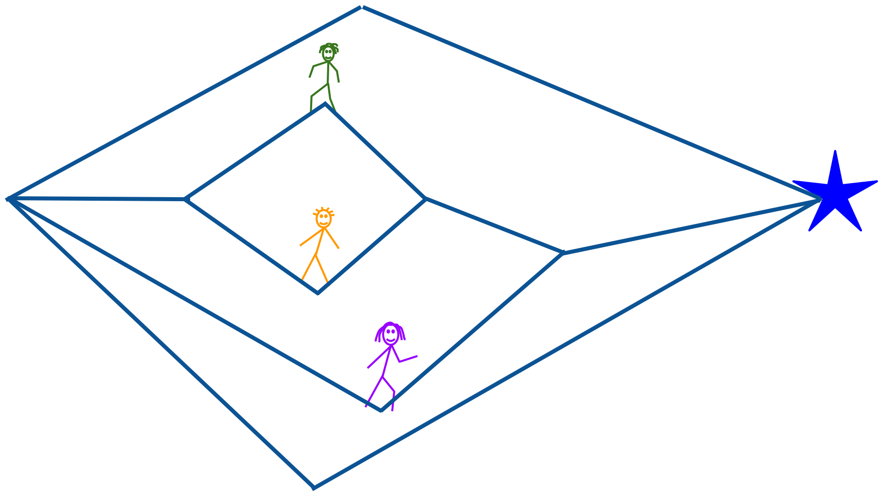
Informational Substitutes



Yiling Chen Harvard
Bo Waggoner UPenn

TCS+ talk
Nov 9, 2016

Goals for this talk

1. Introduce a theory audience to some fun topics:
    - value of information
    - proper scoring rules
    - prediction markets

2. Describe this paper (Chen, Waggoner FOCS 2016):
    - propose definitions of subs. and comps.
    - characterize “good” and “bad” equilibria of prediction markets
    - connect to complexity of information acquisition

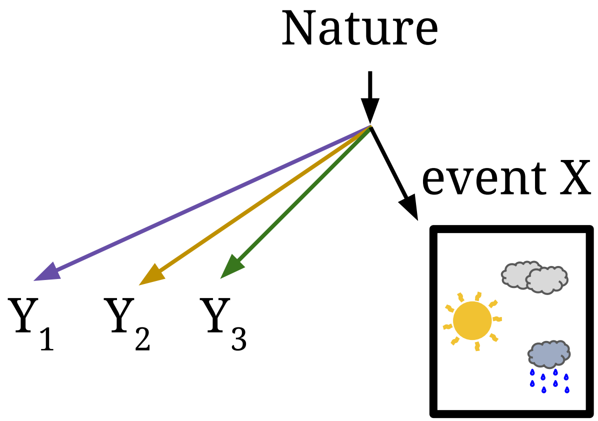
0. Information

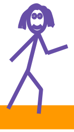
Information (in this talk)

Random variables X, $Y_1, \ldots, Y_n$ jointly distributed, known prior. (finite set of outcomes)

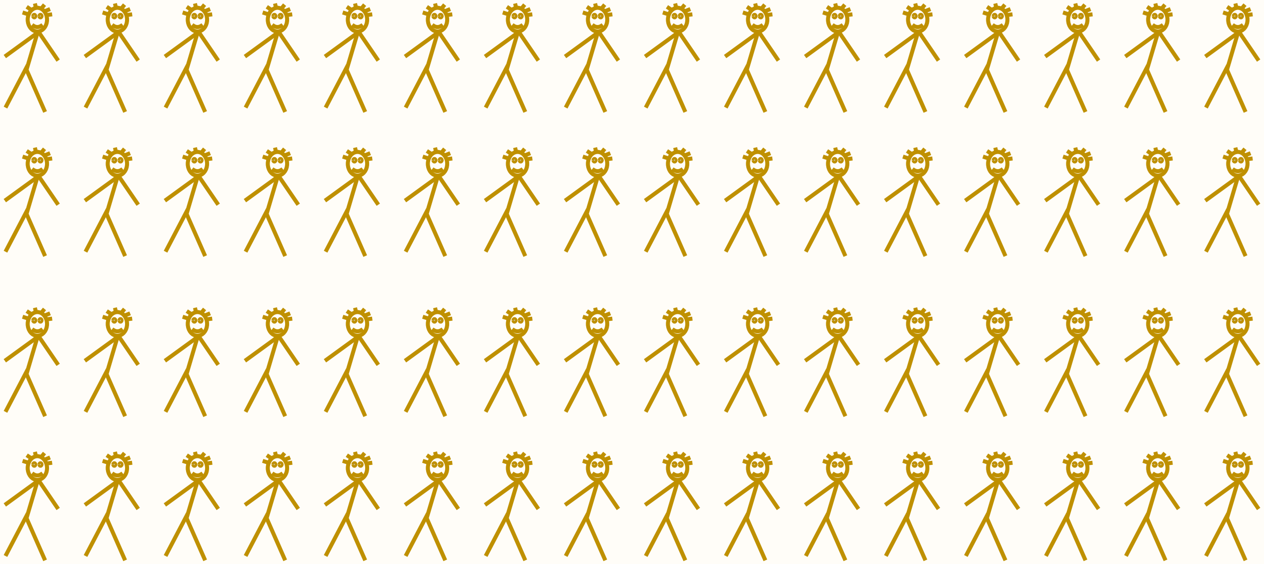
We care about X.

Y_i = "signal" (reveals info. about X).





1. Substitutes



The unreasonable effectiveness of substitutes

Substitutes in economics:

- Market equilibria, stable matchings, …
- [Kelso & Crawford 1982, Roth 1984, Hatfield and Milgrom 2005, …]

Substitutes in computer science:

- Submodularity! [Lehmann+Lehmann+Nisan 2001]
- subs == efficient approx. for **many** problems

Could we also define "substitutes" for information?

And could they also link algorithms and game theory?

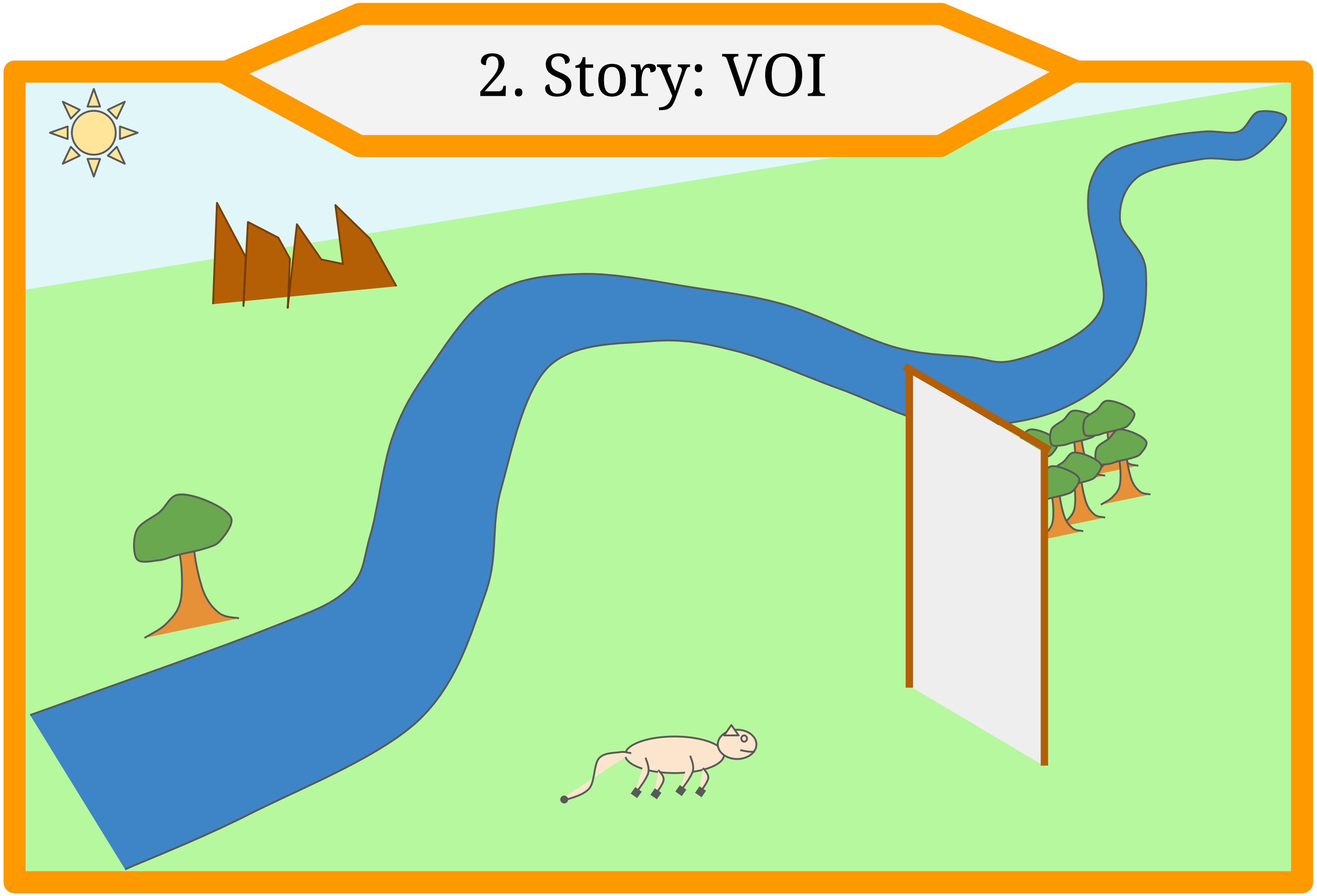
Challenges for defining informational S&C

