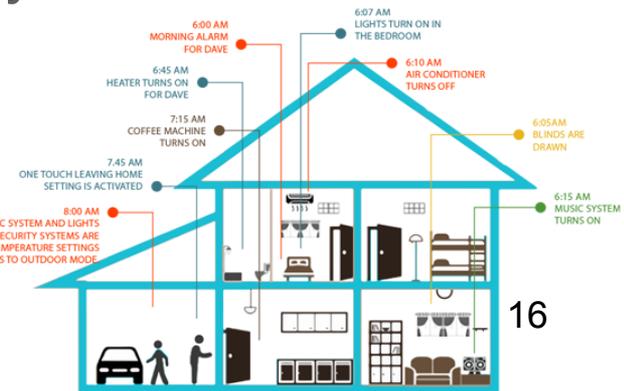
**LOOK INSIDE!**  
**Google Apps**  
**Deciphered: Compute in the Cloud to Streamline Your Desktop**

**LOOK INSIDE!**  
**Google Apps**  
**Administrator Guide: A Private-Label Web Workspace**

**LOOK INSIDE!**  
**Googlepedia**  
**The Ultimate Google Resource (3rd Edition)**

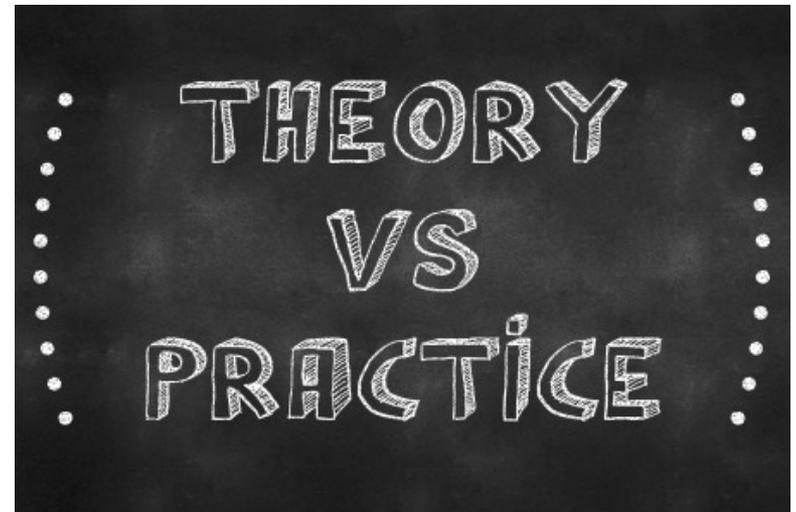
# Active Actor vs. Observer

- Sensing and recording
  - e.g., Activity loggers
  - Modeling and Understanding Human Routine Behavior, Banovic et al. (2016)
- Takes Action
  - e.g., smart home.
  - Cook, Diane J. "How smart is your home?" *Science* (2012)



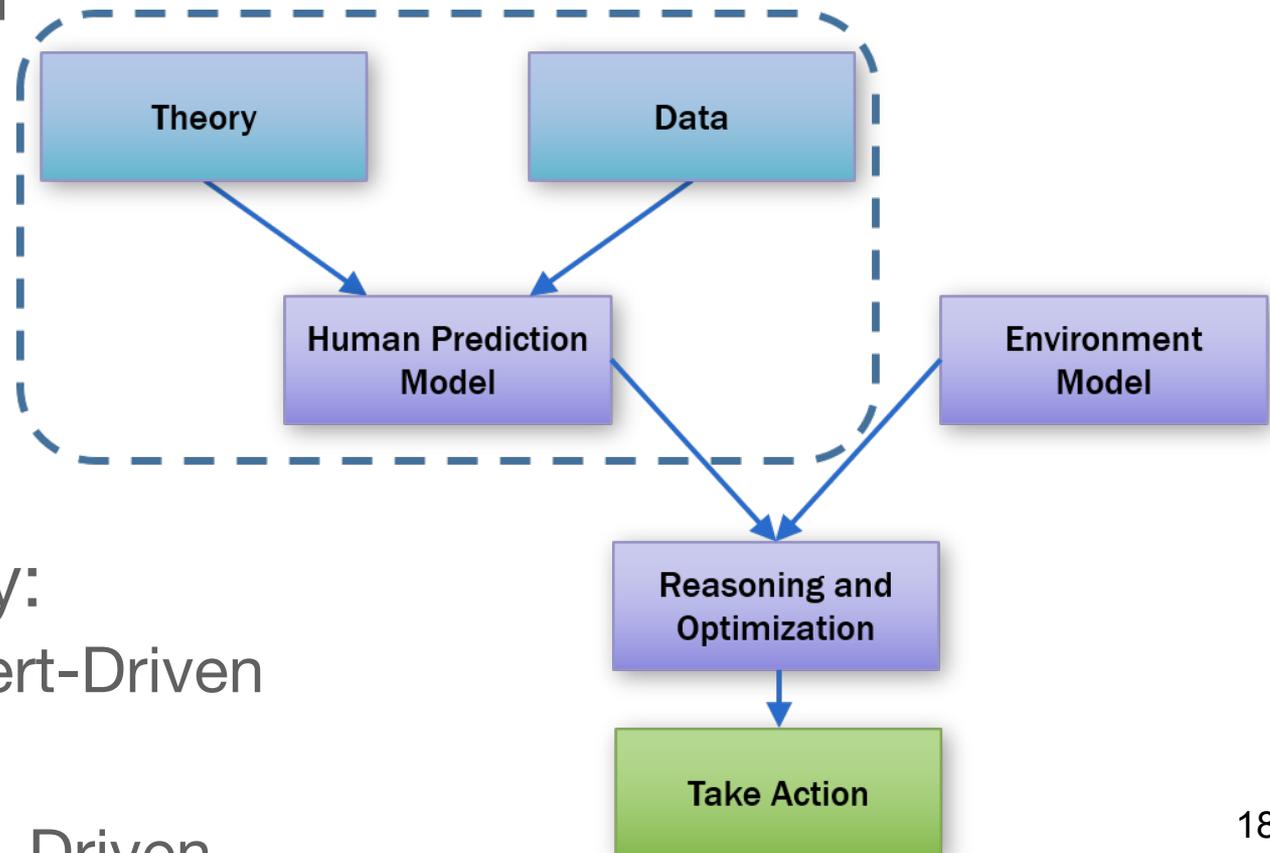
# Are people predictable?

