AI and behavioral economics

- Discovering variables
- Human judgment overfits

Colin Camerer, Caltech

Robert Kirby Prof of Behavioral Economics
Director, T&C Chen Center for Social and Decision Neuroscience

Discovering variables

- Behavioral economics definition
  - include natural limits of computation, willpower and selfishness
- University-structure definition
  - Borrows from neighboring sciences
    - psychophysics (prospect theory), norms (sociology), sociality (psych, anthropology), self control (neuro)

Discovering variables

- Search for predictive variables definition
  - Behavioral economics is open-minded
  - Defaults
    - Reminders
    - Social comparison
    - Cognitive skill
    - anxiety
    - Habit
  - “Nudge” experiments explore this space

Here comes ML

- ML allows exploration of many variables
  - Can give upper bound to how well theory *could* do--complete (Kleinberg et al 2017) or clairvoyant (economic value; Camerer et al QJE 2004)
  - Can discover new variables
- Two examples:
  - Predicting initial play in 3x3 matrix games *(bound)*
  - Semi-structured bargaining *(new)*

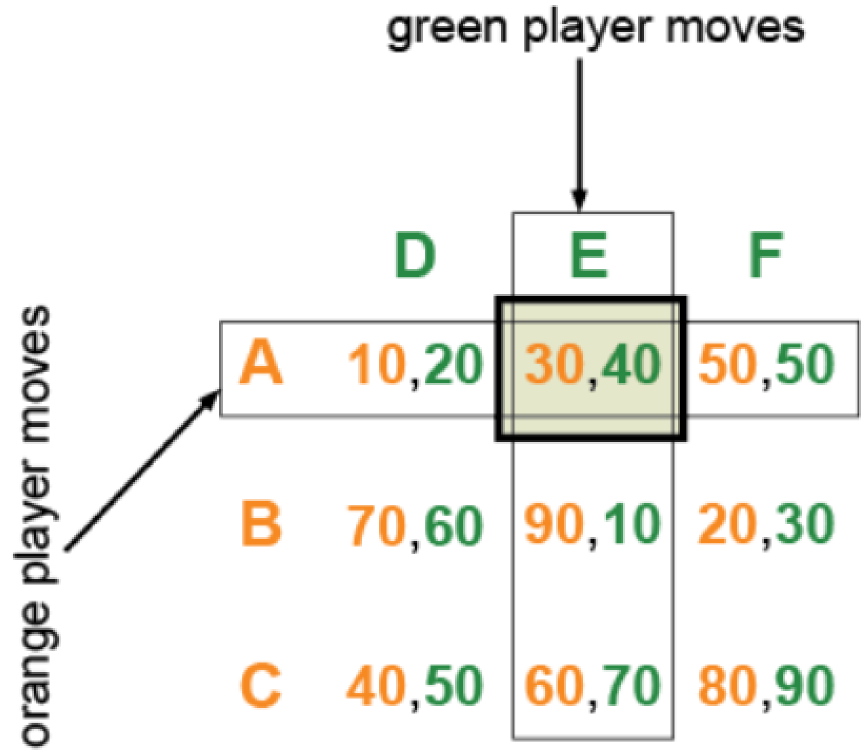
Theory value as % of "clairvoyant" maximum (Camerer Ho Chong QJE 04)

TABLE VIII
ECONOMIC VALUE OF VARIOUS THEORIES

Data set	Stahl and Wilson	Cooper and Van Huyck	Costa-Gomes et al.	Mixed	Entry
Observed payoff	195	586	264	328	118
Clairvoyance payoff	243	664	306	708	176
<u>Economic value</u>					
Clairvoyance	48	78	42	380	58
Cognitive hierarchy (Common τ)	13	55	22	132	10
Nash equilibrium	5	30	15	−17	2
<u>% Maximum economic value achieved</u>					
Cognitive hierarchy (Common τ)	26%	71%	52%	35%	17%
Nash equilibrium	10%	39%	35%	−4%	3%

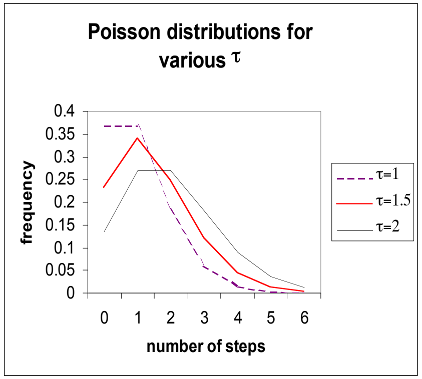
Ex 1: Initial play in 3x3 games

(Fudenberg, Liang 2017; cf. Hartford, Wright, Leyton-Brown 2016)



