Fast, Frugal and Focused:

When less information leads to better decisions

Gregory Wheeler

Munich Center for Mathematical Philosophy
Ludwig Maximilians University of Munich

Konstantinos Katsikopoulos

Adaptive Behavior and Cognition Group
Max Planck Institute for Human Development

MCMP Colloquium, June 25, 2014

the total evidence norm

Rational Choice

Naive Bayes

Neural Networks

Linear Regression

Dynamic Programming

*...in most situations we might as well **throw away** our information and toss a coin.*

- Richard Bellman

the total evidence norm

Naive Bayes

Neural Networks

Linear Regression

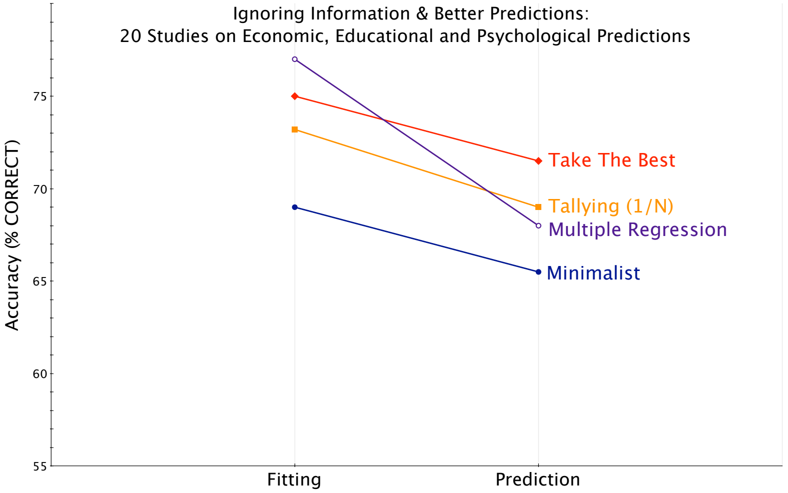
Rational Choice

Dynamic Programming

Bounded Rationality

*...in most situations we might as well **throw away** our information and toss a coin.*

- Richard Bellman



Czerlinski, Gigerenzer, & Goldstein (1999)

heuristic structure and strategic biases

Take-the-Best

(Gigerenzer & Goldstein 1996)

Tallying $(1/N)$

(Dawes 1979)

Search Rule: Look up cues in random order

Stopping Rule: After m $(1 < m \leq N)$ cues, stop the search.

Decision Rule: Predict that the alternative with the higher number of positive cue values has the higher criterion value.

Bias: *ignore weights*

heuristic structure and strategic biases

Take-the-Best

(Gigerenzer & Goldstein 1996)

Search Rule: Look up the cue with the highest validity

Stopping Rule: If cue values differ $(+/-)$, stop search. If not, look up next cue.

Decision Rule: Predict that the alternative with the positive cue value has the higher criterion value.

Bias: *ignore cues*

Tallying (1/*N*)

(Dawes 1979)

Search Rule: Look up cues in random order

Stopping Rule: After m $(1 < m \leq N)$ cues, stop the search.

Decision Rule: Predict that the alternative with the higher number of positive cue values has the higher criterion value.

Bias: *ignore weights*

outline

- 1^{st} Result
- A Puzzle
- 2^{nd} Result (Puzzle Solved!)
- Coherentism and Heuristics

decision task and setup

Forced choice paired comparison task

Decide which of two alternatives, A and B, has the larger value on some numerical criterion, C, given their values on n cues $X_1, \ldots, X_n$.

