NORTH ATLANTIC TREATY ORGANIZATION

SCIENCE AND TECHNOLOGY ORGANIZATION

AC/323()

www.sto.nato.int

STO TECHNICAL REPORT

PUB REF STO-MP-SAS-114-PPI

ANNEX I
Optimal Experimental Design Theory, Asymmetric Cost Structures, and the Value of Information

Jonathan D. Nelson

Optimal Experimental Design Theory, Asymmetric Cost Structures, and the Value of Information

Jonathan D. Nelson

NATO SAS-114 Meeting
Kastellet, Copenhagen, December 6th, 2016

"There is nothing so practical as a good theory". --Lewin, 1951

→ the entropy content in this talk is a preview of *Crupi*, Nelson, Meder, Cevolani, & Tentori (submitted). For questions on it, or if you wish to cite it, please contact Prof. Vincenzo Crupi (vincenzo.crupi@unito.it).

contact: jonathan.d.nelson@gmail.com,
http://jonathandnelson.com/

abc Center for Adaptive Behavior and Cognition

Max-Planck-Institut für Bildungsforschung
50
50 Years of Research on Human Development

Why should the Intelligence Community care...

- ... about theory of what makes an investigation useful?
  - statisticians, mathematicians, and philosophers have thought a lot
  - state of the art performance in many domains (classification trees, image registration, predicting eye movements)
- ... about the psychology of information?
  - current ideas of human psychology are out of date / simplistic / not specific enough to be helpful ("confirmation bias")
  - usually people decide what information to collect or analyze
  - psychology needs to be characterized, to understand discrepancies between human intuition and normative valuation of information

Part 1 of 3: history and state of the art of the math

Finding a useful experiment (test, question)

Domain	Hypotheses	Questions	Answers
Science	Theories	Experiments	Possible results
Categorization	Individual categories	Features to view	Forms of features
Medical diagnosis	Possible diseases	Medical tests	+ / - test results
Intelligence Analysis	J is a terrorist (or not)	Reads terrorist pubs? Plays with explosives?	

- we don't have (and can't get) all the info we need
- but carefully selected experiments (tests, investigations, questions) can help

Background: what makes a question (or experiment) useful?

- many ideas in statistics, since 1950s (Good, Lindley, etc)
- there was no overarching rhyme or reason (bag of tricks)
- the most psychologically plausible ideas had to do with expected reduction in uncertainty (or similar)
  (Nelson, *Psych Rev*, 2005)

Core ideas

NB: knowledge assumptions much stronger than from Jonas's talk

- We want to know $K=\{k_1, k_2, \ldots k_n\}$
- We can observe $D=\{d_1, d_2, \ldots d_m\}$
- We know $P(K \times D)$
- How surprising is it if $K=k_i$?
- How uncertain is K, on average?
- How much would knowing $D=d_j$ reduce uncertainty?
- What is the expected uncertainty reduction if we query D?

	d_1	d_2	…	d_m	Σ
k_1					$P(k_1)$
k_2					$P(k_2)$
…					…
k_n					$P(k_n)$
Σ	$P(d_1)$	$P(d_2)$	…	$P(d_m)$	1

What we could quantify with a measure of uncertainty?

- ecosystem health
- income inequality in a society
- uncertainty about
  - the true category
  - a patient's disease
  - the best scientific hypothesis
- expected information gain of an experiment
  (expected reduction from prior to posterior uncertainty)

What is uncertainty?

(not the plenary smorgasbord from Bjørn Isaksen, but …)

- not knowing for sure
  (Popper-esque)
- the number of possibilities minus 1
  (smells like a heuristic)
- the probability of guessing incorrectly
  (Bayes's error)
- expected surprise
  (handles all of the above, and many more!)