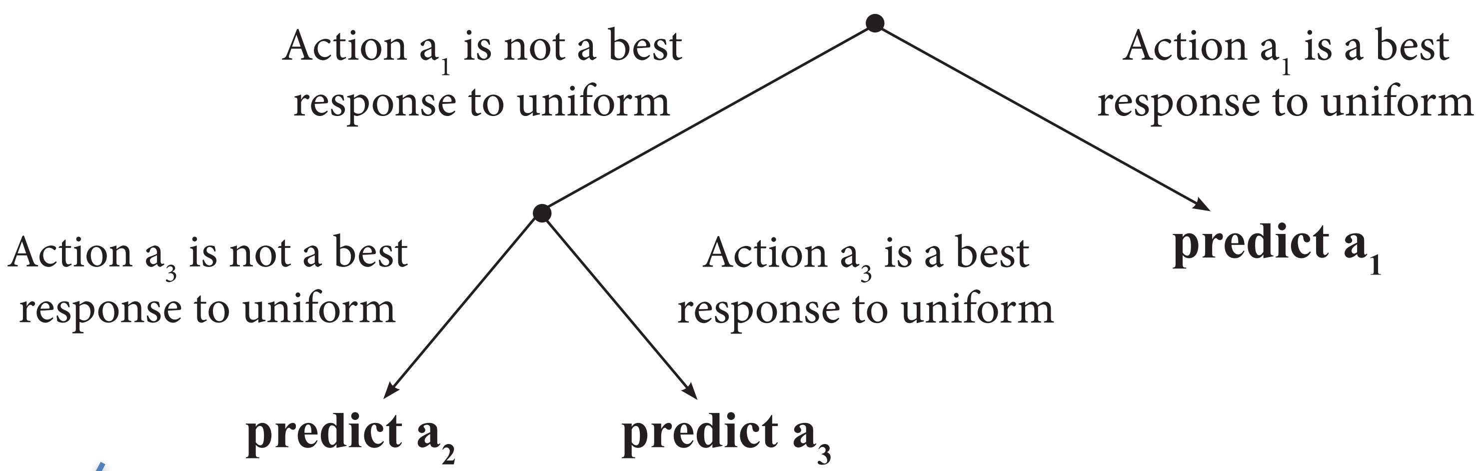
Poisson CH
(Camerer+ QJE 2004)

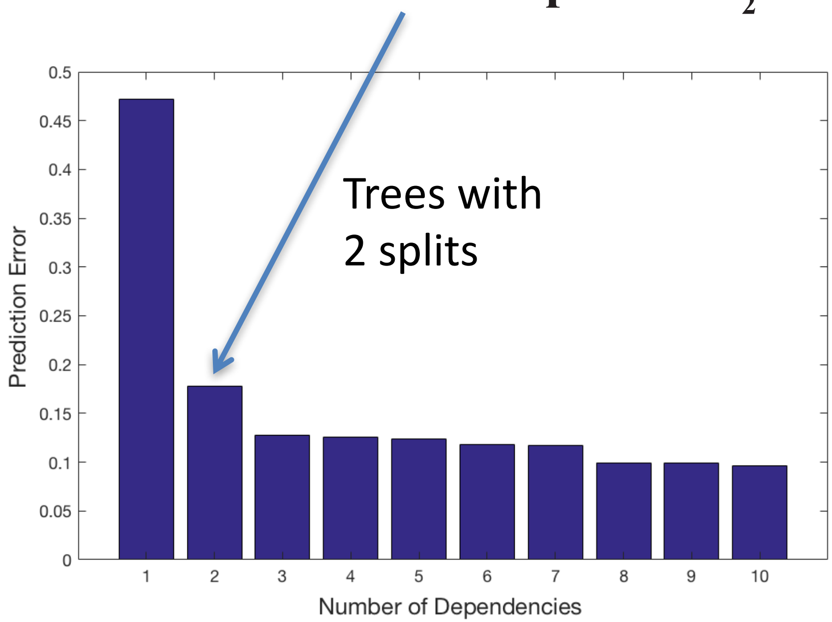
$$P_k(a_i) = \sum_{h=0}^{k-1} \frac{\pi_\tau(h)}{\sum_{h=0}^{k-1} \pi_\tau(h)} P_h(a_i)$$



ML
(88 features)

- *Maximizing total payoffs:* Indicator for whether there exists an action $a_2 \in A_{\text{col}}$ such that
$$u_1(a_1, a_2) + u_2(a_1, a_2) = \max_{a \in A}(u_1(a) + u_2(a)).$$
- *Max-max:* Indicator for whether the row player would choose a_1 if he could also choose the column player's action; that is, whether there exists some action $a_2 \in A_{\text{col}}$ such that
$$(a_1, a_2) \in \operatorname{argmax}_{a \in A} u_1(a).$$
- *Max-min:* Indicator for whether action a_1 maximizes the lowest possible payoff the row player might obtain; that is, whether
$$a_1 \in \operatorname{argmax}_{a_1' \in A_{\text{row}}} \min_{a_2 \in A_{\text{col}}} u_1(a_1', a_2).$$


Action a_1 is not a best response to uniform
Action a_1 is a best response to uniform
predict a_1
Action a_3 is not a best response to uniform
Action a_3 is a best response to uniform
predict a_2
predict a_3


Trees with 2 splits
Prediction Error
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
4
5
6
7
8
9
10
Number of Dependencies

	Error	Completeness
Naive Benchmark	0.6667	0
Uniform Nash	0.5507	33.66%
	(0.0055)	
Poisson Cognitive Hierarchy Model	0.3838	82.02%
	(0.0197)	
Prediction rule based on game features	0.3360	95.88%
	(0.0056)	
"Best possible"	0.3218	1