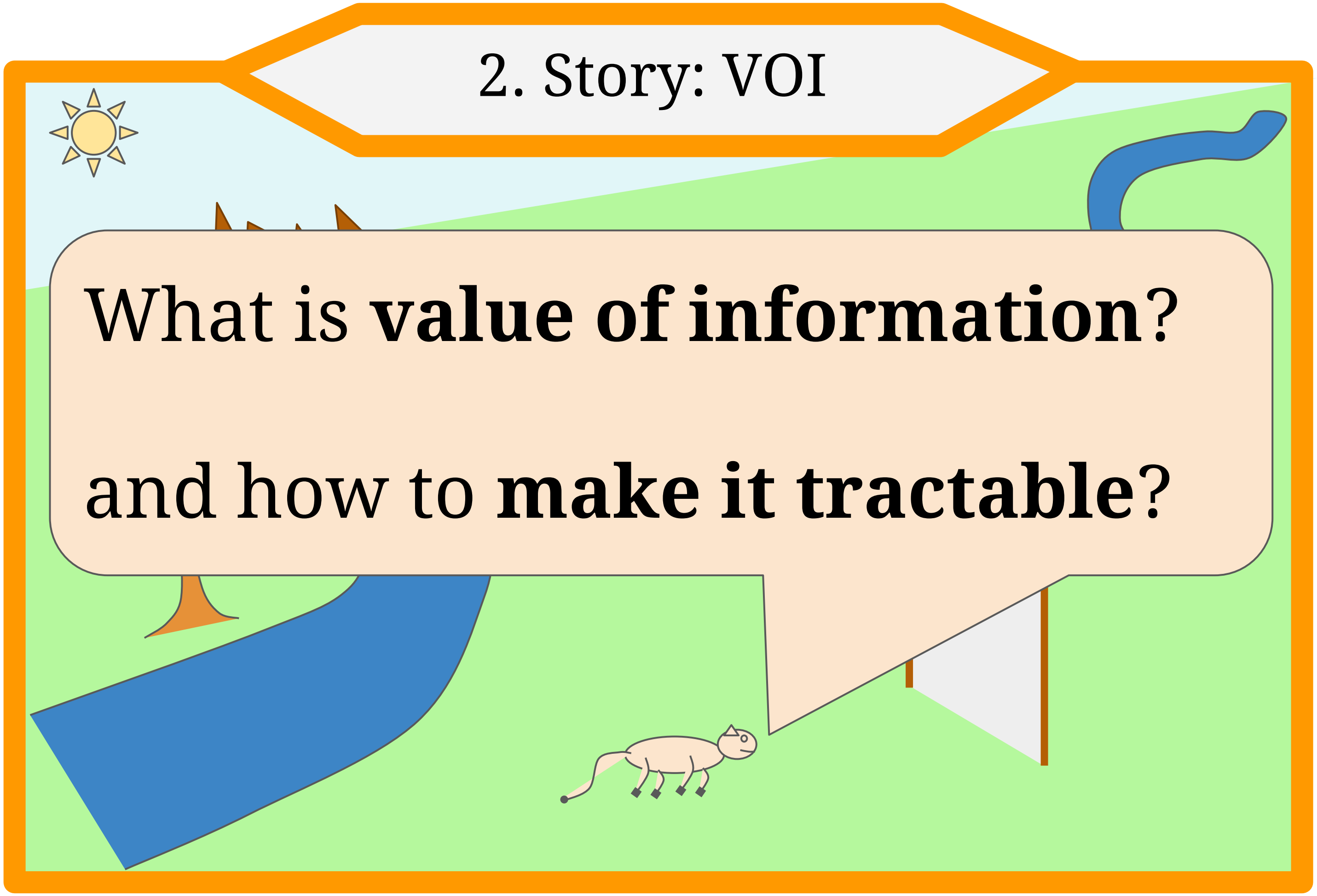
Items

- Valuation function f is given
- f(i) does not depend on f(j)
- “marginal value” is straightforward

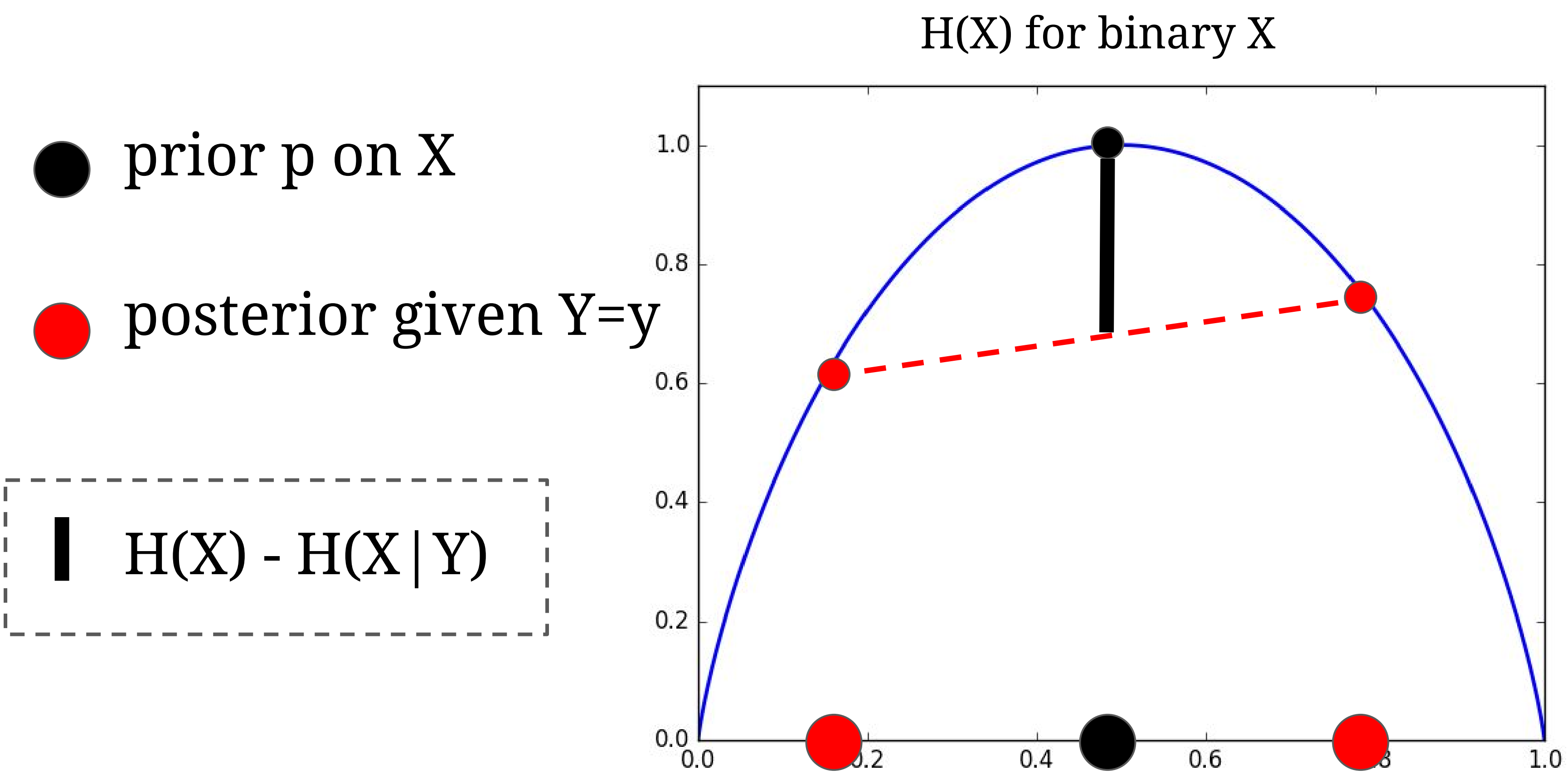
Information

- What is the “value” of a set of pieces of information?
- Two pieces of information may be correlated, redundant, …
- What is a “marginal” piece of information?


2. Story: VOI


2. Story: VOI
What is **value of information**?
and how to **make it tractable**?

The Story part 1: Shannon 1948



Example where Shannon does not apply

X = rain or no rain

Y = weather forecast

I hate carrying my umbrella….

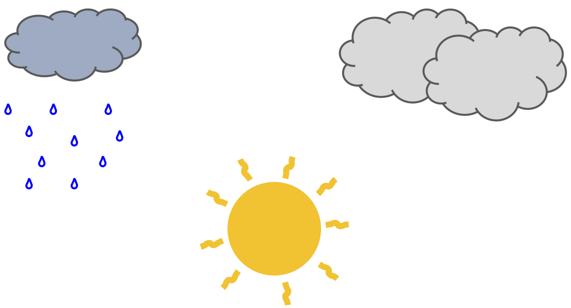
But if it rains on me → I melt

The Story part 2: Howard 1966

1. Known prior p on X

2. Select decision d

2. Nature draws x ~ p



3. Get utility u(d, x).

V(∅) = "expected utility when deciding optimally with **no signals**"

The Story part 2: Howard 1966

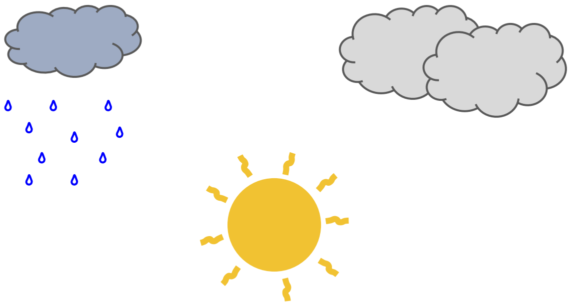
1. Known prior p on X

1.5. **Observe Y**, Bayesian update to p_y

---

2. Select decision d

2. Nature draws x ~ p_y



---

3. Get utility u(d, x).

V(**Y**) = "expected utility when deciding optimally after **observing Y**"