Some (weak) requirements for any entropy function

- definitions:
  - K is a random variable $K = \{k_1, k_2, \dots k_n\}$, where $n \geq 2$
  - ent(K) is the uncertainty about the value that K will take
- we would like an entropy function such that
  - ent(K) ≥ 0
  - if $\max_{\{i=1:n\}} P(k_i) = 1$, then ent($K$) = 0
  - maximal (ties allowed) if $P(k_1) = P(k_2) \dots = P(k_n) = 1/n$, for any n
  - permutation invariant: reordering the P(k_i) does not change ent(K)
  - extensible: addition of zero-probability k_i does not change ent(K)
  - broader than Shannon, Tsallis, Renyi, Arimoto, even Sharma-Mittal

→ the entropy content in this talk is a preview of *<u>Crupi</u>*, Nelson, Meder, Cevolani, & Tentori (submitted). For questions on it, or if you wish to cite it, please contact Prof. Vincenzo Crupi (vincenzo.crupi@unito.it).

Isn't Shannon entropy the correct uncertainty measure?

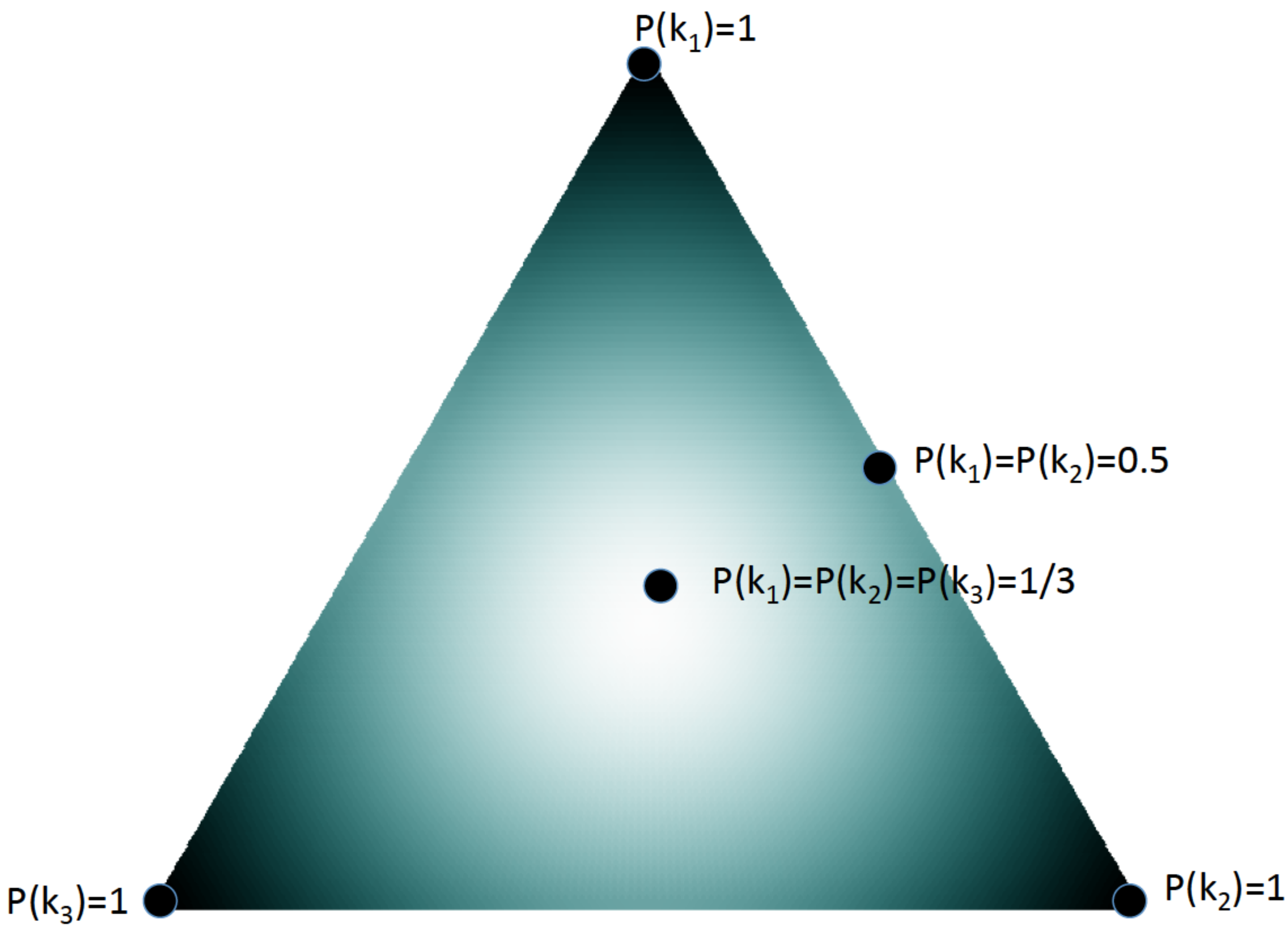
Axiomatic characterizations of entropy also go back to Shannon. In his view, this is "in no way necessary for the theory" but "lends a certain plausibility" to the definition of entropy and related information measures. "The real justification resides" in operational relevance of these measures. --Imre Csiszár (2008)