Table 3: Predicting the realized action in play of lab games

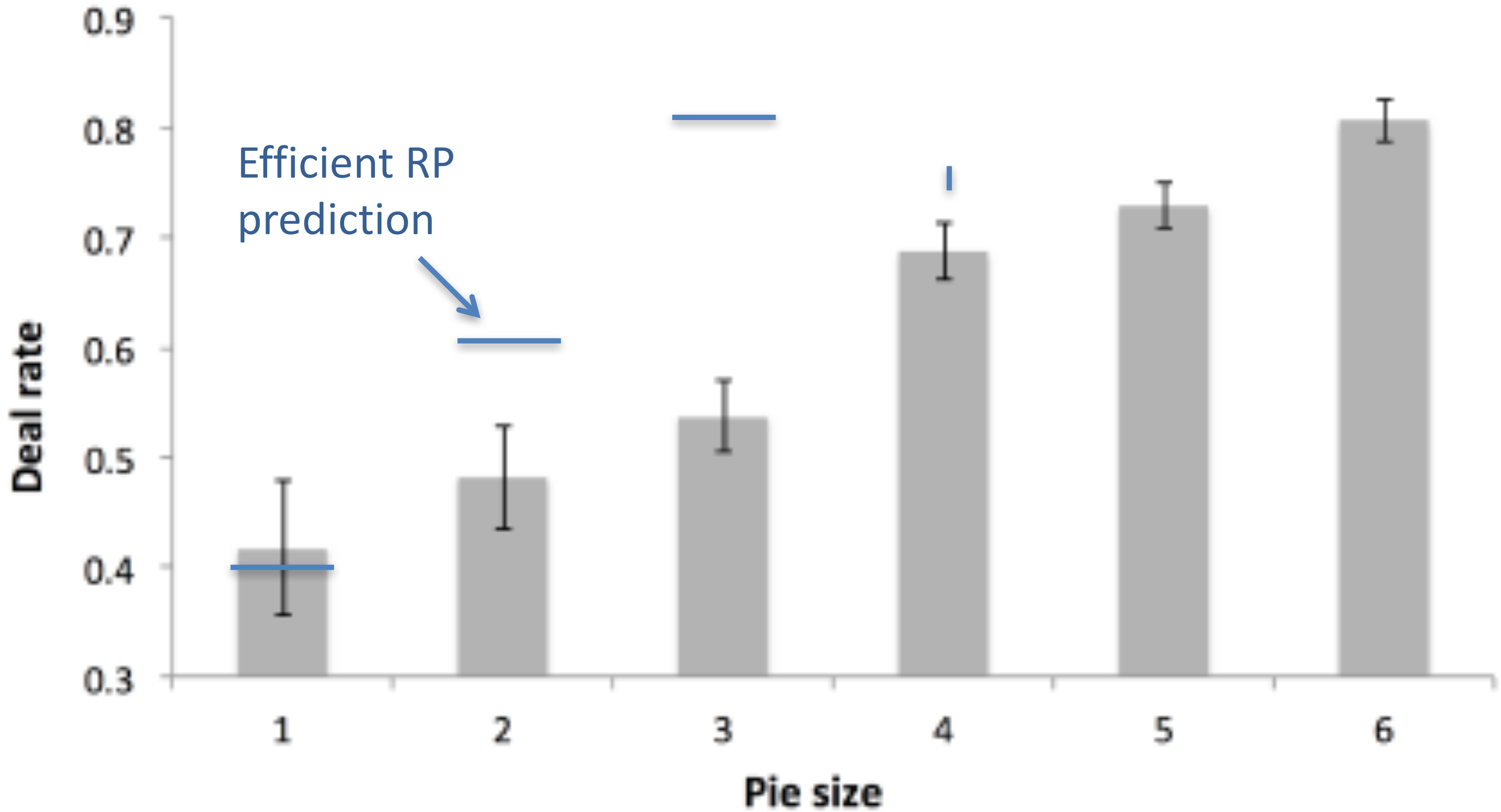
	Error	Completeness
Naive Benchmark	0.6667	0
PCHM	0.3838	82.02%
	(0.0197)	
PCHM with Risk Aversion	0.3531	90.92%
	(0.0133)	
Five-Split Decision Tree	0.3556	90.20%
	(0.0062)	
Unrestricted Decision Tree	0.3360	95.88%
	(0.0056)	
"Best possible"	0.3218	1

Table 5: Introduction of risk aversion improves the cognitive hierarchy prediction error.

Ex 2: Semi-structured bargaining with private information (Camerer, Nave, Smith *Mgt Sci* in press)