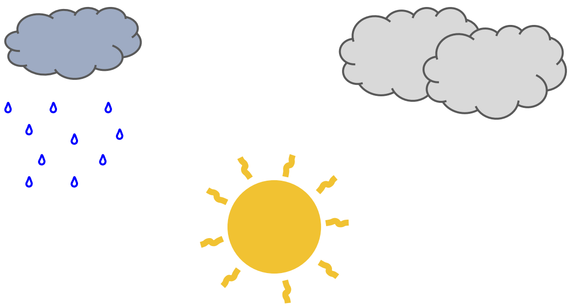
The Story part 2: Howard 1966

1. Known prior p on X

1.5. **Observe Y**, Bayesian update to p_y

2. Select decision d

2. Nature draws $x \sim p_y$



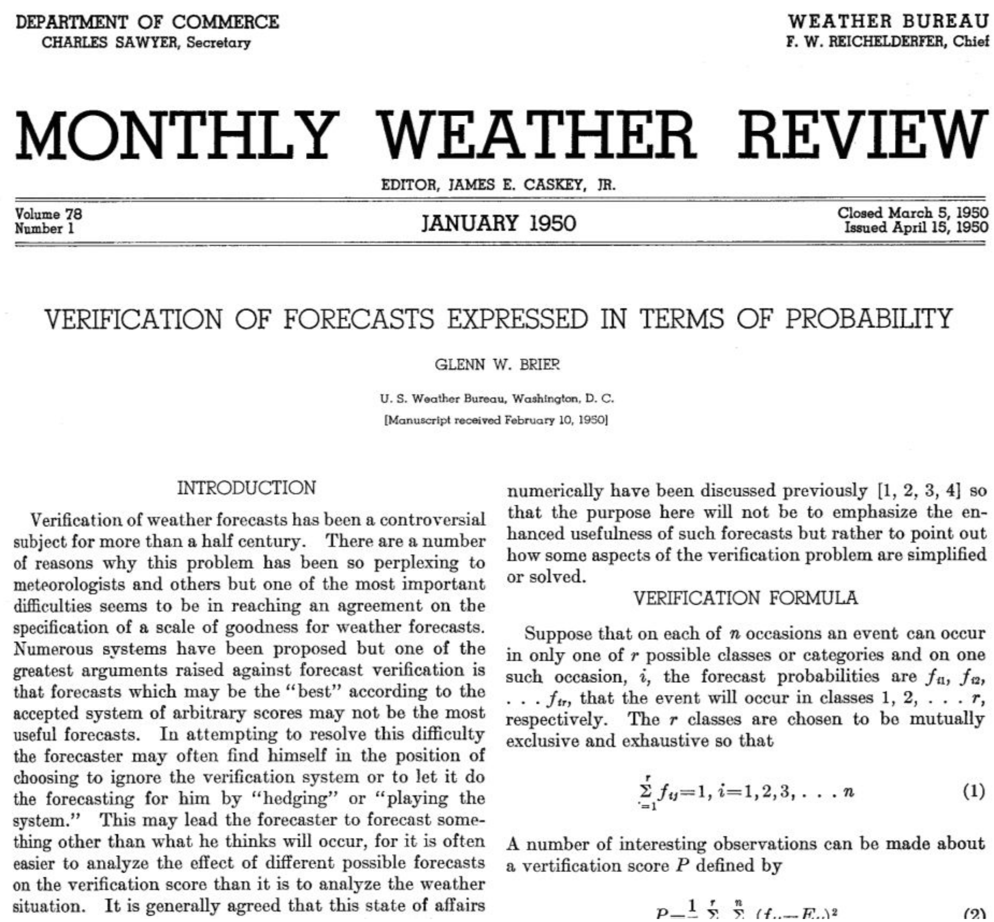
3. Get utility u(d, x).

V(**Y**) = "expected utility when deciding optimally after **observing Y**"

V(Y) - V(∅) = "marginal value of Y"

But: how to make VOI tractable?

Our approach: apply ideas from **proper scoring rules**


DEPARTMENT OF COMMERCE
CHARLES SAWYER, Secretary

WEATHER BUREAU
F. W. REICHELDERFER, Chief

MONTHLY WEATHER REVIEW

EDITOR, JAMES E. CASKEY, JR.

Volume 78
Number 1

JANUARY 1950

Closed March 5, 1950
Issued April 15, 1950

VERIFICATION OF FORECASTS EXPRESSED IN TERMS OF PROBABILITY

GLENN W. BRIER

U. S. Weather Bureau, Washington, D. C.

[Manuscript received February 10, 1950]

INTRODUCTION

Verification of weather forecasts has been a controversial subject for more than a half century. There are a number of reasons why this problem has been so perplexing to meteorologists and others but one of the most important difficulties seems to be in reaching an agreement on the specification of a scale of goodness for weather forecasts. Numerous systems have been proposed but one of the greatest arguments raised against forecast verification is that forecasts which may be the "best" according to the accepted system of arbitrary scores may not be the most useful forecasts. In attempting to resolve this difficulty the forecaster may often find himself in the position of choosing to ignore the verification system or to let it do the forecasting for him by "hedging" or "playing the system." This may lead the forecaster to forecast something other than what he thinks will occur, for it is often easier to analyze the effect of different possible forecasts on the verification score than it is to analyze the weather situation. It is generally agreed that this state of affairs numerically have been discussed previously [1, 2, 3, 4] so that the purpose here will not be to emphasize the enhanced usefulness of such forecasts but rather to point out how some aspects of the verification problem are simplified or solved.

VERIFICATION FORMULA

Suppose that on each of n occasions an event can occur in only one of r possible classes or categories and on one such occasion, i, the forecast probabilities are f_{i1}, f_{i2}, . . . f_{ir}, that the event will occur in classes 1, 2, . . . r, respectively. The r classes are chosen to be mutually exclusive and exhaustive so that

$$\sum_{j=1}^{r} f_{ij}=1,\ i=1,2,3,\ldots n \qquad (1)$$

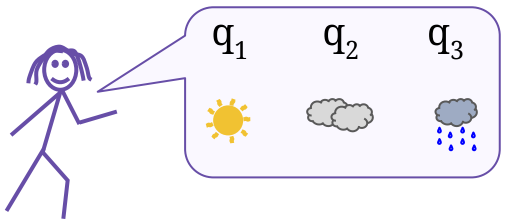
A number of interesting observations can be made about a vertification score P defined by

$$P=\frac{1}{n}\sum_{j=1}^{r}\sum_{i=1}^{n}(f_{ij}-E_{ij})^2 \qquad (2)$$

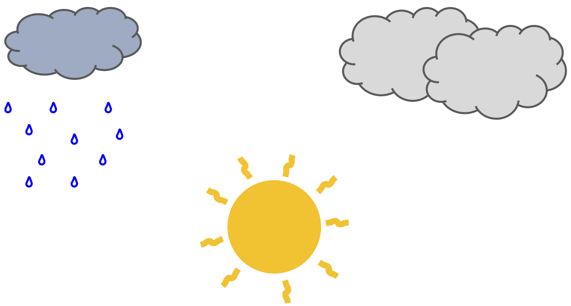
The Story part 3: Savage 1971, "scoring rules"

1. Known prior p on X

2. Select **prediction q**



2. Nature draws x ~ p



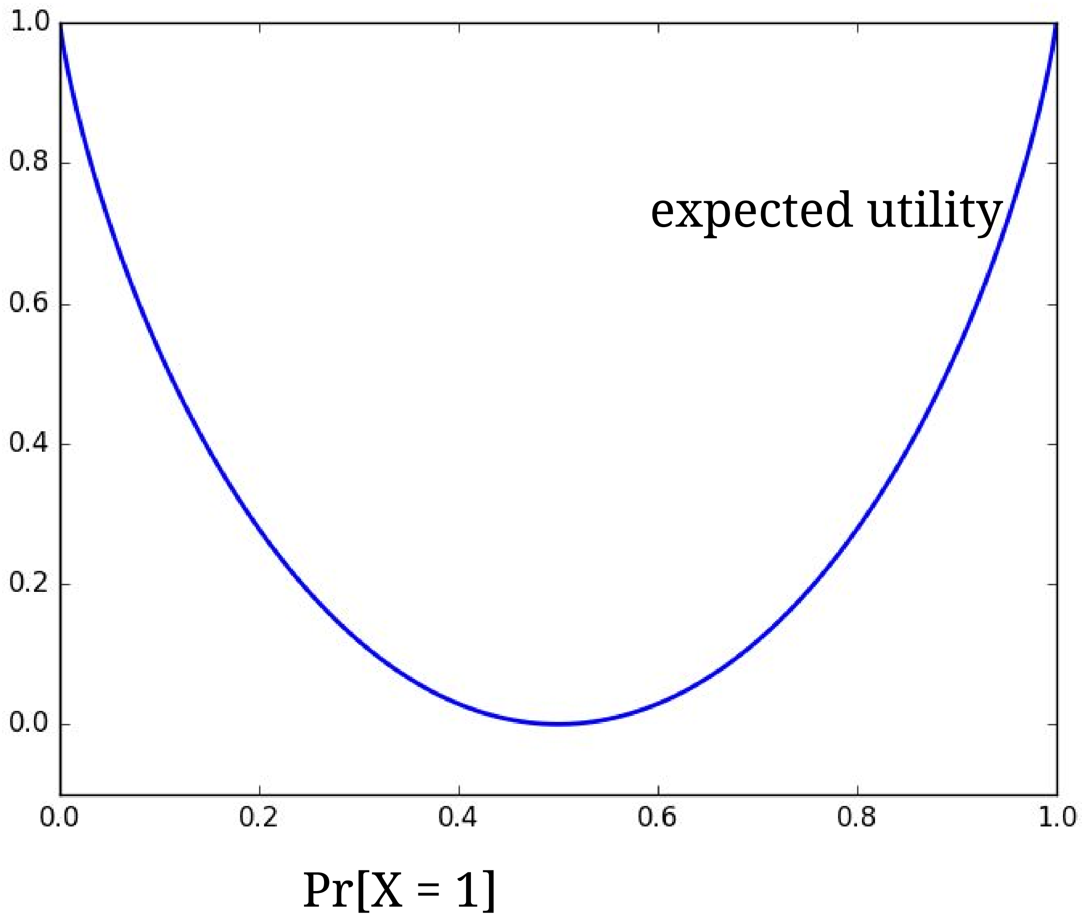
3. Get utility **S(q, x)**.

"Proper scoring rule" - optimal prediction is true belief

Example: $S(q, x) = \log q(x)$.