- Expert-Driven
  - Decision-Theory
  - Game-Theory
  - etc.
- Data-Driven
  - Statistics
  - Machine learning
  - etc.

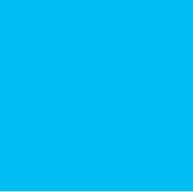


# Human Prediction Model

- High-level illustration of the integration of human prediction models in intelligent agent design



- Theory:
  - Expert-Driven
- Data:
  - Data-Driven



# *Expert-Driven Paradigm*

# Decision Theory

## (Single decision maker)

**Decision Theory = Probability theory + Utility Theory**  
*(deals with chance)      (deals with outcomes)*

### –Fundamental idea

- The MEU (Maximum expected utility) principle
- Weigh the utility of each outcome by the probability that it occurs



# 行動経済学の背景・役割

- ・ 行動経済学と伝統的な経済学の違い
  - 伝統的な経済学：「ホモ・エコノミカス」(合理的経済人)
  - 行動経済学：感情や他人に左右され、ときに間違いを犯す限定合理的な人間観 ← バイアスに支配  
(2002年、2013年、2017年、ノーベル経済学賞受賞)
  - 例：現在性バイアス、確実性バイアス、現状維持バイアス

➔ プロスペクト理論 (Prospect Theory)

# Prospect Theory (1)

## — 設問1

どうしても欲しいと思っていた商品だが、100万円と高価なため今まで購入を躊躇していた。その商品が以下のような2つのキャンペーンを実施し、どちらか一方だけを選ぶことができる。

a. キャンペーンA. 100万円が無条件で半額となり、50万円で購入できる

b. キャンペーンB. 100万円が50%の確率で全額免除となる

## — 設問2

あなたには200万円の借金がありましたが、この度消費者金融が創立100周年記念により借金減額キャンペーンを開催。選択肢は2つあり、どちらか一方だけを選ぶことができる。

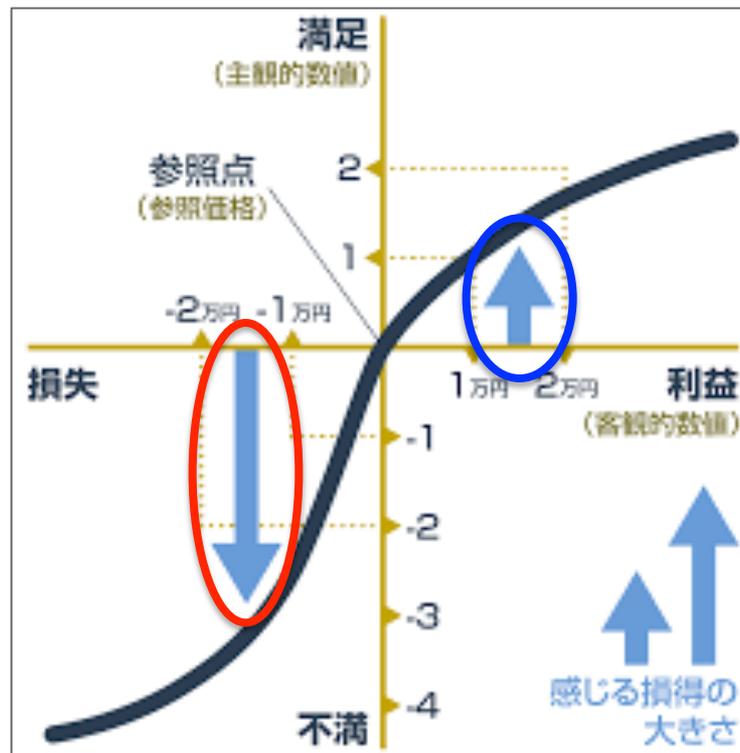
a. キャンペーンA：200万円の借金が無条件で50%減額され、返済額が100万円になる

b. キャンペーンB：200万円の借金が50%の確率で全額帳消しになる

# Prospect Theory (2)

## 特徴

- ・ 損失をそれと同じ規模の利得よりも重大に受け止める
- ・ わずかな確率であっても発生する可能性があるケースを強く意識する



# Prospect Theory (3)

## 設問1

- ある深刻な伝染病が我が国を襲ったとします。この病気は放っておけば600人の犠牲者を生みます。この病気への対処プログラムとしては2つあり、プログラムAでは200人が救われ、プログラムBを採用した場合には、3分の1の確率で600人全員が救われるが3分の2の確率で誰も救うことができません。AとBのどちらを採用しますか。なお、AとBの両方のプログラムを同時に採用することはできません。

## 設問2

- ある深刻な伝染病が我が国を襲ったとします。この病気は放っておけば600人の犠牲者を生みます。この病気への対処プログラムとしては2つあり、プログラムAでは400人が死亡し、プログラムBを採用した場合には、3分の1の確率で全員が救われるが3分の2の確率で600人全員が死亡します。AとBのどちらを採用しますか。なお、AとBの両方のプログラムを同時に採用することはできません。

## 結果

- 設問1：被験者の3分の2が「プログラム**A**」を選択
- 設問2：被験者の3分の2が「プログラム**B**」を選択

# Prospect Theory (4)

- 設問への回答が異なる理由
  - 2つの設問とも内容は同じ
    - プログラムAを実行すれば200人が生存、400人が死亡
    - プログラムBを実行すれば、確率3分の2で生存者0、確率3分の1で死者0人
    - 問題設定の方法(フレーミング) → 被験者のリスク認識と意思決定を変化