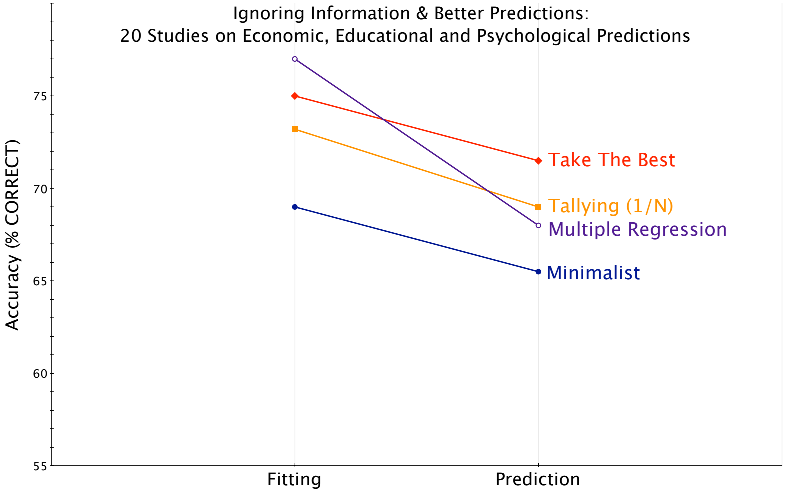
decision task and setup

Forced choice paired comparison task

Decide which of two alternatives, A and B, has the larger value on some numerical criterion, C, given their values on n cues $X_1, \ldots, X_n$.

Perfect Discrimination Assumption

Each cue discriminates among the alternatives.



Czerlinski, Gigerenzer, & Goldstein (1999)

accuracy as a function of size of training sample

Leave-one-out Cross Validation

- There are $n + 1$ inferences to be made in population
  - Training sample: n
  - Test sample: 1

accuracy as a function of size of training sample

Leave-one-out Cross Validation

- There are $n + 1$ inferences to be made in population
  - Training sample: n
  - Test sample: 1
- Cross-validation is repeated $n + 1$ times:
  - Each 1 of $n + 1$ inferences comprises the test sample once.

accuracy as a function of size of training sample

Leave-one-out Cross Validation

- There are $n + 1$ inferences to be made in population
  - Training sample: n
  - Test sample: 1
- Cross-validation is repeated $n + 1$ times:
  - Each 1 of $n + 1$ inferences comprises the test sample once.
- Labeling which cues do best
  - v^* maximum cue validity in $n + 1$ trials; (X^* is that cue).
  - v^{**} second maximum cue validity.

accuracy as a function of size of training sample

Leave-one-out Cross Validation

- There are $n + 1$ inferences to be made in population
  - Training sample: n
  - Test sample: 1
- Cross-validation is repeated $n + 1$ times:
  - Each 1 of $n + 1$ inferences comprises the test sample once.
- Labeling which cues do best
  - v^* maximum cue validity in $n + 1$ trials; (X^* is that cue).
  - v^{**} second maximum cue validity.

Cue covariation

ρ is covariation between cues X^* and X^{**}:

$$\Pr(X^* \wedge X^{**} \text{are correct on trial } t) - \Pr(X^* \text{is correct on trial } t) \cdot \Pr(X^{**} \text{is correct on trial } t)$$

α: single-cue predictive accuracy measured by leave-one-out validation

- Size of training sample → $\frac{1}{n+1}$

$$\alpha = \frac{1}{2}\left(v^* \cdot \left(1 + v^* - \frac{1}{n+1}\right) + \rho\right)$$

α: single-cue predictive accuracy measured by leave-one-out validation

- Size of training sample

$$\alpha = \frac{1}{2}\left(v^* \cdot \left(1 + v^* - \frac{1}{n+1}\right) + \rho\right)$$

- Cue Covariation

α: single-cue predictive accuracy measured by leave-one-out validation

- Size of training sample

$$\alpha = \frac{1}{2}\left(v^* \cdot \left(1 + v^* - \frac{1}{n+1}\right) + \rho\right)$$

- Cue Covariation
- v^*: maximum cue validity in population of $n+1$ trials

α: single-cue predictive accuracy measured by leave-one-out validation

- Size of training sample

$$\alpha = \frac{1}{2}\left(v^* \cdot \left(1 + v^* - \frac{1}{n+1}\right) + \rho\right)$$