Entropy as expected surprise

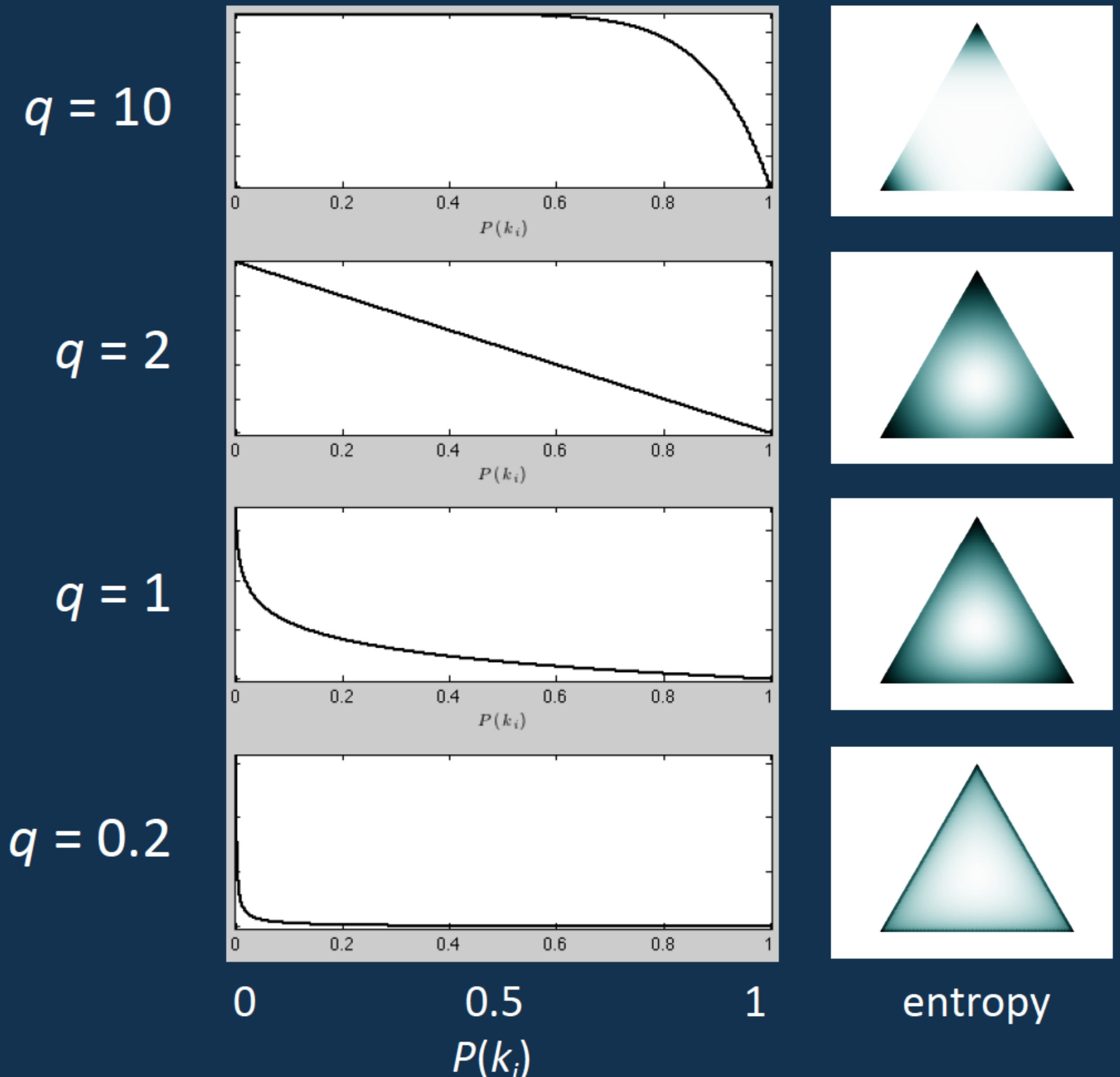
- entropy in *K* is average surprise: $\mathrm{ent}(K) = \sum_{i=1}^{n} [P(k_i)\ \mathrm{surp}(k_i)]$