バイアス

- » 設問1：「利得の局面」(何人救えるか?)を表現  
→ 価値関数が危険回避的

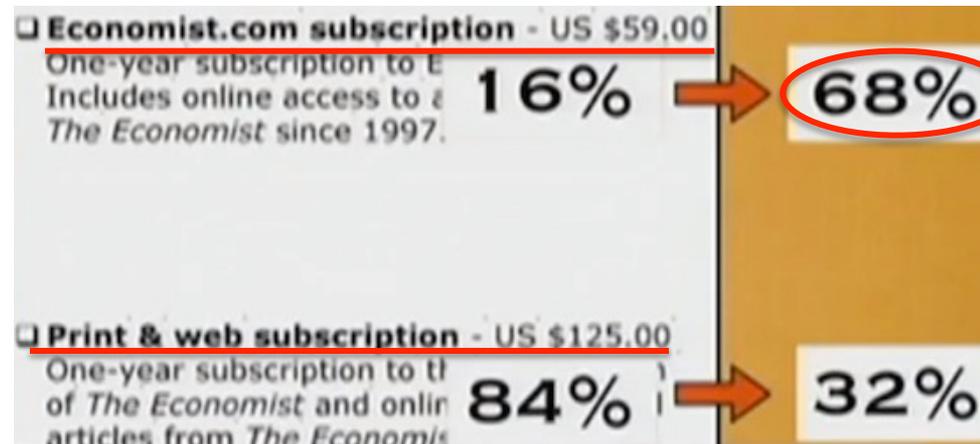
$$v(200) > \frac{2}{3}v(0) + \frac{1}{3}v(600)$$

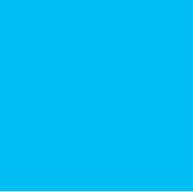
- » 設問2：「損失の局面」(何人死亡するか?)を表現  
→ 価値関数が危険志向的

$$v(-400) < \frac{2}{3}v(-600) + \frac{1}{3}v(0)$$

# バイアス：コントラスト

- Reflective Systemの検証
    - “コントラスト”のバイアス
      - 左：Print & web subscriptionを選択(84%)→右：真ん中の選択肢(0%)を削除→webのみを選択(68%)
      - 商品選択・意思決定におけるバイアス(“ナッジ”)の検証
- ➡ ユーザの健康を考慮したレストランのメニューの選択





# ***Data-Driven Paradigm***

# Setting

- Supervised
  - Labeled decision making settings
    - In setting  $x$ , person  $y$ , choose  $z$
  - Construct a model  $F(\cdot)$ 
    - Such that  $F$  approximates the real decision making outcomes
    - Generalizes to new decision making settings
- Unsupervised
  - Unlabeled decision making settings
    - In setting  $x$ , person  $y$
  - Identifying underlying structure
    - Clusters
    - Association rules
- Reinforcement Learning

# Who to Model? What to Model?



# Deep networks should answer all our prayers!? (1)

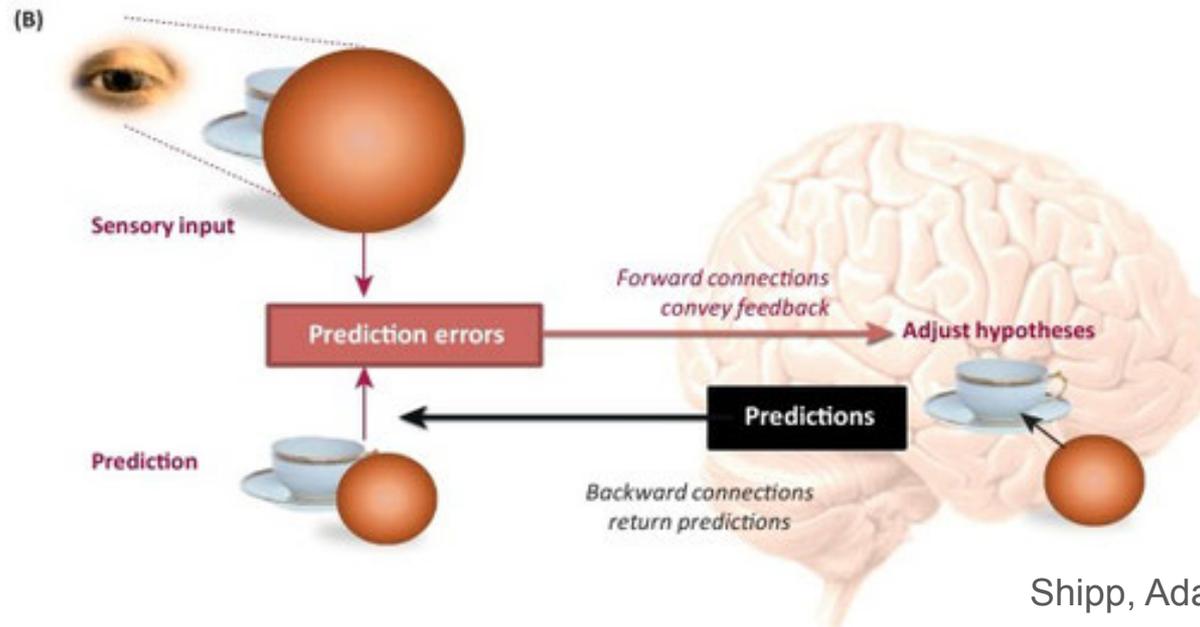
- It's (supposed to be) easy!
  - **Deep learning** has demonstrated the possibility of stunning predictive performance via learning features and digesting large amounts of data
  - Could we automatically search for decision-making models?
- However, all that glitters is not gold. At least not yet.
  - Prob. 1 - Deep Learning is a dark art
    - Not a “one-size-fits-all” solution
    - Not an “off-the-shelf” solution
    - **A LOT of tricks.**

# Deep networks should answer all our prayers!?! (2)

- Prob. 2 - Where is big data coming from?
  - Observational data:
    - » Posts to social media sites
    - » Digital pictures and videos
    - » GPS trails
    - » Transaction records
    - » cell phones
    - » Traffic
    - » Etc.
- Hard to Obtain for Many Decision Settings!
- Task and Goal Oriented Data Collection is Difficult!