- Cue Covariation
- v^*: maximum cue validity in population of $n+1$ trials
- Assumptions:

α: single-cue predictive accuracy measured by leave-one-out validation

- Size of training sample

$$\alpha = \frac{1}{2}\left(v^* \cdot \left(1 + v^* - \frac{1}{n+1}\right) + \rho\right)$$

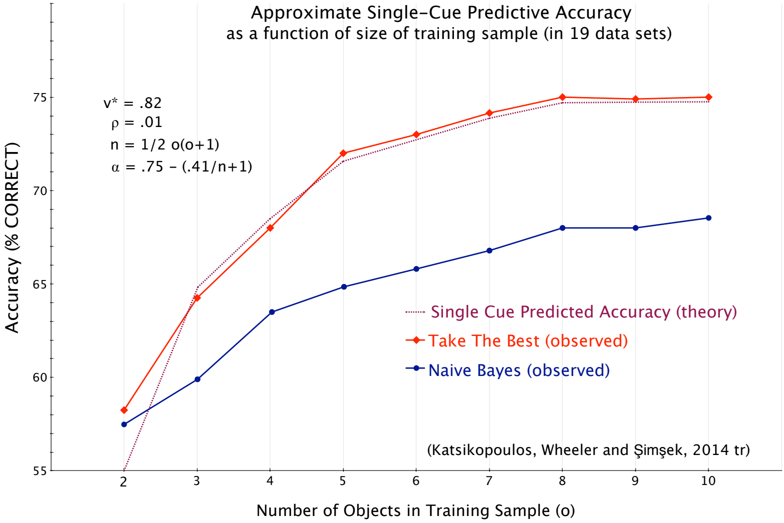
- Cue Covariation
- v^*: maximum cue validity in population of $n+1$ trials
- Assumptions:
    - Perfect Discrimination Assumption

α: single-cue predictive accuracy measured by leave-one-out validation

- Size of training sample

$$\alpha = \frac{1}{2}\left(v^* \cdot \left(1 + v^* - \frac{1}{n+1}\right) + \rho\right)$$

- Cue Covariation
- v^*: maximum cue validity in population of $n+1$ trials
- Assumptions:
  - Perfect Discrimination Assumption
  - when $v^* - v^{**} = \dfrac{1}{n+1}$ and v^* otherwise.


Approximate Single–Cue Predictive Accuracy
as a function of size of training sample (in 19 data sets)
v* = .82
ρ = .01
n = 1/2 o(o+1)
α = .75 – (.41/n+1)
Accuracy (% CORRECT)
75
70
65
60
55
2
3
4
5
6
7
8
9
10
Single Cue Predicted Accuracy (theory)
Take The Best (observed)
Naive Bayes (observed)
(Katsikopoulos, Wheeler and Şimşek, 2014 tr)
Number of Objects in Training Sample (o)

single variable decision rules

A Brunswikian Question

Under what *environmental conditions* do "single reason" rules perform well?

- high ρ?
- low ρ?
- some other structural feature?

Egon Brunswik

single variable decision rules

A Brunswikian Question

Under what environmental conditions do “single reason” rules perform well?

single variable decision rules

A Brunswikian Question

Under what environmental conditions do "single reason" rules perform well?

- Cues are highly intercorrelated (Hogarth & Karelaia 2005)
  - average pairwise cue correlation $\rho_{X_i X_j}$
- Cues are independent (Baucells, Carrasco & Hogarth 2008)
- Cues are conditionally independent (Katsikopoulos & Martignon 2006)

single variable decision rules

A Brunswikian Question

Under what environmental conditions do "single reason" rules perform well?