- then if $\mathrm{surp}(k_i)$ = _____, we get _____ entropy
  - $\mathrm{surp}(k_i)$ = $(1 - P(k_i))$, Quadratic entropy (Gini, 1912)
  - $\mathrm{surp}(k_i)$ = $\ln \frac{1}{P(k_i)}$, Shannon (1948) entropy
  - $\mathrm{surp}(k_i)$ = $\ln_q \frac{1}{P(k_i)}$, Tsallis (1988) entropy

Shannon entropy of K=[k_1, k_2, k_3]. Black=none, white=max



Tsallis surprise and Tsallis entropy, for various degrees *q*:



Rényi (1961) entropy: different expectations of surprise:

- Rényi: instead of averaging the surprise values themselves, use a (magic) function of those surprise values to average them, in the General Theory of Means framework:

$$\mathrm{ent}(K) = \ln\left\{\sum_{i=1}^{n}\left[P(k_i)\ \mathrm{e}^{(1-r)\left(\ln\frac{1}{P(k_i)}\right)}\right]\right\}^{1-r}$$

Tsallis, Rényi, Sharma-Mittal, and Generalized Means

- General theory of means for self-weighted entropies:

$$\mathrm{ent}(K) = g^{-1}\left\{\sum_{i=1}^{n}\left[P(k_i)\, g\big(\mathrm{surp}(k_i)\big)\right]\right\}$$