# Predictive codingの概要 (1)

- Predictive coding (予測符号化)
  - 人間は外界からの感覚的データだけから知覚を形成しているわけではない
  - 知覚や運動におけるトップダウンの予測が重要な役割を果たしている ➡ 自由エネルギー原理へ



# Predictive codingの概要 (2)

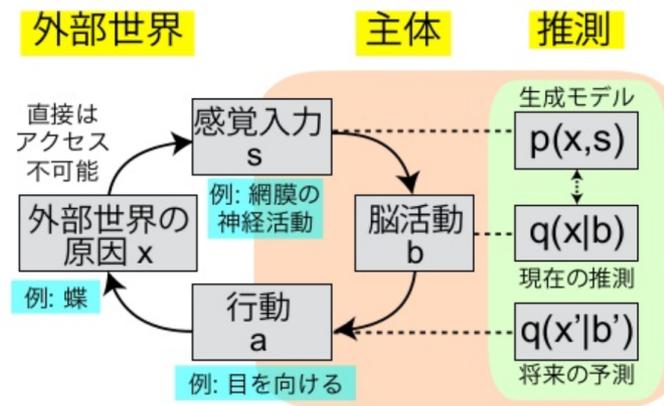
- Predictive codingの数学的基盤
  - 予測的符号化の理論では、脳における推論をベイズ統計学の原理で説明する
    - 知覚に関する内的モデルによる予測を確率分布として表現  
→ ベイズ統計学の事前分布
    - 感覚信号が確率分布として入力され、事前分布とのズレが予測誤差 → 感覚信号はベイズ統計学では観測・尤度 (likelihood)
    - 我々が主観的に経験する知覚は、この事前分布から感覚信号の入力を受けて事後分布が計算される一連の過程



**ベイズフィルタ**

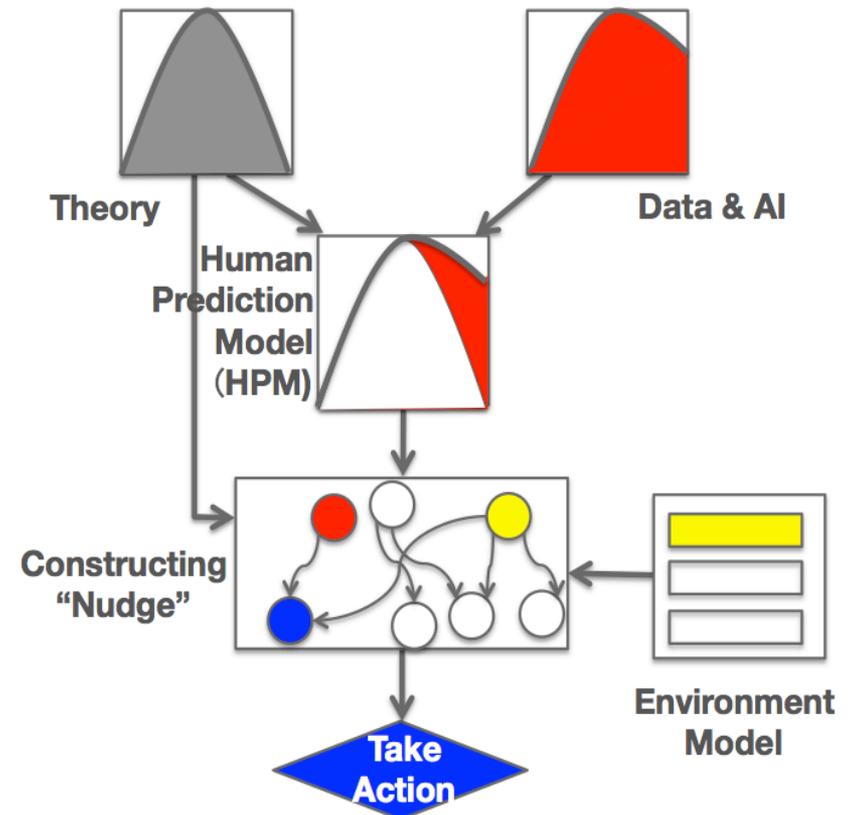
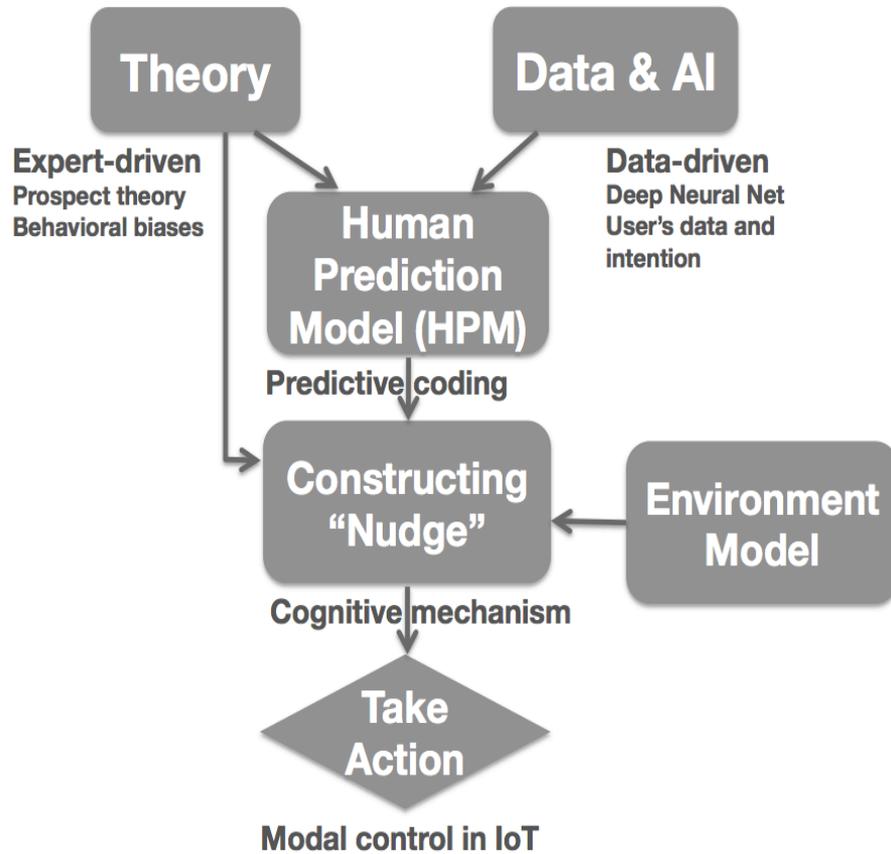
# Predictive codingの概要 (3)