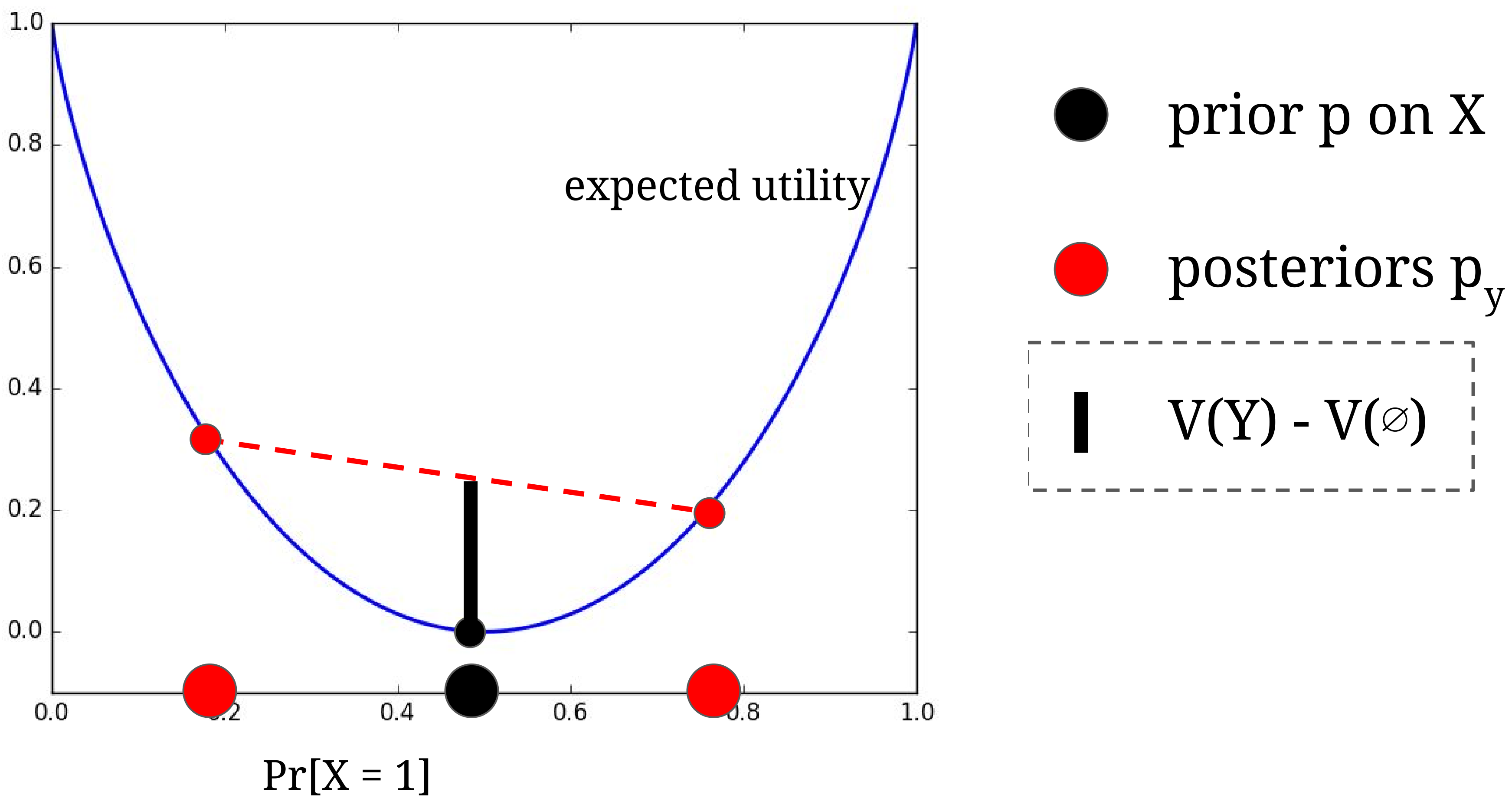
Savage: scoring rules ⟷ convex functions!

Example: S(q, x) = log q(x).



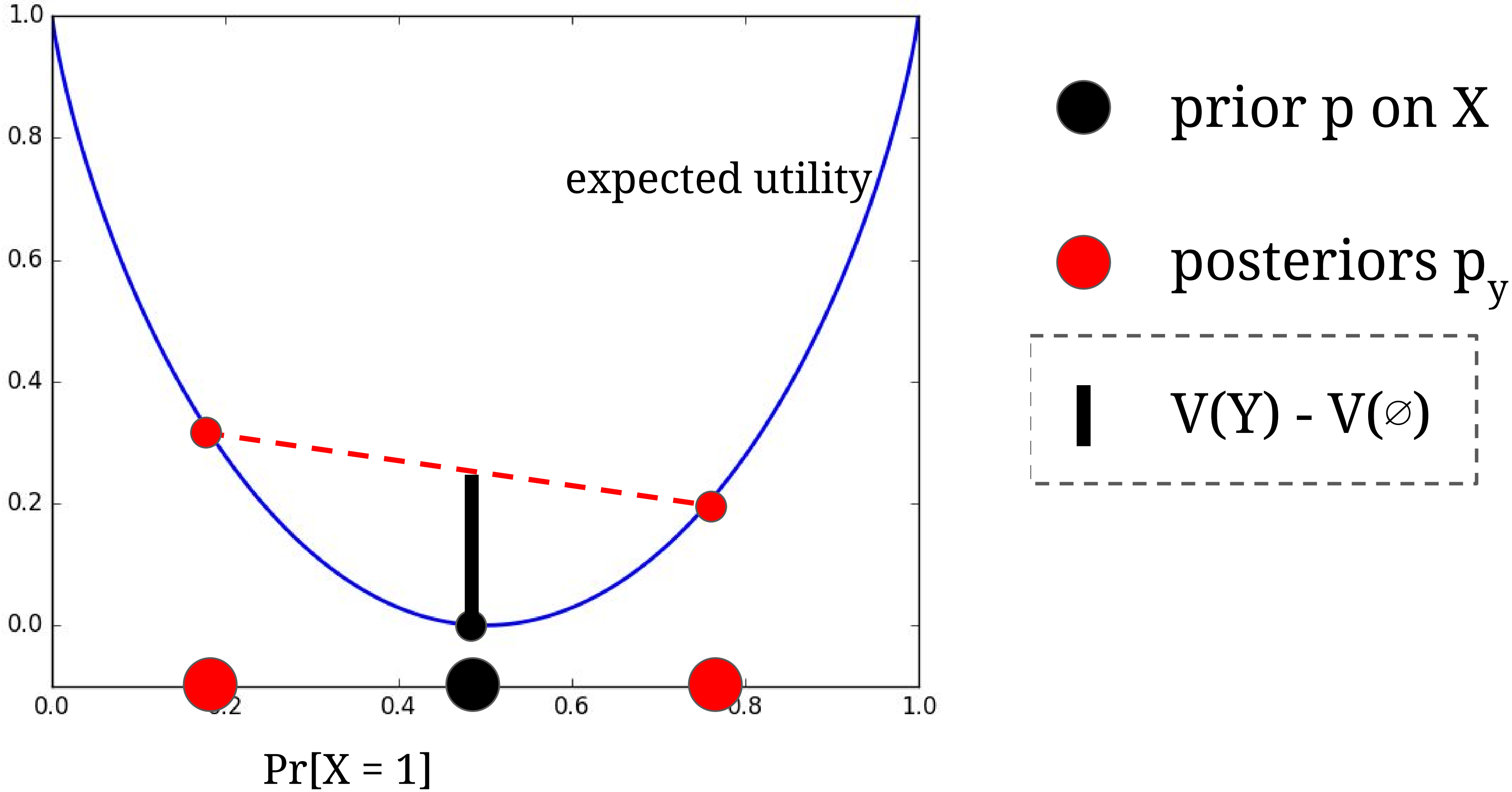
Savage: scoring rules ⟵⟶ convex functions!

Example: $S(q, x) = \log q(x)$.

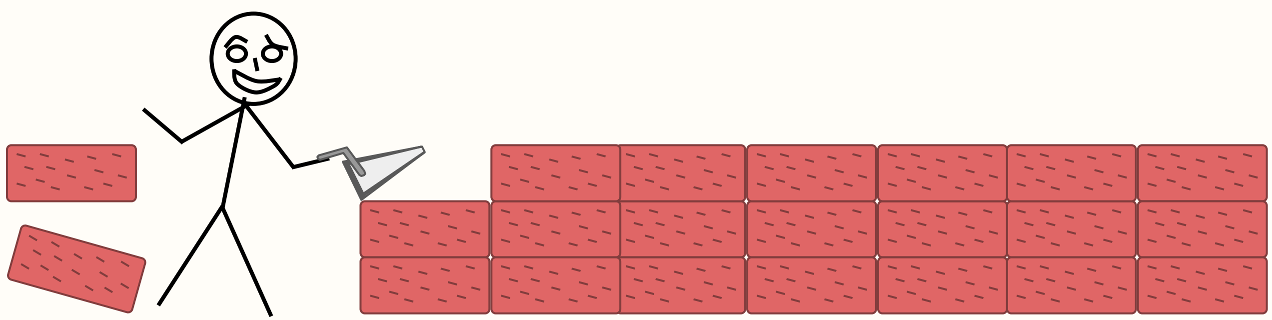


DECISION PROBLEMS ⟷ convex functions!

Example: S(q, x) = log q(x).