Figure 2: Deal rates and mean payoffs across pie sizes

(a) Deal rates by pie size



predicting disagreements ROC

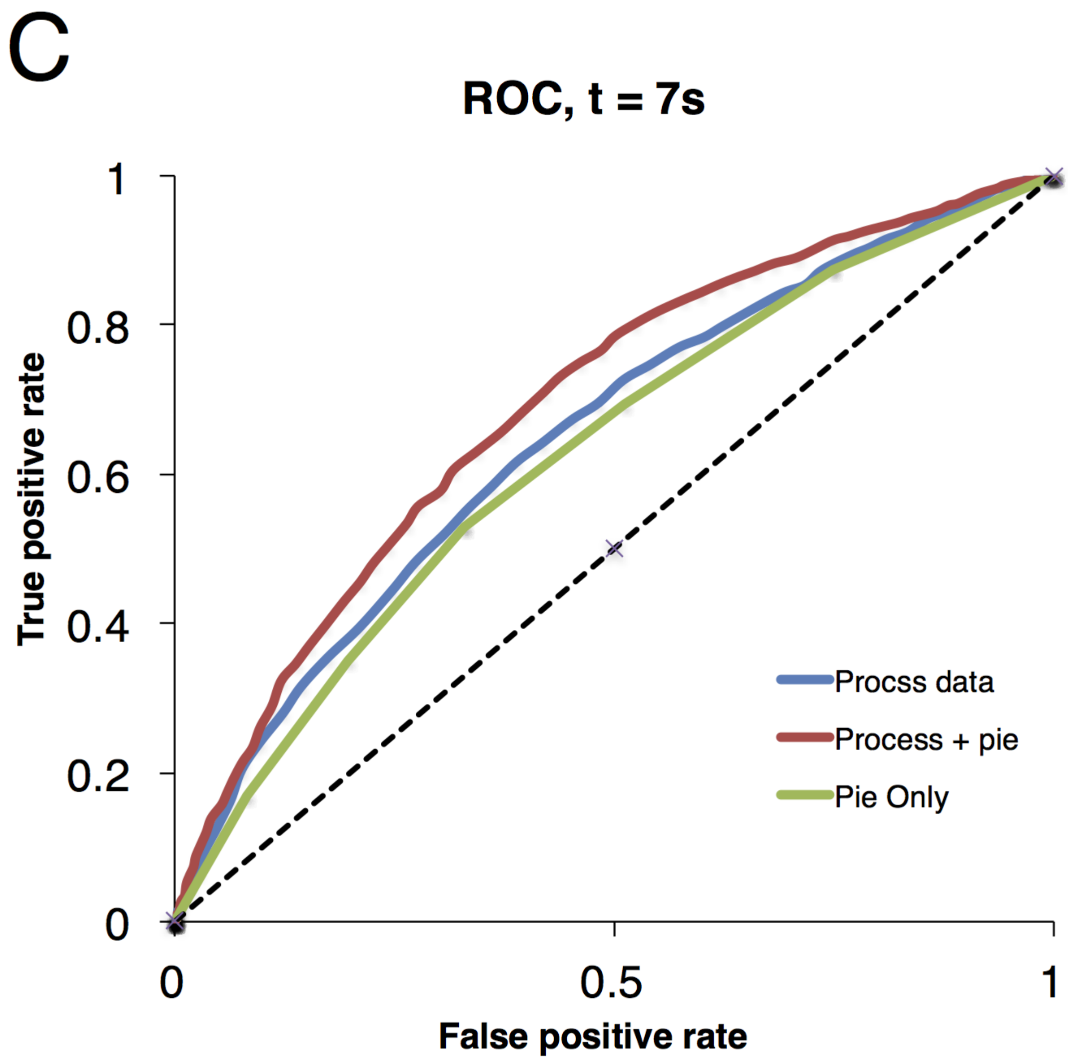
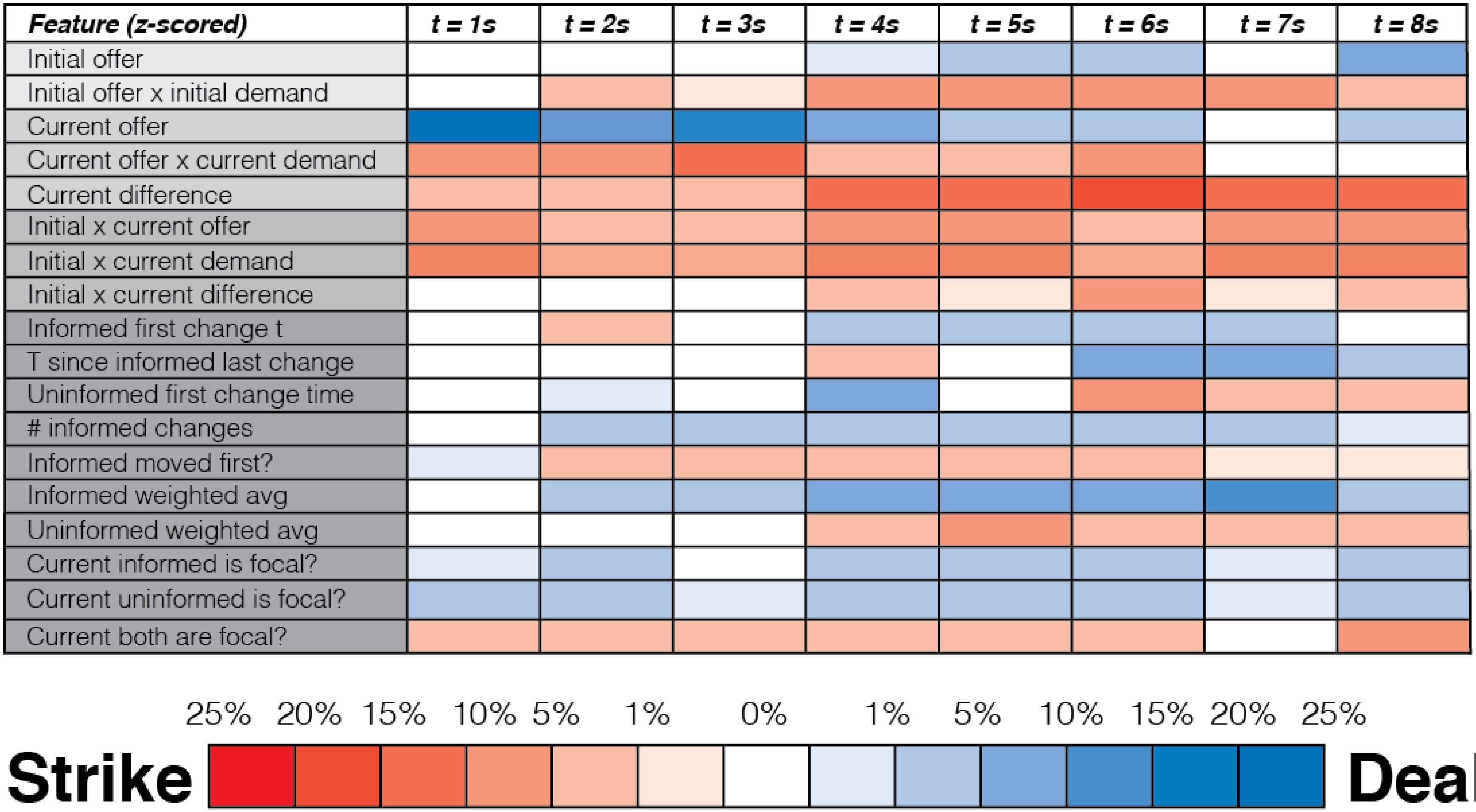
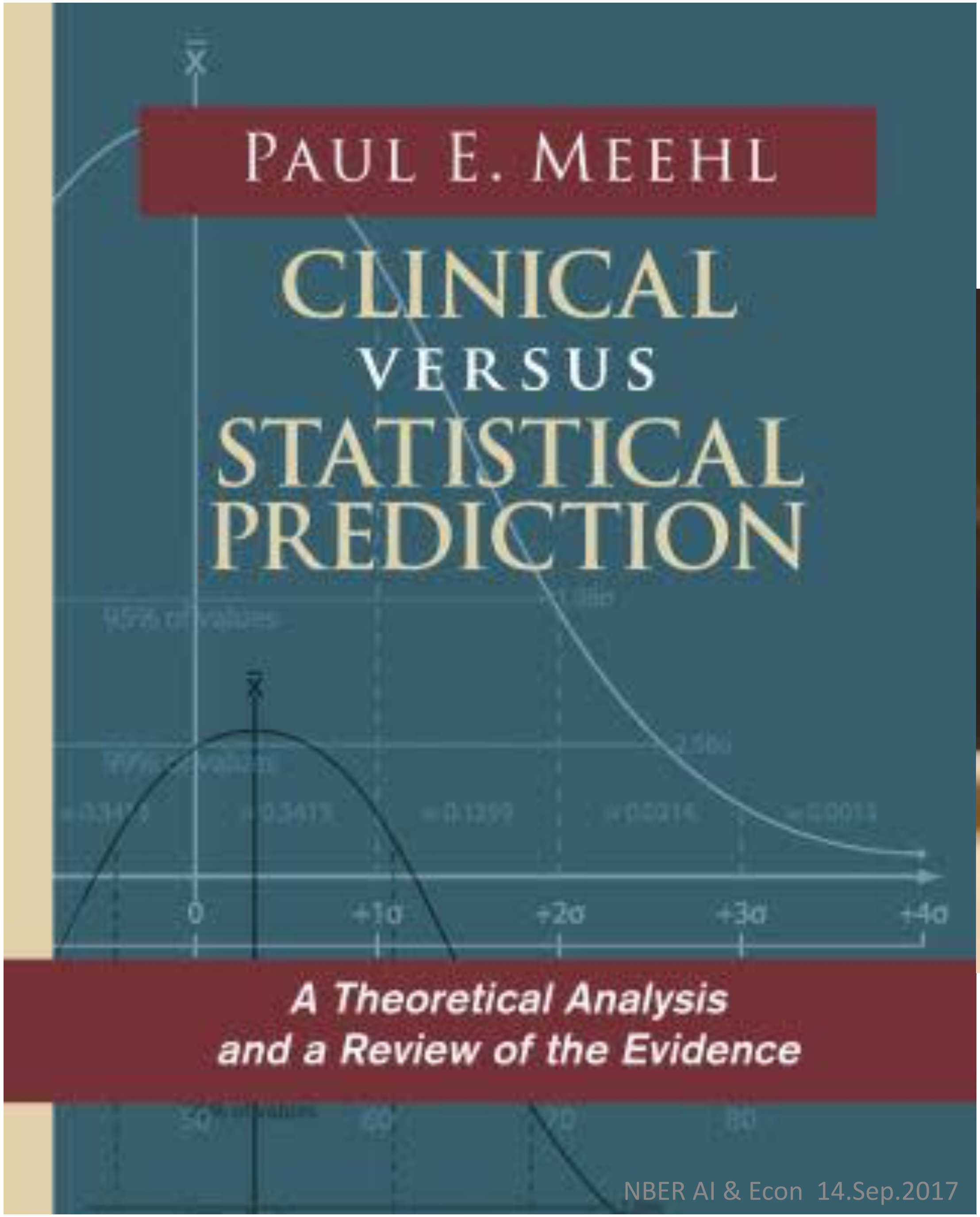