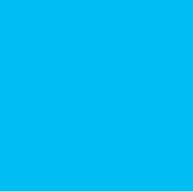
- 自由エネルギー原理 (Free-Energy Principle)
    - 外界に関する生成モデルと現在の認識から計算される変分自由エネルギーを最小化するために
      - 脳状態を変えることによって正しい認識に至る過程 (perceptual inference)
      - 行動によって感覚入力を変えることによって曖昧さの低い認識に至る過程 (active inference) の二つを組み合わせていると考える
- 自由エネルギー原理 (FEP) [Clip slide](#)



「現在の推測 $q$ 」だけでなく、  
「反実仮想的な将来の予測 $q$ 」も生成している。  
(ここでは $x'$ ,  $b'$ は将来の $x$ ,  $b$ についての表示)

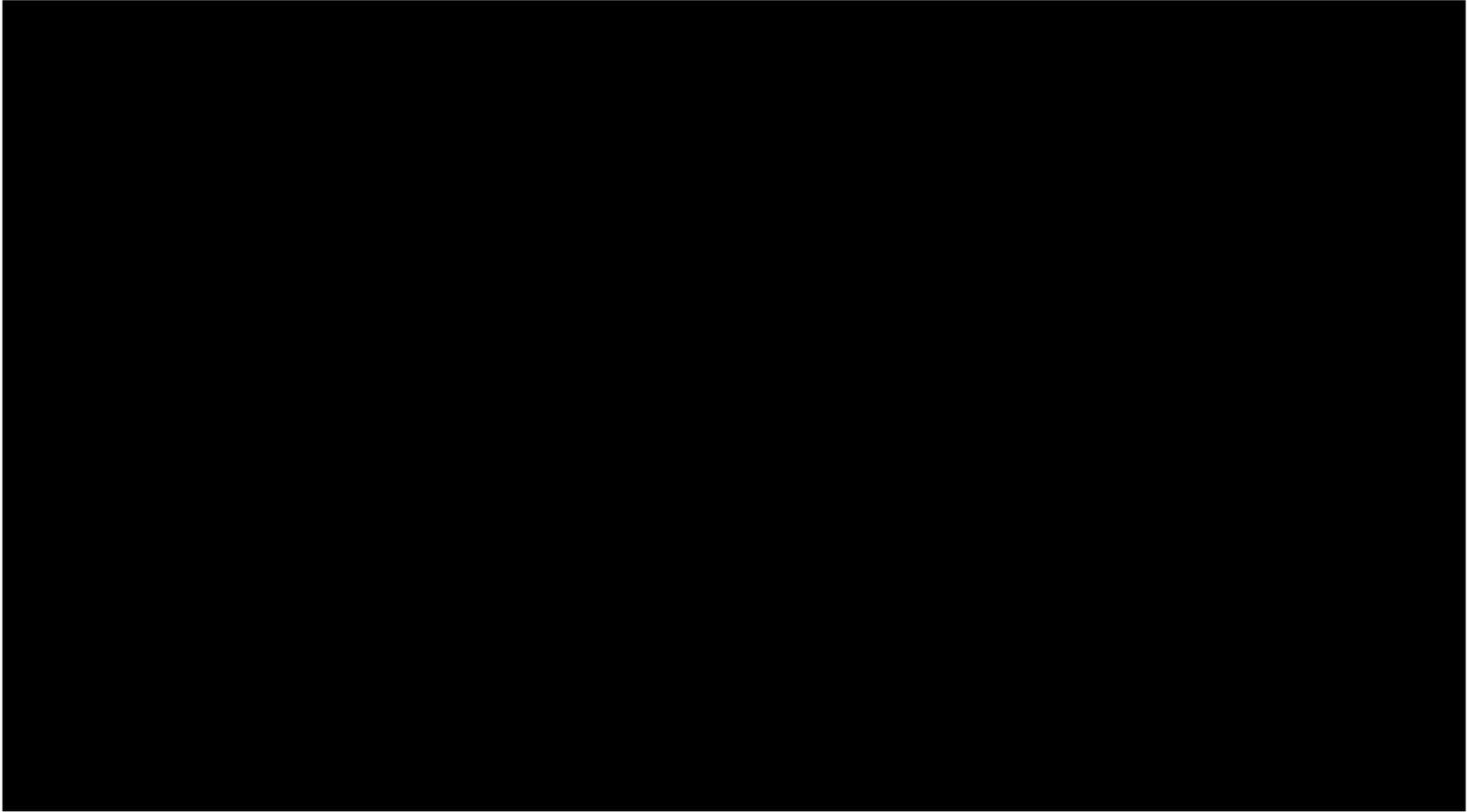
# 我々の提案するエージェントモデル





# ***Trust and Decision-Making***

# You Won't Believe What Obama Says In This Video!



*"President Trump is a total and complete dipshit"* (20秒あたり) 37  
「トランプ大統領は救いようのないマヌケだ」