3. Definitions



Our definitions

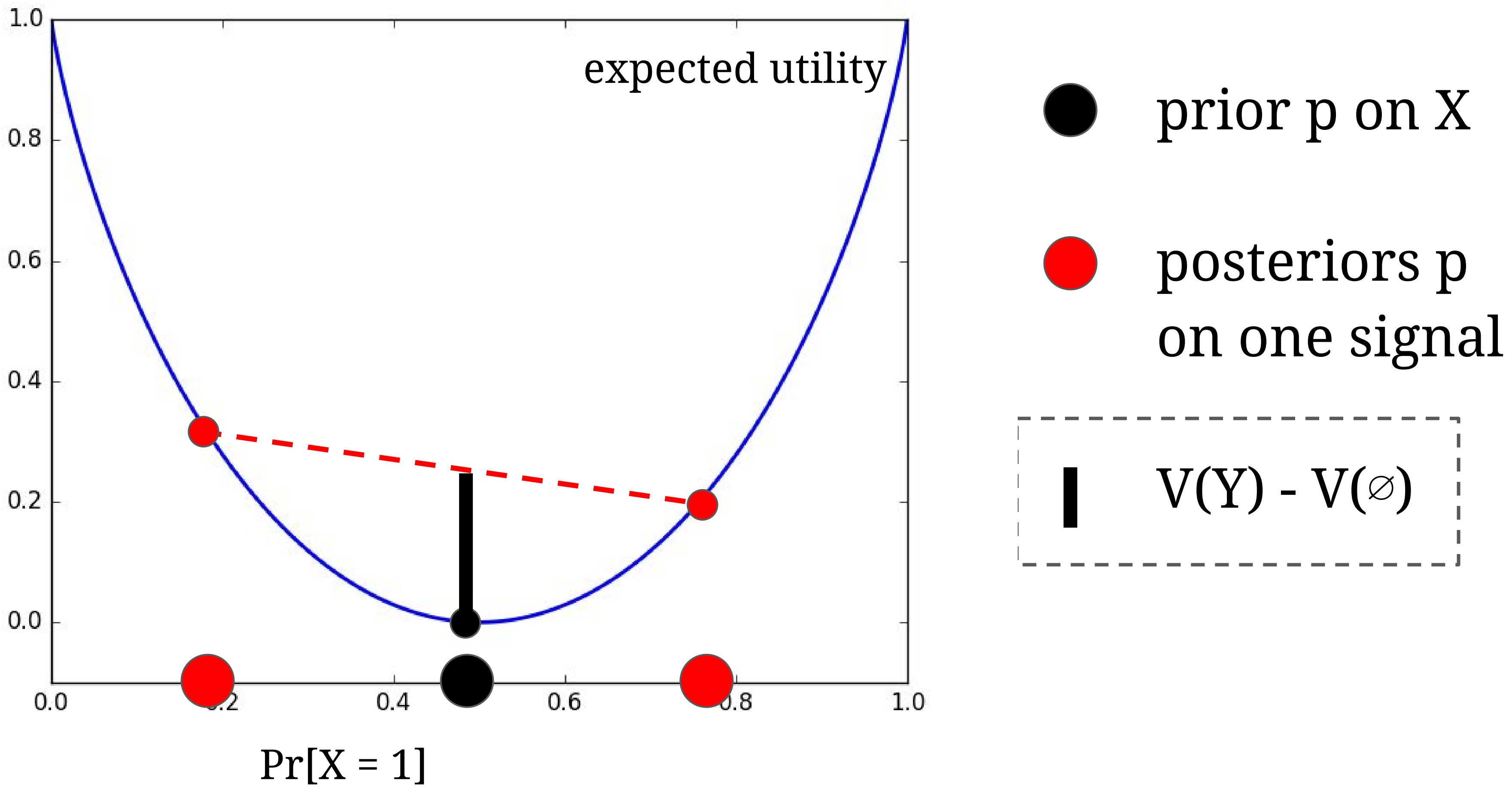
$Y_1 \ldots Y_n$ are **substitutes** for u if V is **submodular**:
For $A \subseteq B \subseteq \{Y_1 \ldots Y_n\}$,

$$V\big(A \cup \{Y_i\}\big) - V(A) \geq V\big(B \cup \{Y_i\}\big) - V(B).$$

- complements = supermodular
- depends on **both decision prob AND info structure**

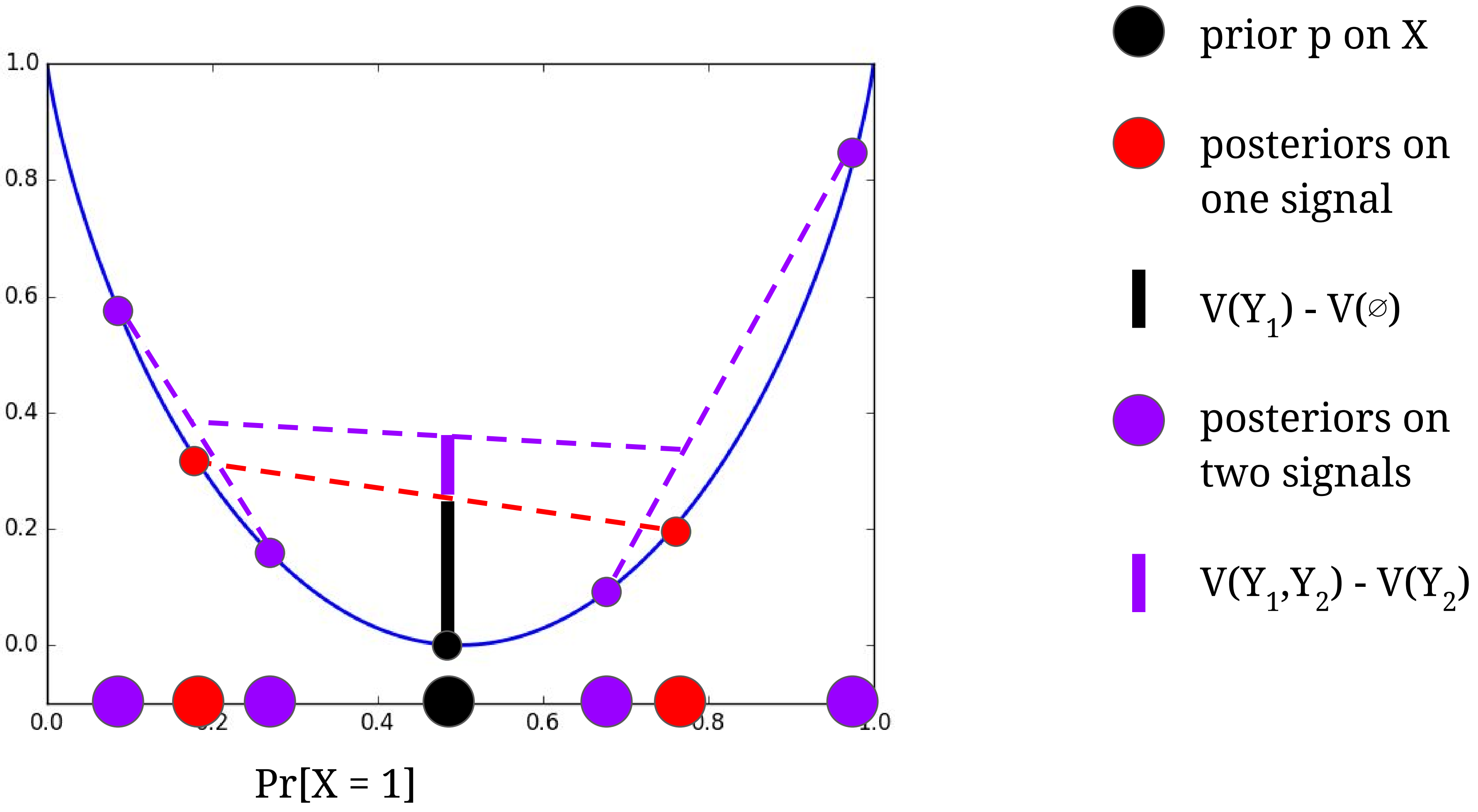
Visualizing an example of substitutes

Example: $S(q, x) = \log q(x)$. Y_1, Y_2 i.i.d. conditioned on x.

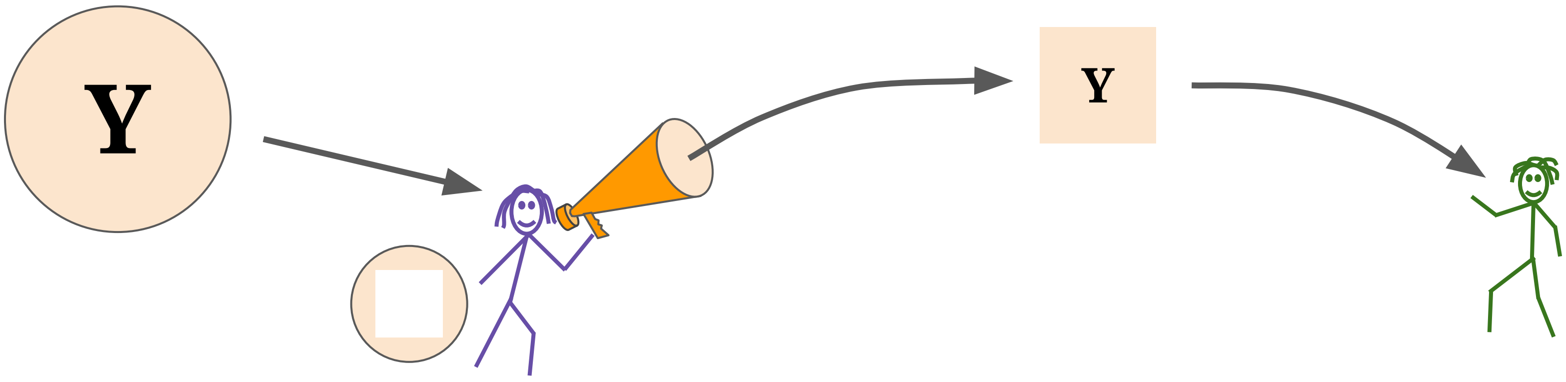


Visualizing an example of substitutes

Example: $S(q, x) = \log q(x)$. Y_1, Y_2 i.i.d. conditioned on x.



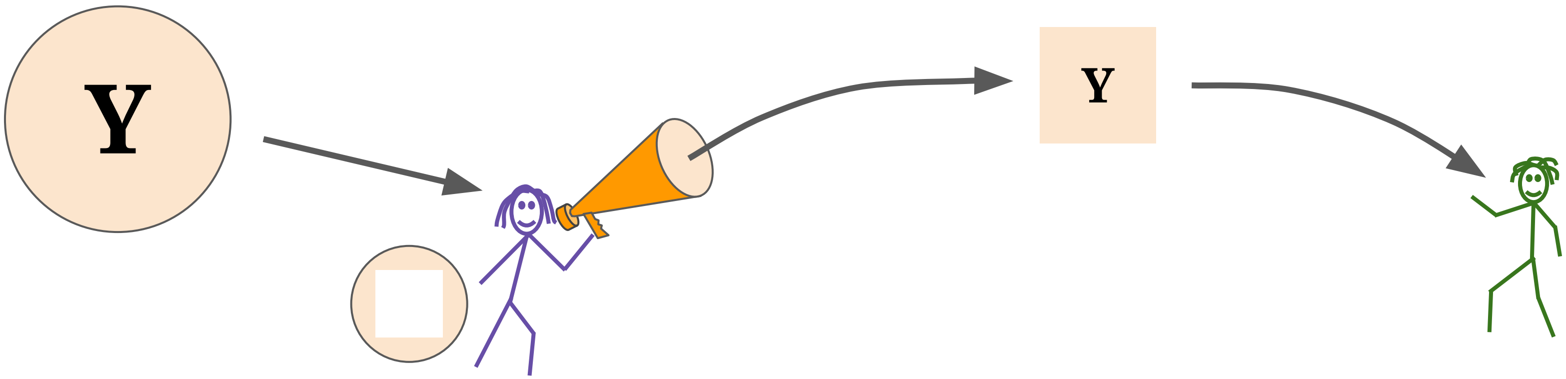
Roadblock: Information is divisible!