Figure 8: Bargaining process features selected by the classifier for outcome prediction (deal=1) and their estimated marginal effects. (Pie sizes are excluded.)



II: Human and ML prediction

- history
- hypothesis:
  - Some human judgment patterns can be understood as imperfect ML



Paul Meehl 1920-2003
Univ Minnesota

scope of “clinical”

- Psychiatric diagnosis
- Homicidality
- Juvenile delinquency
- Recidivism
- Academic performance
- Graduate PhD admissions

background on "bootstrapping"

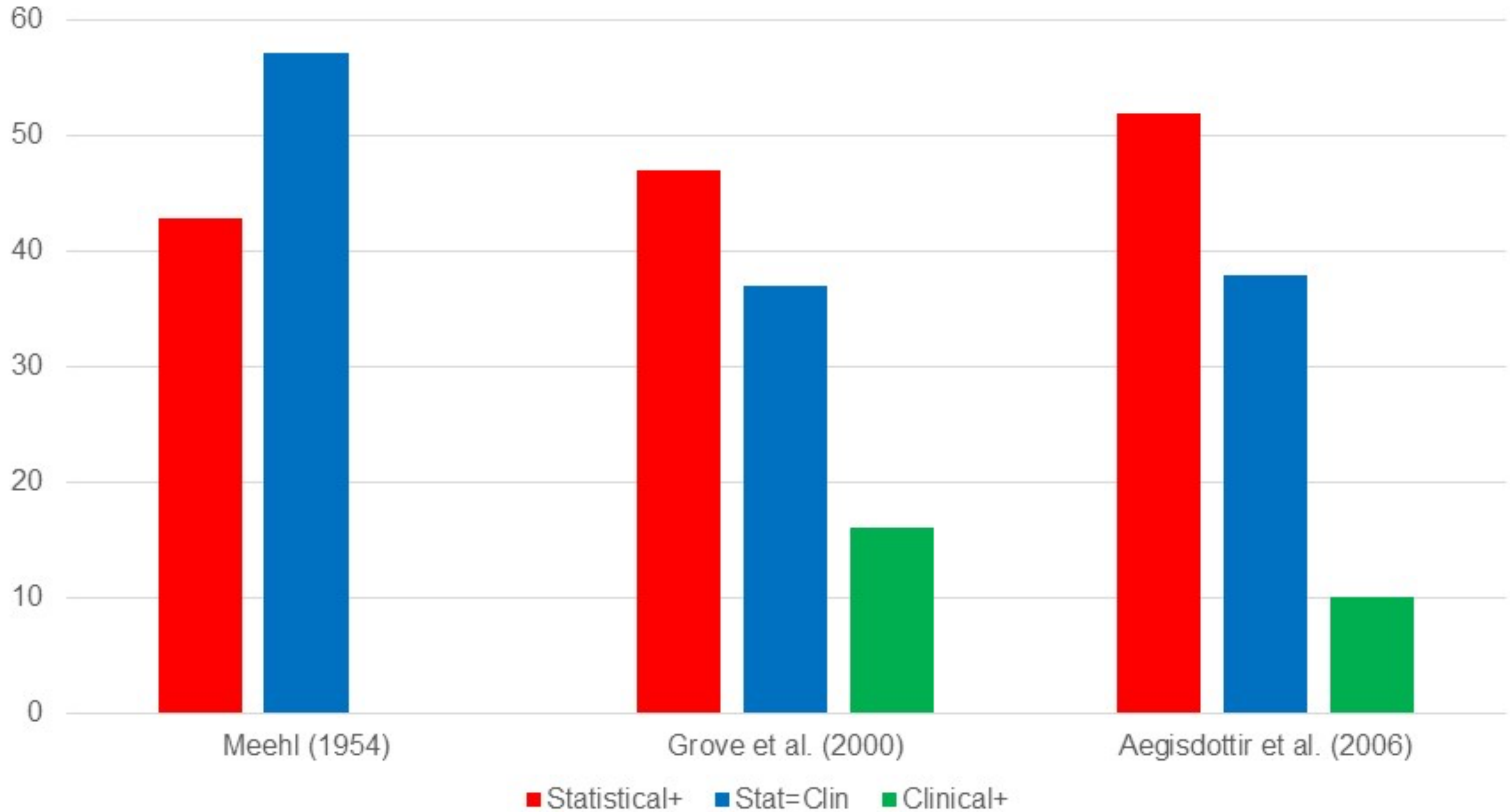
- Meehl (1954):

  "what I expected to be a floor turned out to be a ceiling"

- Unstructured interviews and clinical judgment can be *notoriously* unreliable
  - "Bootstrap" (=fit judgments to X_i, discard ε)
  - $\approx$ 10% > clinical
  - But there is *some* reliable intuition (omitted variables) in bootstrap residuals $\approx 1/3$ of $\sigma^2(\varepsilon)$ (Camerer unpub'd thesis `81; compare test-retest with bootstrap)

Clinical vs. Statistical Prediction

Results of 3 Meta-Analyses



Typical effect size -.15 (***no*** subsamples>0)