- Tsallis:
  $g(x)=x$, $\mathrm{surp}(k_i)= \ln_q (1/P(k_i))$

$$\mathrm{ent}(K) = \sum_{i=1}^{n}\left[P(k_i)\ln_q \frac{1}{P(k_i)}\right]$$

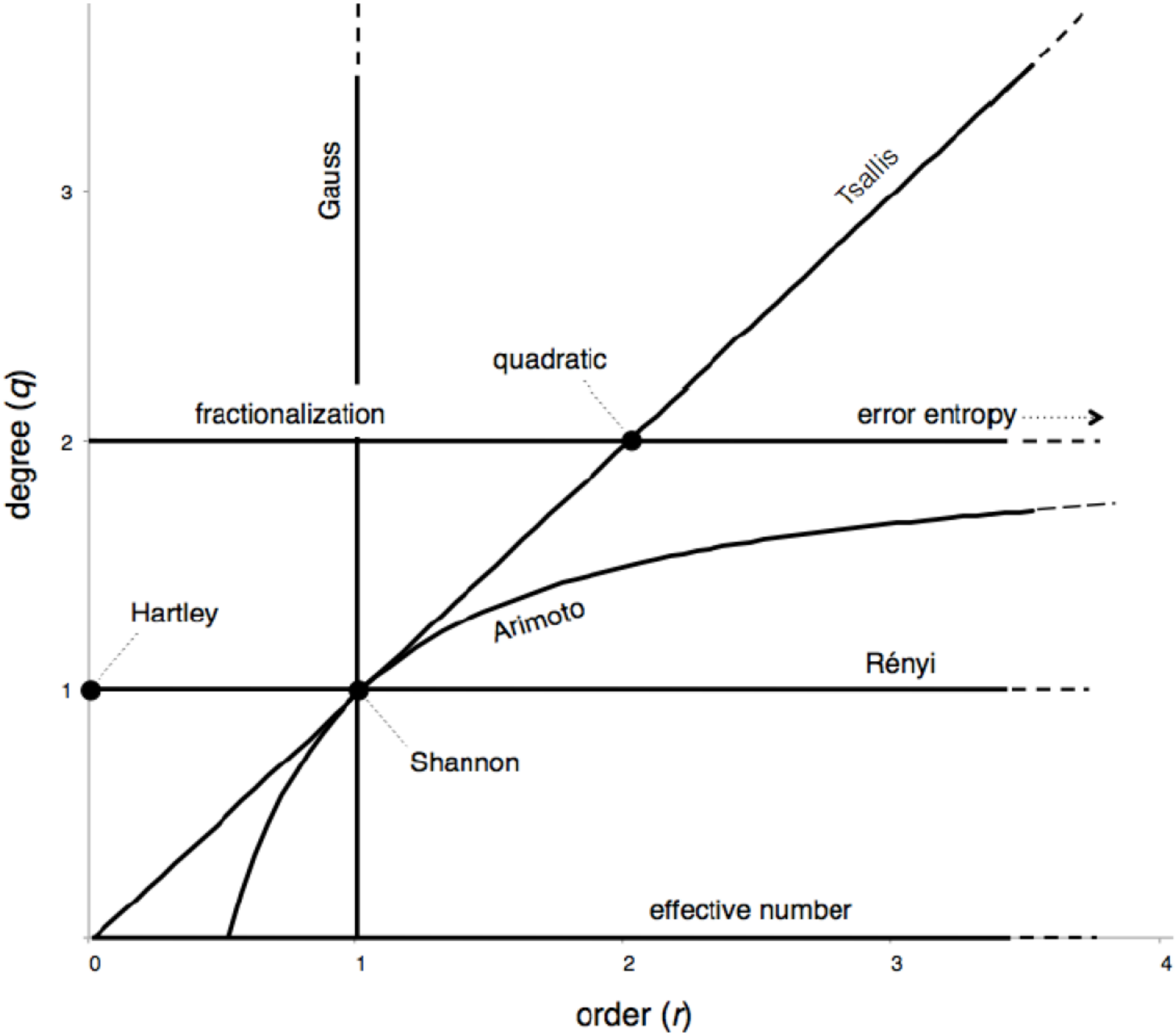
- Rényi:
  $g(x)=e^{(1-r)x}$, $\mathrm{surp}(k_i)= \ln (1/P(k_i))$

$$\mathrm{ent}(K) = \ln\left\{\sum_{i=1}^{n}\left[P(k_i)\, e^{(1-r)\left(\ln\frac{1}{P(k_i)}\right)}\right]\right\}^{1-r}$$

- Sharma-Mittal:
  combine Rényi + Tsallis:
  r is order, *q* is degree
  - set $\mathrm{surp}(k_i) = \ln_q 1/P(k_i)$
  - set $g(x) = \ln_q \exp_r x$