“Half the truth is often a great lie.”

- Benjamin Franklin

Roadblock: Information is divisible!



“Half the truth is often a great lie.”
- Benjamin Franklin

Example: Alice observes entire stock market,
but strategically reports one stock’s performance.

Pragmatic solution

$Y_1 \ldots Y_n$ are **strong substitutes** for u if:
For $A \subseteq B \subseteq \{Y_1 \ldots Y_n\}$ and **any randomized function f,**

$$V\big(A \vee \mathbf{f(Y_i)}\}\big) \text{ - } V(A) \geq V\big(B \vee \mathbf{f(Y_i)}\big) \text{ - } V(B).$$

where $A \vee f(Y_i)$ is the signal conveying both A and $f(Y_i)$.

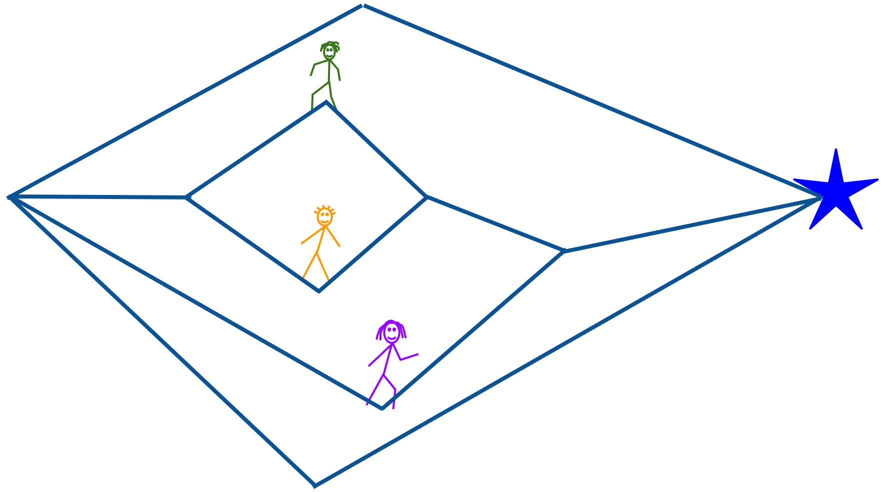
Context: “signal lattices”

Lattice: partially ordered set closed under:

- **meet** $A \wedge B$: greatest C that is less than both
- **join** $A \vee B$: least C that is greater than both

Example: subsets. (meet = intersection, join = union)

For signals: meet = “observe both”, join = “common knowledge”.



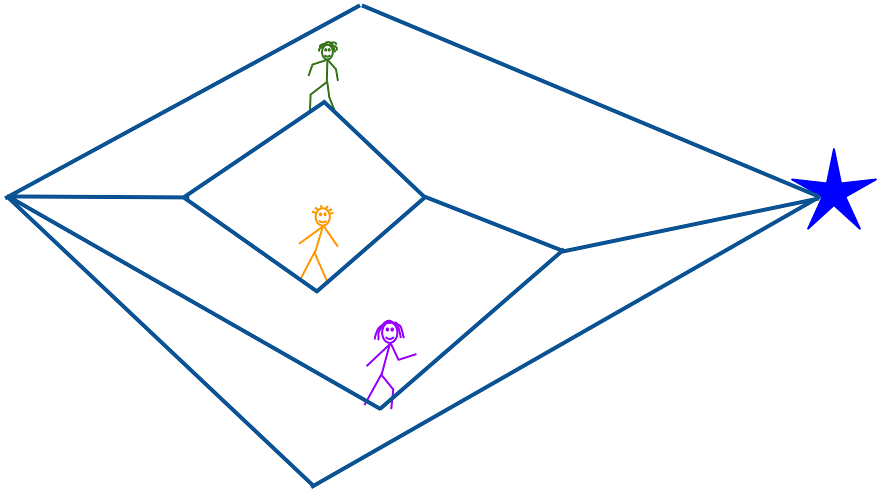
The “continuous” signal lattice

Idea: given X and $Y_1 \dots Y_n$,

The set of **randomized functions of sets of Y_is**

form a lattice of signals

ordered by informativeness (about X).



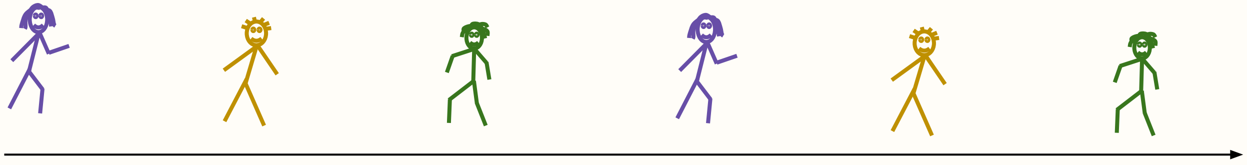
-- Remainders --

Two separate applications:

- Markets (for information).
  substitutes $\longleftrightarrow$ good equilibria

- Algorithms.
  complexity of optimal info. acquisition

4. Prediction markets



Idea / motivation

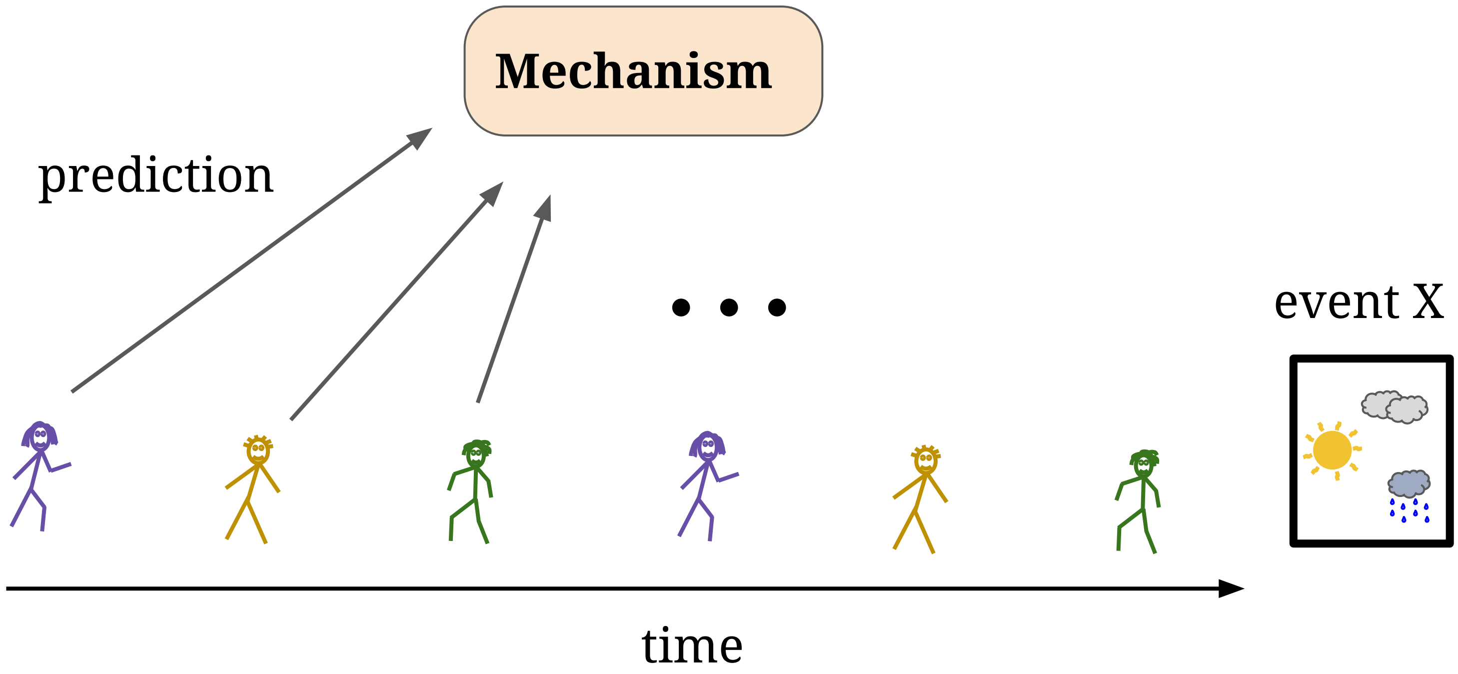
Each agent has a signal Y_i.

Goal: aggregate into prediction about X **quickly**.

Idea / motivation

Each agent has a signal Y_i.

Goal: aggregate into prediction about X **quickly**.