Grove Psych Assess. 2000
Ægisdóttir Counseling Psychologist 06

History of skepticism

- Strong bias ***against*** statistical >> clinical 1954-20??
  - almost no traction (except: bank credit scoring)
  - Why?
    - Clinicians thought to have ‘intuition’
      - Interactions
      - “broken leg cues” (rare, highly diagnostic)
  - “the question of whether the actuarial approach is superior to the clinical is tantamount to asking whether the sperm is more important than the ovum” (Zubin, 1956, p627)
  - small training sets

History of skepticism (cont'd)

- sporadic, informal discussions of
  - selective labelling (eg Dawes '79 PhD admissions)
  - decision→ payoff
    - what is clinician's objective function?*
- now: Large training sets → ML reproduces possible 'intuition' well
  - Interactions
  - Broken-leg cues

*cf. Einhorn, JPersAssess 86

properties of human judgment

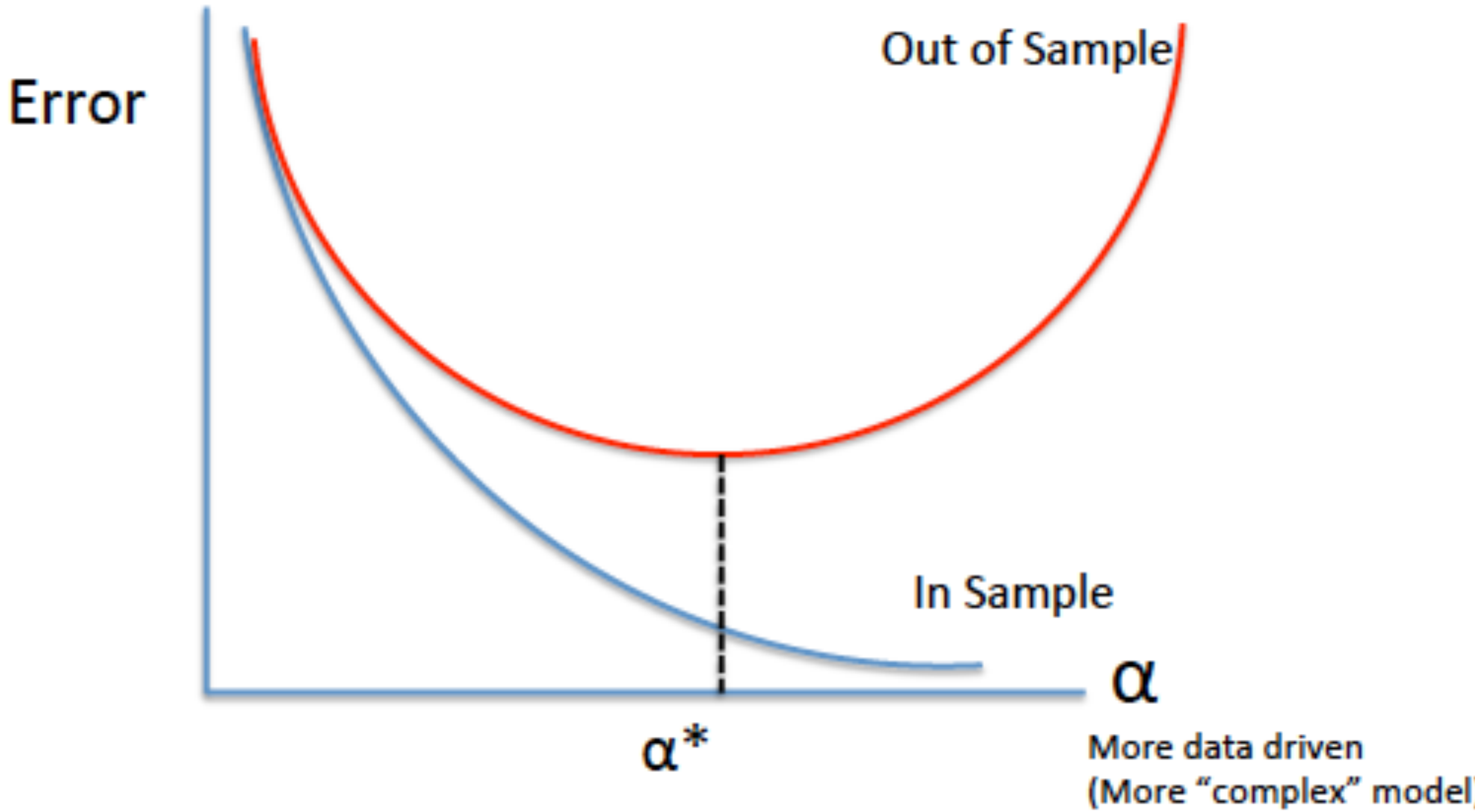
- we do not intuitively accept sparsity
  - (sex – fights)/wk and marital satisfaction r=.40-.81 (Dawes 1979)
  - (GRE+quality+GPA) and PhD success r=.48 (Dawes 1971)
  - (HS) $\cap$ (steady job) $\cap$ (no baby unwed) = no poverty (Jencks)

we do not like sparsity (cont'd)

- Obsession with personal interviews
  (e.g. ASSA hotel meetings)
- Outside >> "inside" view (Kahneman, Lovallo *Mgt Sci* 1993)
- Clustering >> each case unique
- …outside view throws away information

overconfidence and overfitting

- Humans: prediction CIs are too narrow
- ML: Overfitted prediction CIs are too narrow
  (i.e., degraded fit in test/holdout samples)
- Humans: more information increases confidence, not predictive accuracy
  - Clinical accuracy 26-28% (chance=20%)
    confidence 33-53% (Oskamp 1965)