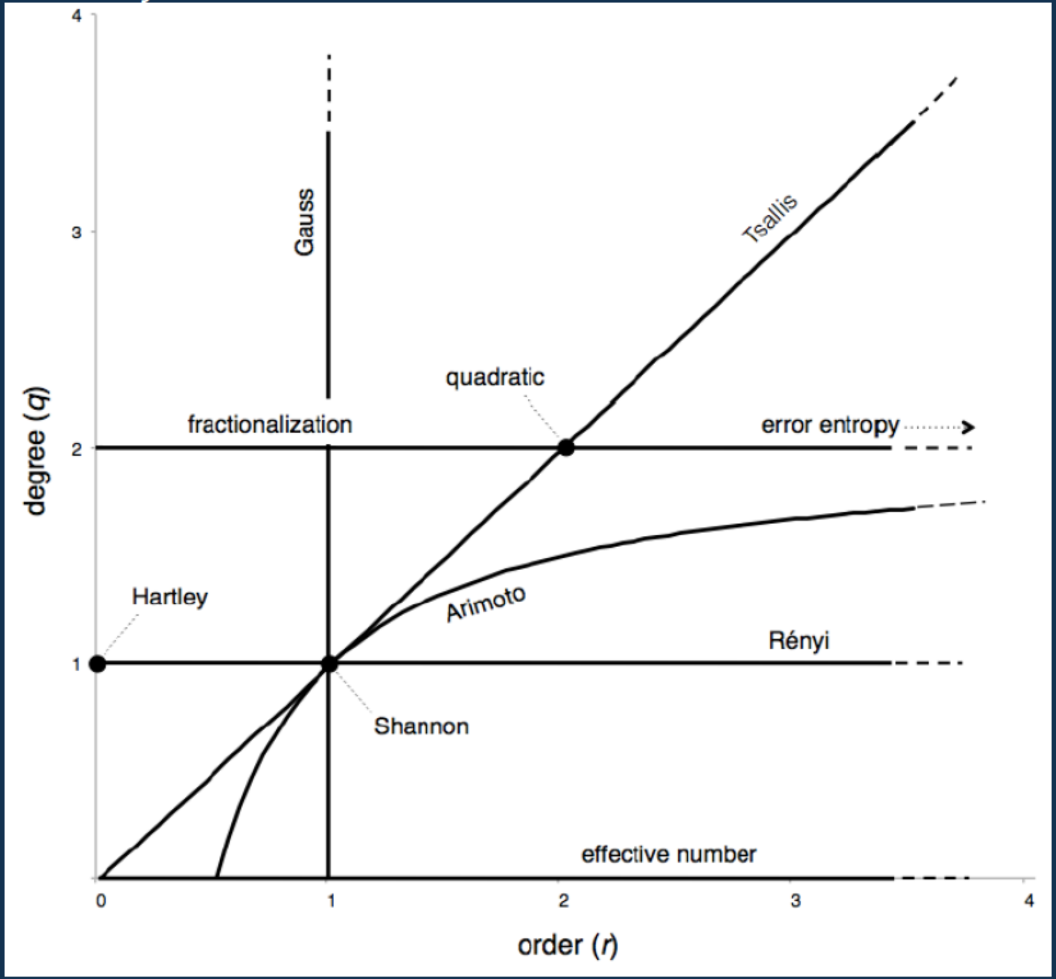
$$\mathrm{ent}(K) = \frac{1}{q-1}\left[1-\left(\sum_{i=1}^{n}P(k_i)^r\right)^{\frac{q-1}{r-1}}\right]$$

Sharma-Mittal entropies



The value of an experiment (question)