# DeepFake

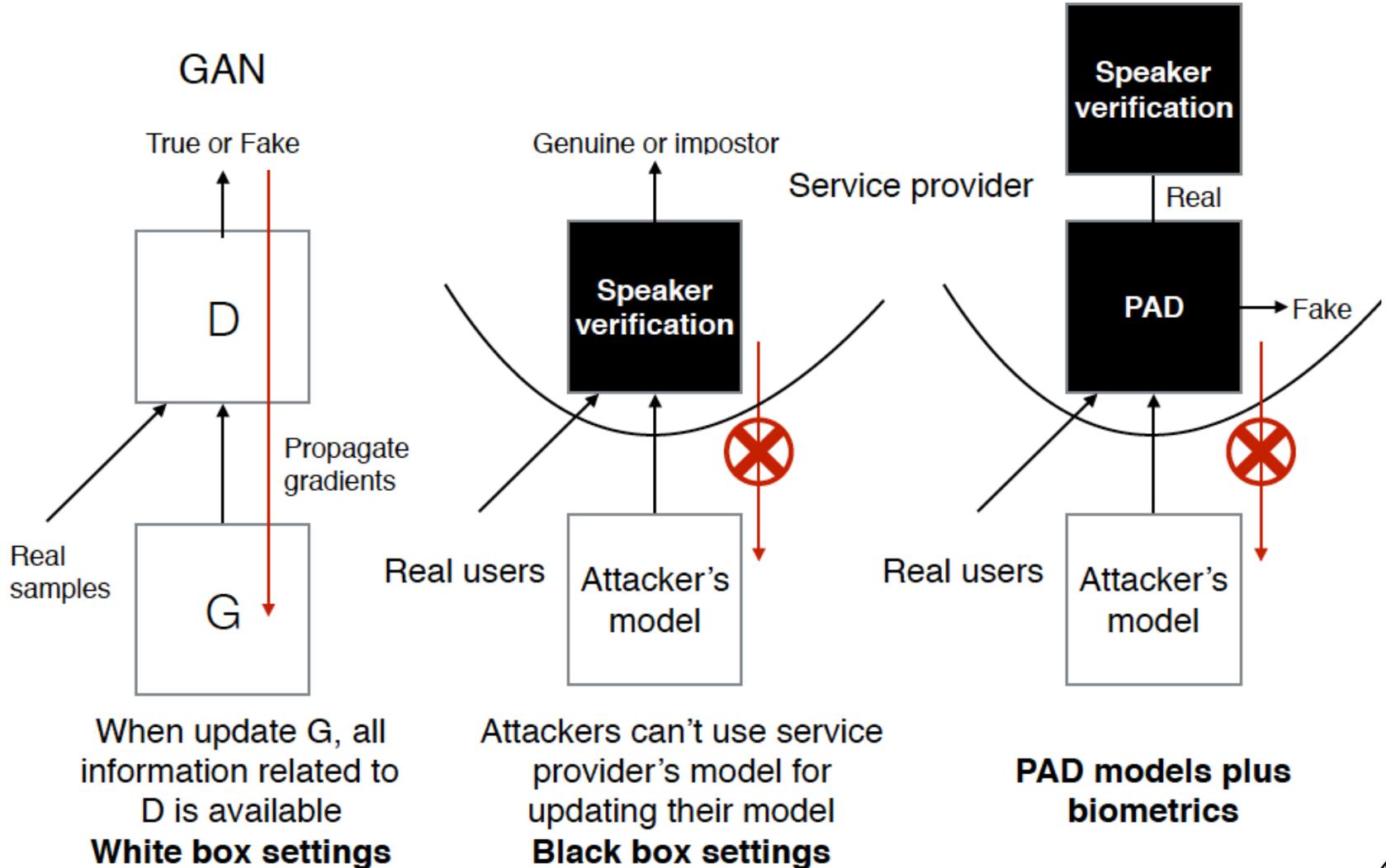
- 2017年 有名人の顔に入れ替えたポルノ映画がRedditに登場→アルゴリズムDeepFakeがGitHubに
- DeepFakeを見破るAIプログラム
  - 「瞬き」が不自然な箇所を検出
- Media Forensics (DARPA)
  - 音声や動画など、デジタル生成された情報の信憑性と正確性をチェックできるツールを開発
- AI同士の“文化戦争”
  - すでに人間には判別できない. . .

# Media Forensics (J. Yamagishi, 2019)

- 話者照合の脆弱性
  - 話者照合：声を使った手軽な生体認証
  - 声の模倣技術（音声合成や声質変換）を利用 → 標準話者識別システムに対して300人中288人の詐称に成功
- 話者照合における生体検知
  - 話者照合における防御技術の必要性
    - 生体検知：人間の生体から発声された音声かどうか判断
  - ASVspoof challengeという国際的なコンテスト
  - 音声合成と声質変換による合成音声と生体との識別精度：99%以上
  - スピーカから再生した音声と生体との識別精度：96%以上

すでに人間には  
判別できない！

# Spoofing/PAD and GAN





しかし、ここで扱っているのは...

However, ...

In Their Targets in ASVspoof challenge

- ***Detection of computer-generated*** fake speech/facial videos/reviews

- Note that human-generated fake media or the validation of semantics/contents of the media (***fact check***) are the out of the scope of the talk

# Flood of “Fake News”

## フェイクニュースの氾濫

We are living in uncertain, confusing times – when it can be hard to know what to believe

Fake news



Internet hoaxes  
'alternative facts'

'post truth'

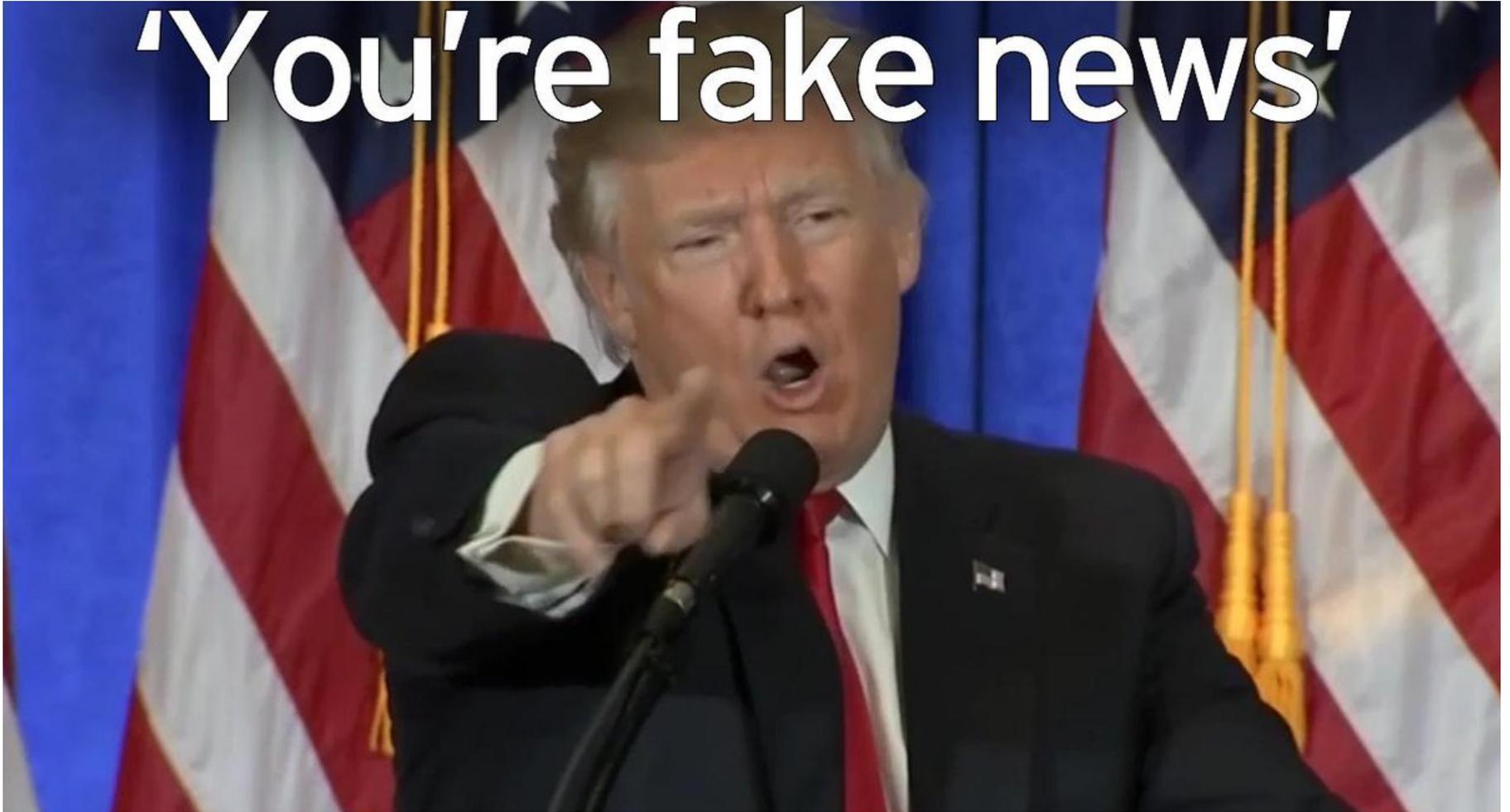
<https://news.yahoo.com/newshack/newshack/bbc-fakenews.html>