- Cues are highly intercorrelated (Hogarth & Karelaia 2005)
  - average pairwise cue correlation $\rho_{X_i X_j}$
- Cues are independent (Baucells, Carrasco & Hogarth 2008)
- Cues are conditionally independent (Katsikopoulos & Martignon 2006)

central idea of focused correlation

$$Cov[X_1, \ldots, X_n \mid C = c] - Cov[X_1, \ldots, X_n]$$

central idea of focused correlation

$$\exp(Cov[X_1, \ldots, X_n \mid C = c] - Cov[X_1, \ldots, X_n])$$

central idea of focused correlation

$$\exp(Cov[X_1, \ldots, X_n \mid C = c] - Cov[X_1, \ldots, X_n])$$

&

Let all RVs be indicator functions

central idea of focused correlation

$$\exp(Cov[X_1, \ldots, X_n \mid C = c] - Cov[X_1, \ldots, X_n])$$

&

Let all RVs be indicator functions

$$For_c(x_1, \ldots, x_n) := \frac{\frac{\Pr(x_1, \ldots, x_n \mid c)}{\Pr(x_1 \mid c) \cdots \Pr(x_n \mid c)}}{\frac{\Pr(x_1, \ldots, x_n)}{\Pr(x_1) \cdots \Pr(x_n)}}$$

single-cue accuracy as a function of criterion predictability and focused correlation

- Criterion predictability

$$v_1 = \sum_{c, x_2, \ldots, x_k} \frac{\Pr(C = c \mid X_1 = c, X_2 = x_2, \ldots, X_k = x_k)}{FOR_C(X_1 = c, X_2 = x_2, \ldots, X_k = x_k)} \cdot \Pr(X_1 = c) \cdot \Pr(X_2 = x_2) \cdots \Pr(X_k = x_k)$$

single-cue accuracy as a function of criterion predictability and focused correlation

- Criterion predictability →

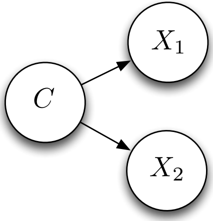
$$v_1 = \sum_{c, x_2, \ldots, x_k} \frac{\Pr(C = c \mid X_1 = c, X_2 = x_2, \ldots, X_k = x_k)}{FOR_C(X_1 = c, X_2 = x_2, \ldots, X_k = x_k)} \cdot \Pr(X_1 = c) \cdot \Pr(X_2 = x_2) \cdots \Pr(X_k = x_k)$$

- Result: single cue accuracy increases when the ratio of criterion predictability to focused cue correlation increases

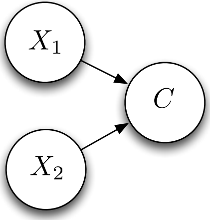
solving the puzzle

Cues should be dependent but conditionally independent given the criterion

Cues should be independent but conditionally dependent given the criterion



$X_1 \not\perp X_2$
$X_1 \perp X_2 \mid C$

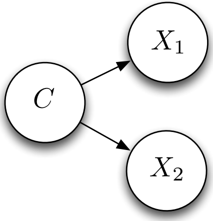


$X_1 \perp X_2$
$X_1 \not\perp X_2 \mid C$

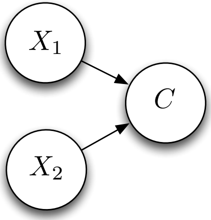
solving the puzzle

Result 2

$$v_1 = \sum_{c, x_2, \ldots, c_k} \frac{\Pr(C = c \mid X_1 = c, X_2 = x_2, \ldots, X_k = x_k)}{FOR_C(X_1 = c, X_2 = x_2, \ldots, X_k = x_k)} \cdot \Pr(X_1 = c) \cdot \Pr(X_2 = x_2) \cdots \Pr(X_k = x_k)$$