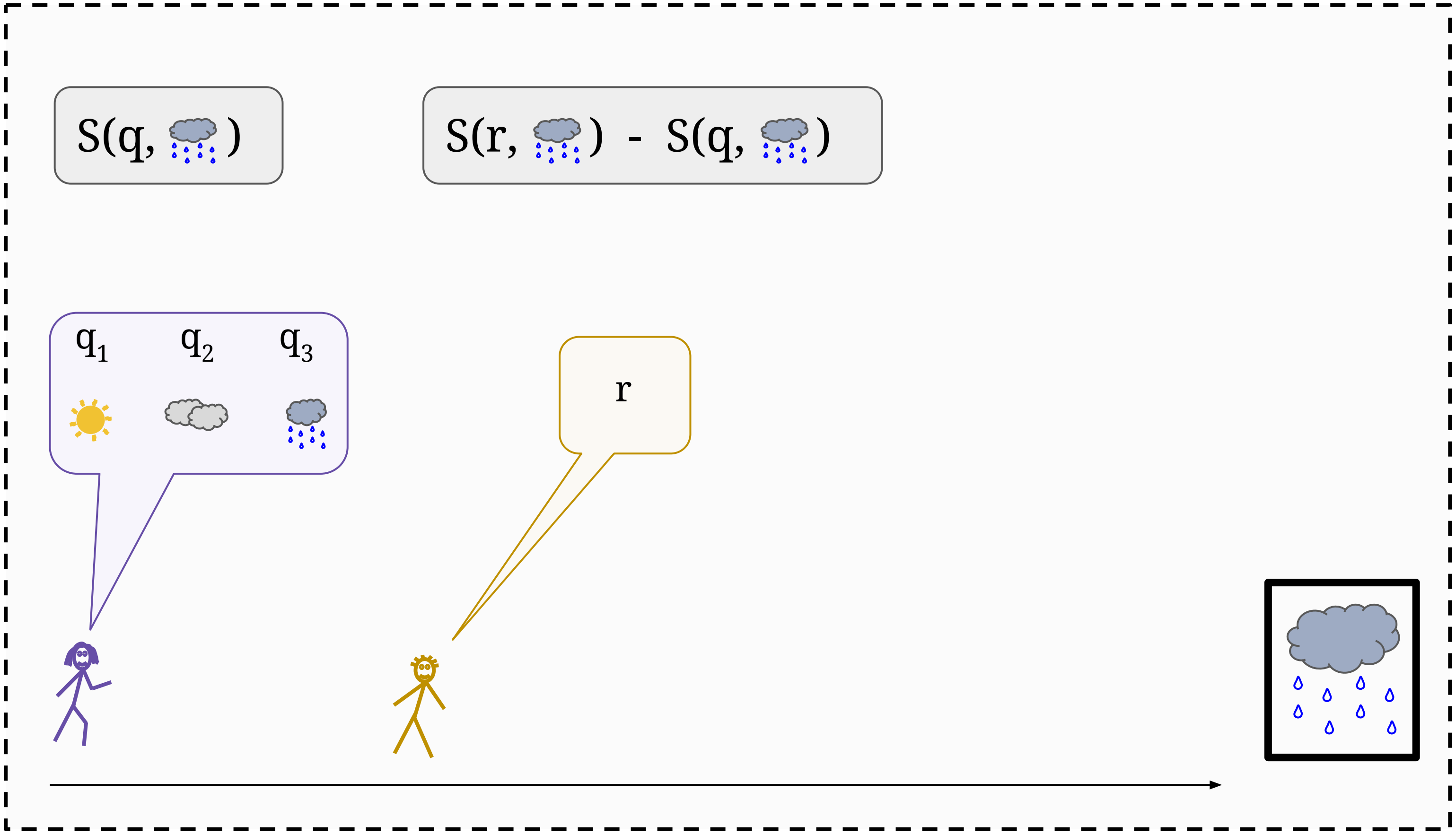
The mechanism [Hanson 2003]*

Only one participant: proper scoring rule! Truthful.

*can also be viewed as buying/selling shares [Abernethy+Chen+Wortmann-Vaughan 2013]

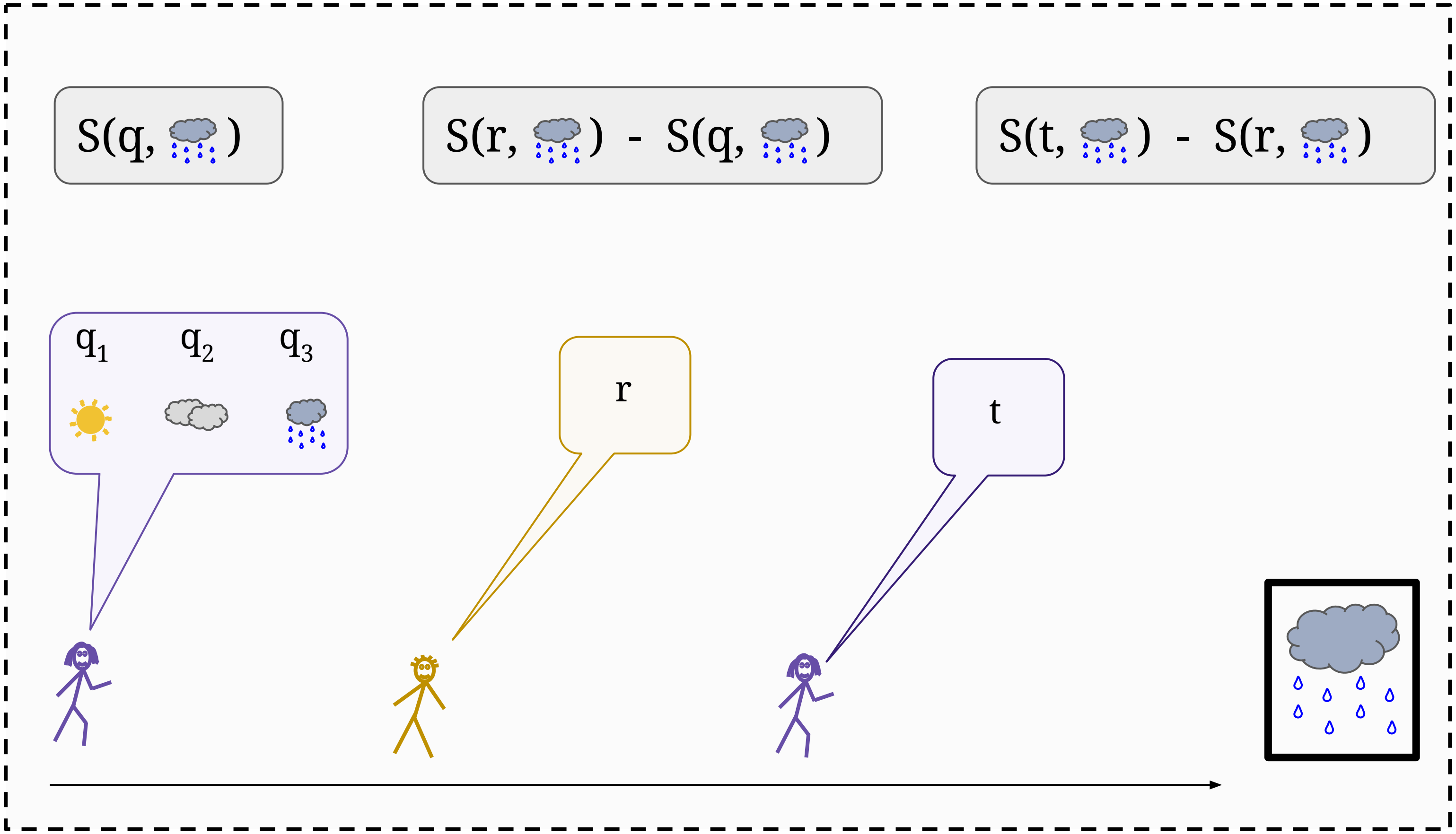
The mechanism [Hanson 2003]

Two participants: “chained” scoring rule! Truthful.



The mechanism [Hanson 2003]

Two participants, three stages: **not understood!**



The mechanism [Hanson 2003]

Two participants, three stages: **not understood!**

Known: for **log scoring rule**, if $Y_1 \dots Y_n$ are…

- conditionally independent on X ⇒ "rush".
  [Chen+Dimitrov+Sami+Reeves+Pennock+Hanson+Fortnow+Gonen 2010]

- independent ⇒ "delay".
  [Gao+Zhang+Chen 2013]

Prediction markets results

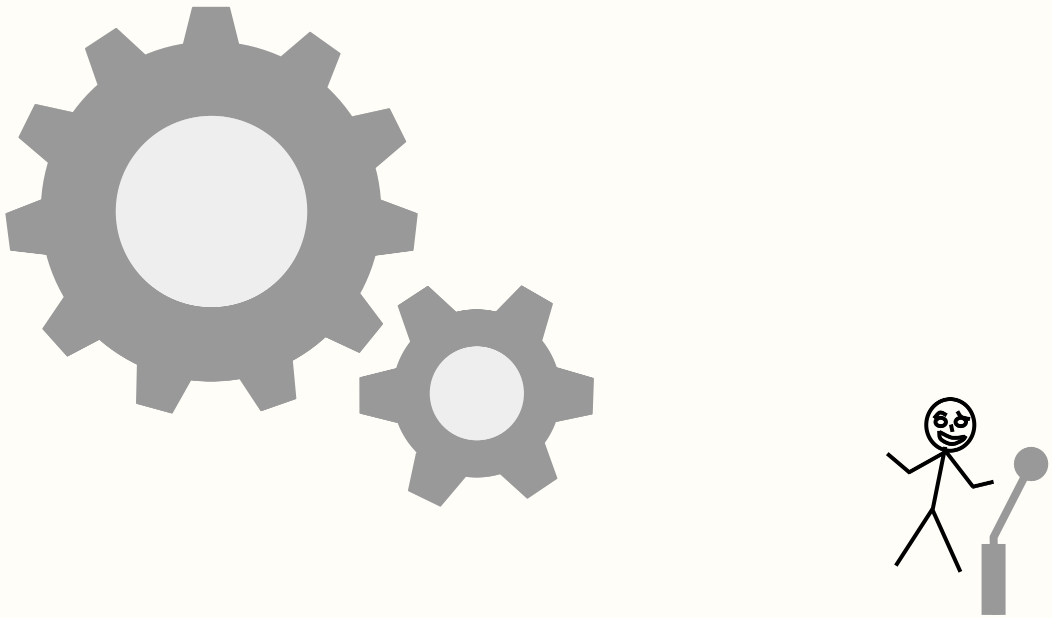
Thm. If and only if signals are strong **substitutes**, the only equilibria are “**all rush**”.

(efficient market hypothesis $\longleftrightarrow$ substitutes)

Thm. If and only if signals are strong **complements**, the only equilibria are “**all delay**”.