Error
Out of Sample
In Sample
α
α*
More data driven
(More "complex" model)

conclusion

- ML can help discover new “behavioral” variables
- Properties of human prediction could be understood as mistaken machine learning
  - not enough sparsity (regularization)
  - do not correct for overfitting → overconfidence

pro-ML

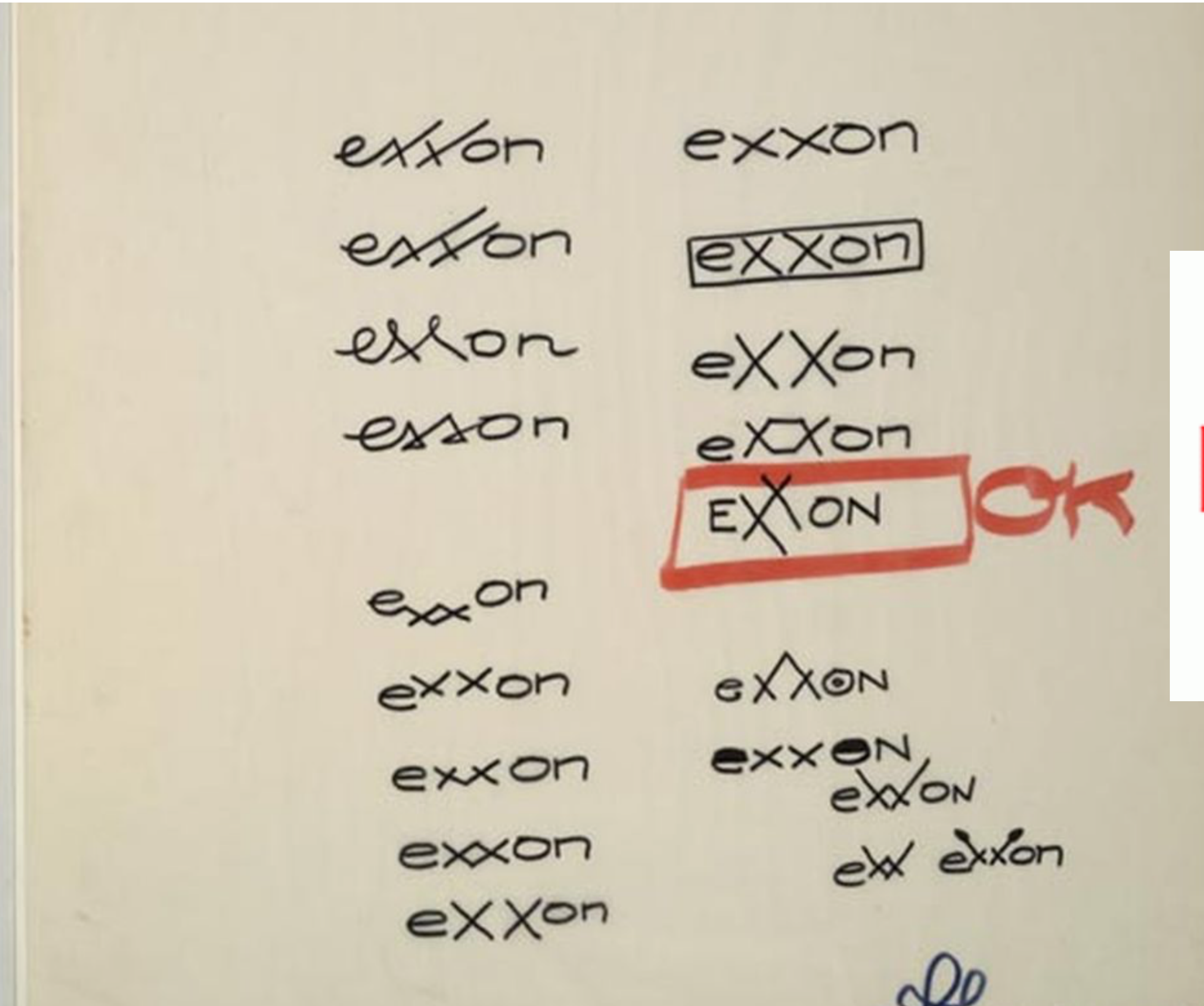
- ML training sets will grow and grow
  - Can self-play around the clock
- Individual- level "human training sets" are constrained by:
  - Genes
  - density of life experience
  - scope of life experience
  - Ability to learn from text, vicarious experience

pro-human

- Human cultural accumulation
- Wisdom of crowds and division of labor
  - ‘group IQ’ can be > max IQ_i
- Cross-domain generalization
  - ML: AlphaGo NN does not inform playing chess
- Wisdom accumulates during a lifetime
  - meta-cognition, dimension reduction (better ideas, more quickly)
- *Can ML do these too??*

Can ML be as *creative* as humans?

- Typical model (e.g. Campbell 1960):
  - large variation of ideas, somehow select the good ones (MAYA)
    - product design, writing sentences, novel plots, music



http://www.theverge.com/2013/11/5/5068132/raymond-loewy-the-man-who-designed-everything