# たとえば...

- ニーチェ (Nietzsche) の言葉
  - 「神は死んだ。そして、我々は生き残った」
  - 「事実というものは存在しない。存在するのは解釈だけである」
- “真実”の証明方法
  - “悪魔の証明”になっていないか？
  - “1次資料”とは何か？
  - “認知バイアス”や“エコーチェンバー”の影響は？
  - “真実”の意味は？

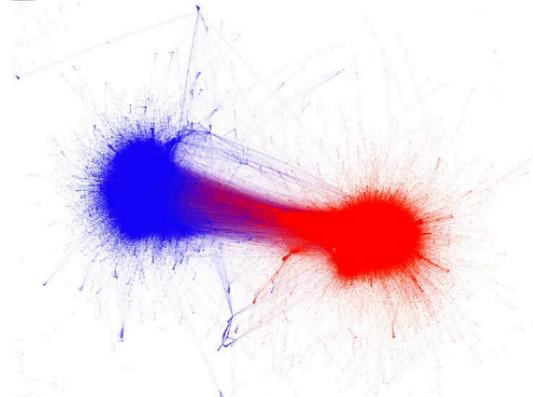
どうすれば…

**'You're fake news'**

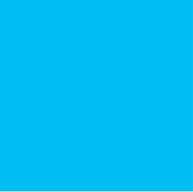


# フェイクニュース拡散の機序

- SNSは情報拡散装置
    - 「偽ニュース」は一部のユーザ(とbot)によって広く、遠くまで拡散する
  - 認知的バイアス
    - 直感、経験、思い込み
    - バックファイヤー効果：自分の世界観に固執
    - エコーチェンバー：似た者同士だけつながる
- 米大統領選2016のエコーチェンバー**