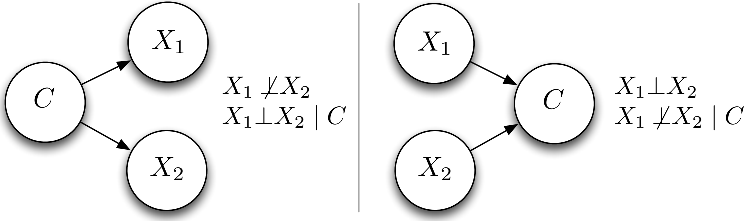
$X_1 \not\perp X_2$
$X_1 \perp X_2 \mid C$



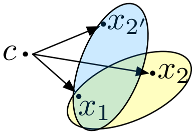
$X_1 \perp X_2$
$X_1 \not\perp X_2 \mid C$

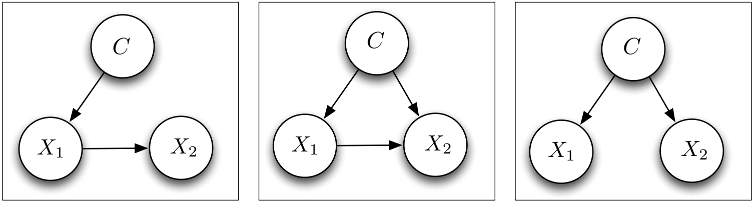
Result 2

$$v_1 = \sum_{c, x_2, \ldots, c_k} \frac{\Pr(C = c \mid X_1 = c, X_2 = x_2, \ldots, X_k = x_k)}{\dfrac{\dfrac{\Pr(x_1, \ldots, x_n \mid c)}{\Pr(x_1 \mid c) \cdots \Pr(x_n \mid c)}}{\dfrac{\Pr(x_1, \ldots, x_n)}{\Pr(x_1) \cdots \Pr(x_n)}}} \cdots \qquad \frac{N}{D}$$



resolving a discontinuity





$$P(X_2) = P(X_2')$$

or

$$\text{(A2)} \begin{bmatrix} P(C|X_1) = P(C|X_2) \\ P(C|X_1) = P(C|X_2') \end{bmatrix}$$

adaptive epistemic norms

- Conditional independence: { lousy for total evidence / good for single cue }
- Robustness of single cue:
  - ✓ cond independent cues
  - ✓ independent cues
  - ✓ deflationary focused corr
  - X inflationary focused corr
- Total evidence coherence: { ✓ inflationary focused corr
- Final Remarks:
  - Coherentism and Heuristics are complementary
  - Adaptive Epistemology

key references

- Baucells, M., JA Carrasco, and R Hogarth (2008): "Cumulative Dominance and Heuristic Performance in Binary Multi-attribute Choice," *Operations Research*, 56:1289–1304.
- Bovens, L. and S. Hartmann (2003). *Bayesian Epistemology*, Oxford Univ Press.
- Olsson, E. (2005). *Against Coherence*, Oxford University Press.
- Hogarth, R. and N. Karelaia (2005). "Ignoring Information in Binary Choice with Continuous Variables: When is less 'more'?" *Journal of Mathematical Psychology*, 49: 115–124.
- Katsikopoulos, K and L Martignon (2006): Naïve Heuristics for Paired Comparison: Some results on their relative accuracy," *Journal of Mathematical Psychology* 50: 488–494.
- Katsikopoulos, K., L. Schooler, and R. Hertwig (2010). "The Robust Beauty of Ordinary Information," *Psychological Review*, 117(4): 1259.
- Schlosshauer, M. and G. Wheeler (2011). "Focused Correlation and the Jigsaw Puzzle of Variable Evidence," *Philosophy of Science*, 78(3): 276–92.
- Wheeler G., and Scheines, R. (2013). "Coherence and Confirmation Through Causation," *Mind*, 122(435): 135-70.
- Wheeler, G. (2009). "Focused Correlation and Confirmation," *The British Journal for the Philosophy of Science*, 60(1): 79–100.
- Wheeler G., (2012). "Explaining the Limits of Olsson's Impossibility Result," *The Southern Journal of Philosophy*, 50(1): 136-50.