(market failure $\longleftrightarrow$ complements)

5. Algorithms



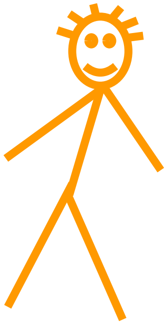
Algorithmic question "SignalSelection"

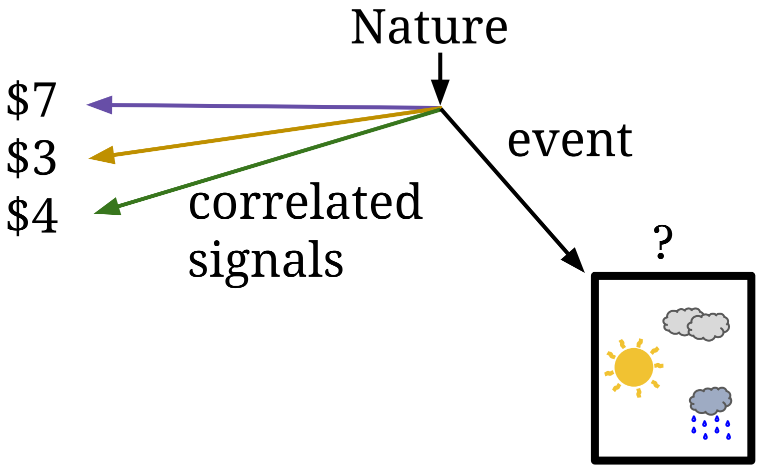
Input:

- utility function u (as an oracle…)
- joint distribution X, $Y_1 \ldots Y_n$ (as an oracle…)
- prices $\pi_1 \ldots \pi_n$ for the signals, budget constraint B

Output:

- which signals to acquire





Complexity results

Reduction: SignalSelection → set function maximization.

Substitutes ⇒ 1-1/e approx in polynomial time

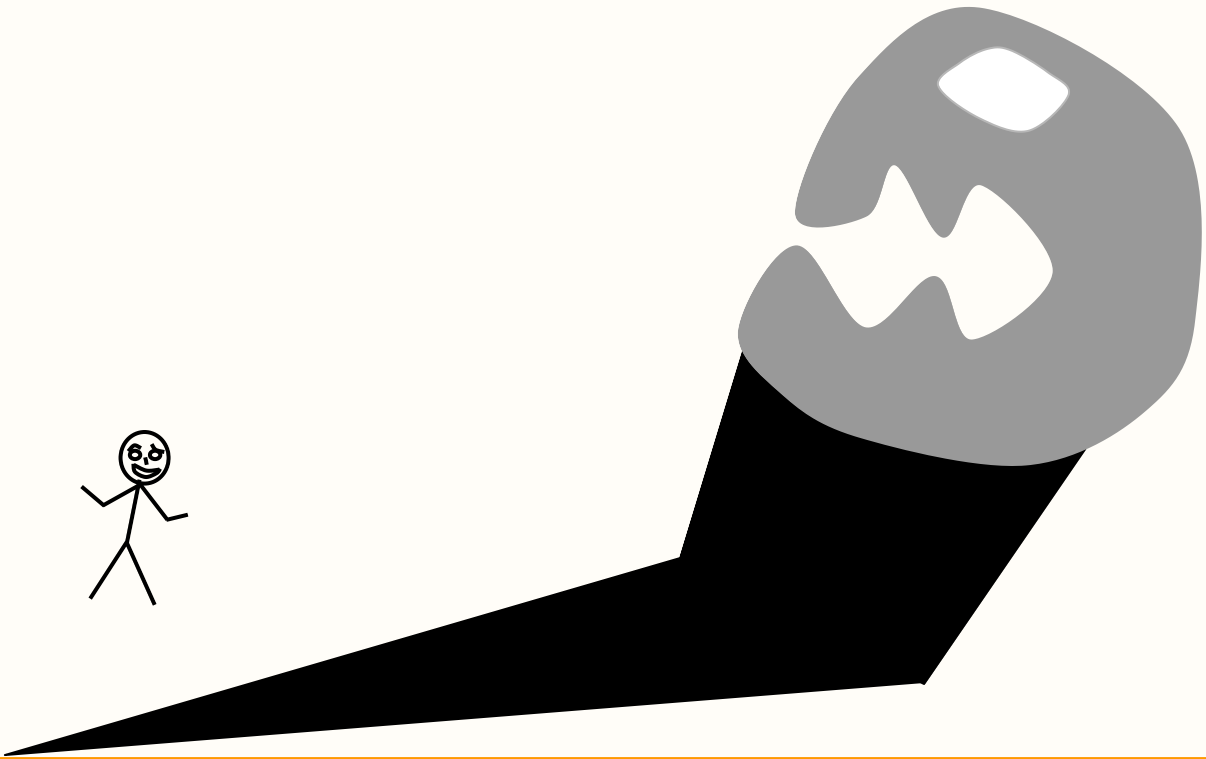
Reduction: set function maximization → SignalSelection.

Comps/generally ⇒ no approx w/ subexp. queries

Notes:

- As in submod. maximization, can handle e.g. matroid constraints.
- Ideas not new here at all! See survey [Krause+Guestrin 2011]
- Model / generality, focus of our question are new

6. Shadows



What do we *know* about subs and comps?

Examples: log scoring rule (Shannon entropy).

Intuition / geometry:

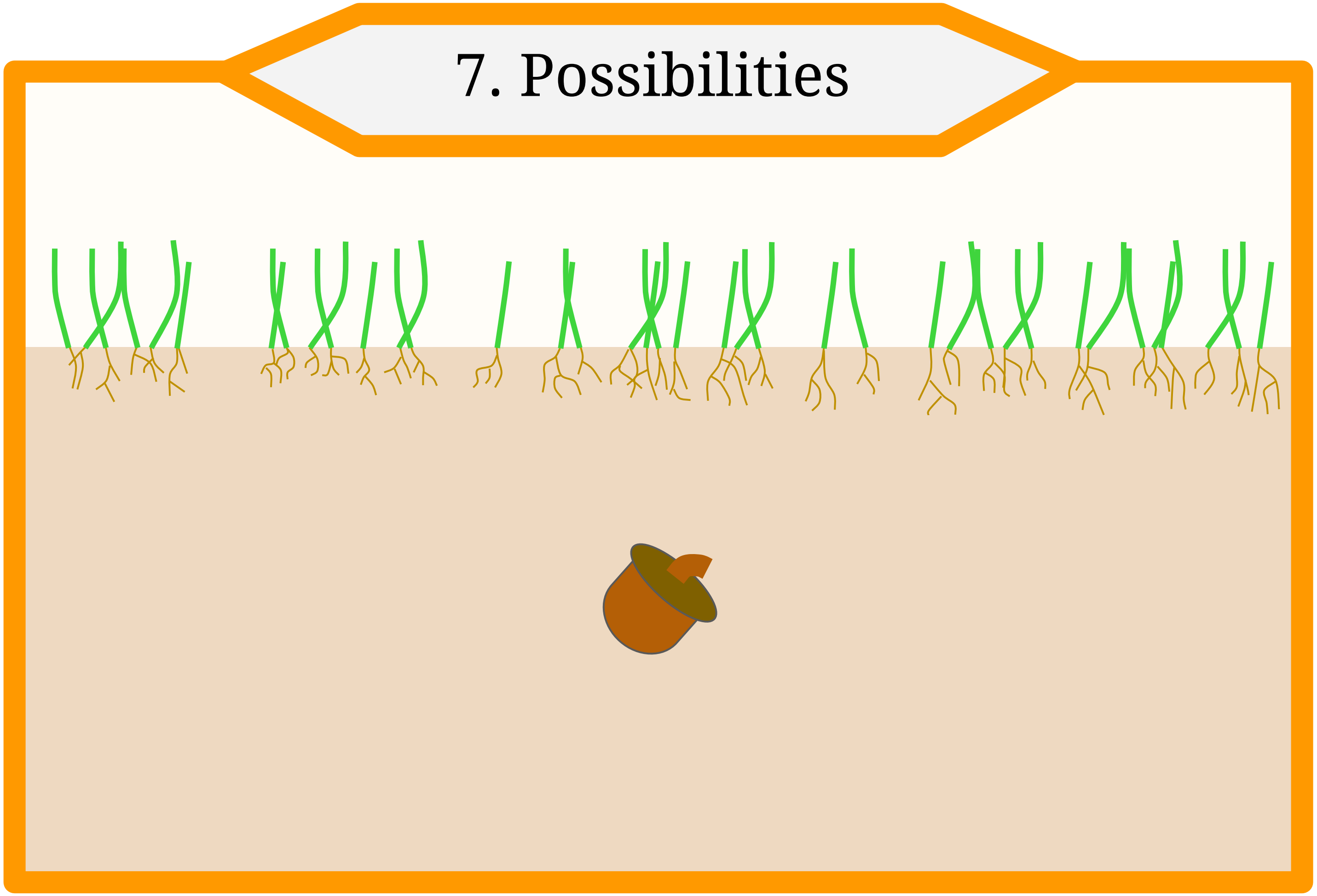
- high curvature in convex G ←→ high marginal value of information
- to increase substitutability, put curvature close to the prior

“Universal” substitutes?

- only with a “trivial-ish” structure

“Universal” complements?

- Yes! XOR of bits


7. Possibilities

Open problems (small selection)

Game theory

- selling information
- signalling
- bundling complements
- other useful applications?

Algorithms

- check if signals are strong substitutes
- compute Alice’s best response in stage one (decompose signal into sub. and comp. components)
- SignalSelection on discrete or continuous lattices!

Structure

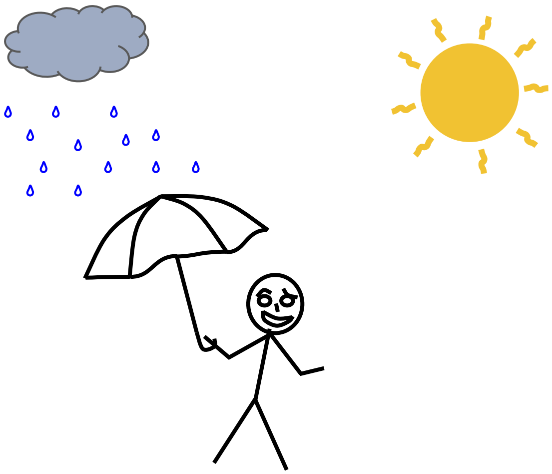
- examples of (classes of) subs and comps
- “more substitutable” signal structures? utilities?
- “universal” substitutes and complements
- connections: e.g. sensitivity of Boolean functions?

Resources

These slides: bowaggoner.com/

Blog posts on proper scoring rules, generalized entropies, ... bowaggoner.com/blog/

Information elicitation tutorial: sites.google.com/site/informationelicitation/



Thanks!