ツイートの拡散に見るリベラル系（青）と保守系（赤）のイデオロギーの分断

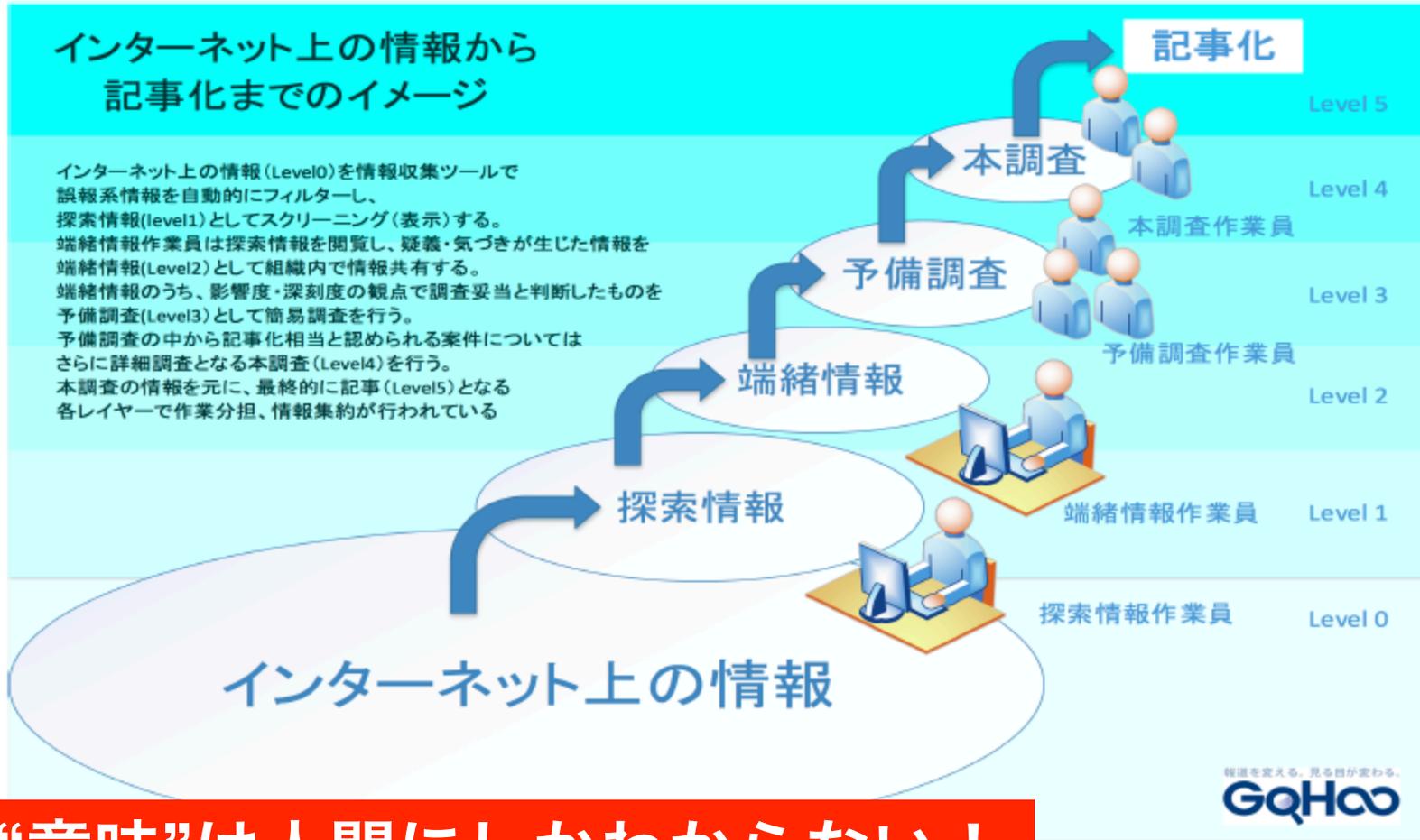


対策は？



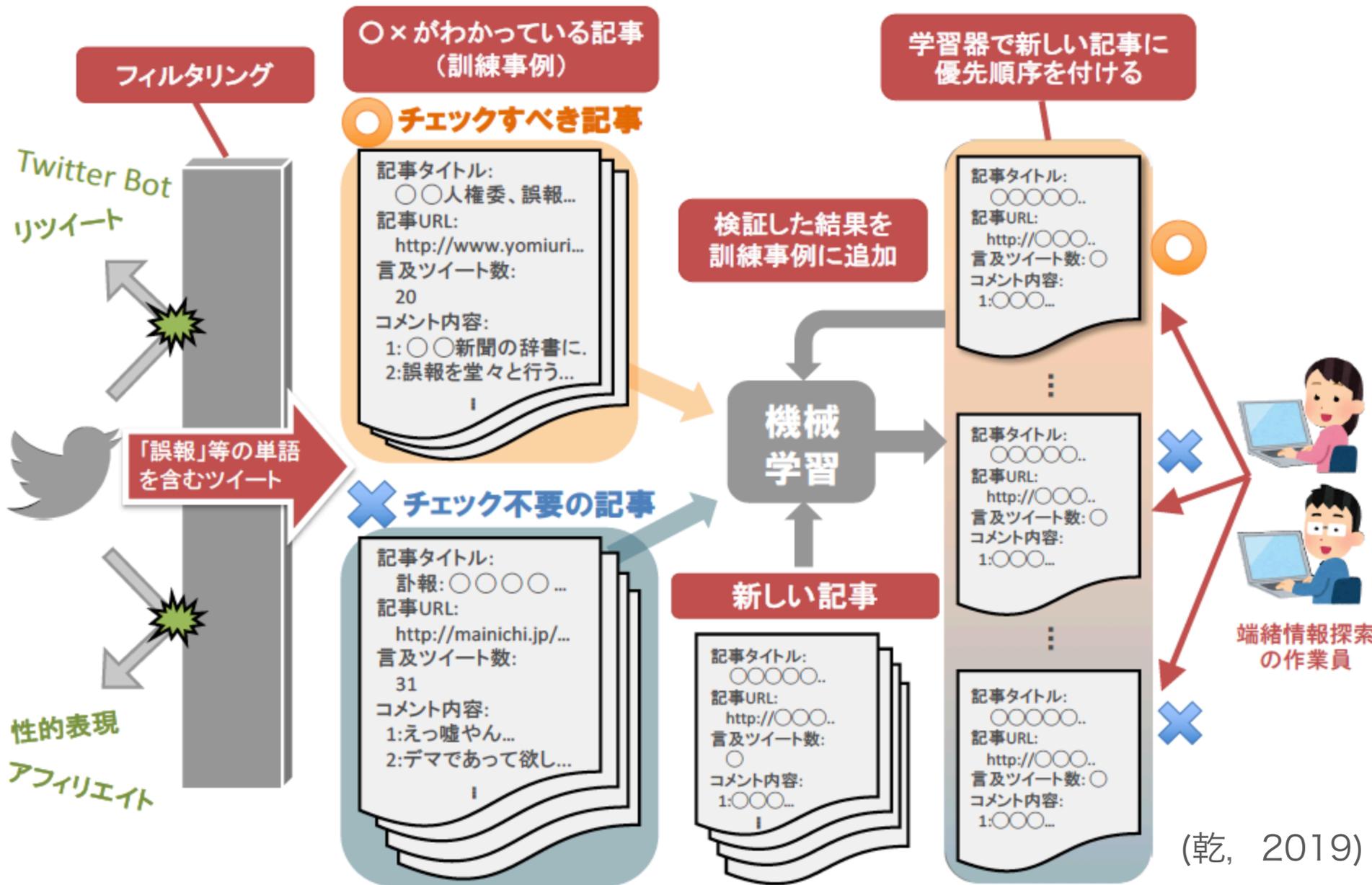
# ねらいはファクトチェックの営みを**支援**すること

## 記事 / 言説を評価するのはあくまで**人間**



**“意味”は人間にしかわからない！**

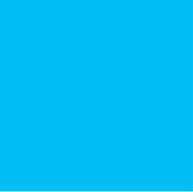
# 自然言語処理・機械学習による記事の絞り込み



# *Outline of Today's Lecture:*

## *Predicting Human Decision-Making*

- HCI, HAI/HRI におけるPHDMの重要性
  - *Should have been: Human-Interacting Agents based on the Predicting Human Decision-Making*
- 意思決定におけるモデルとは？
  - *Expert-Driven Paradigm: Theory*
  - *Data-Driven Paradigm: Practice*
- 知覚・行動・他者の思考の推論(意識)
  - *Predictive Coding*
  - *Trust and Decision-Making*



おわり