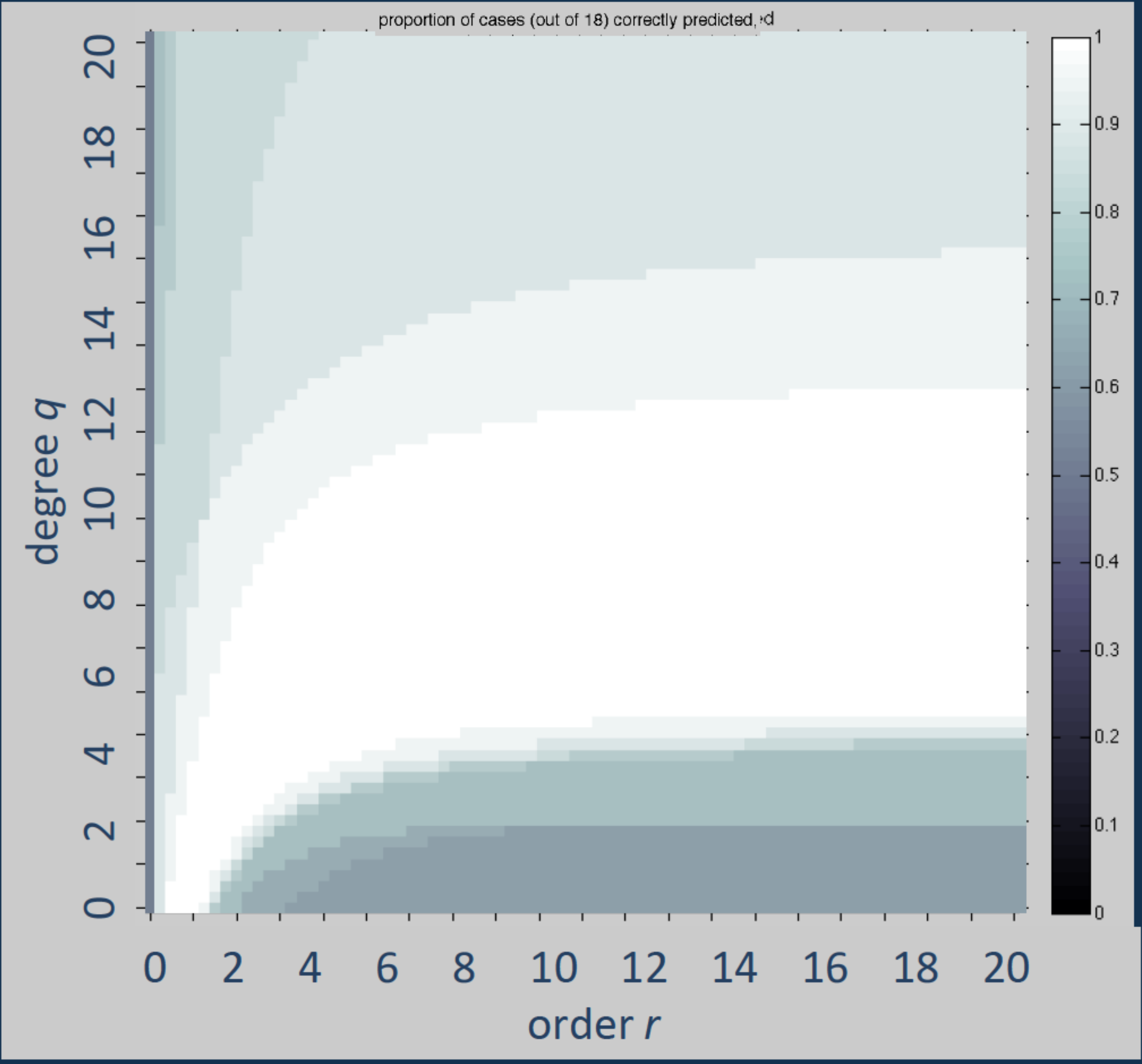
- consider experiment $D = \{d_1, d_2, \ldots d_m\}$, $m \geq 2$
- $eu_{IG}(K,D) = ent(K) - ent(K|D)$,
  $ent(K|D) = sum_{\{j=1:m\}}\, P(d_j)\, ent(K|d_j)$
- each entropy has a corresponding info gain
- which info gain best explains people?



Part 2 of 3: psychology of uncertainty & information

What Sharma Mittal information gain best explains people's choices given words-and-numbers probabilities?

- data from 18 Planet Vuma-type tasks (various papers)
- white = all experiments correctly predicted; black = none correctly predicted
- although individual responses very noisy, something systematic (attention to certainty)

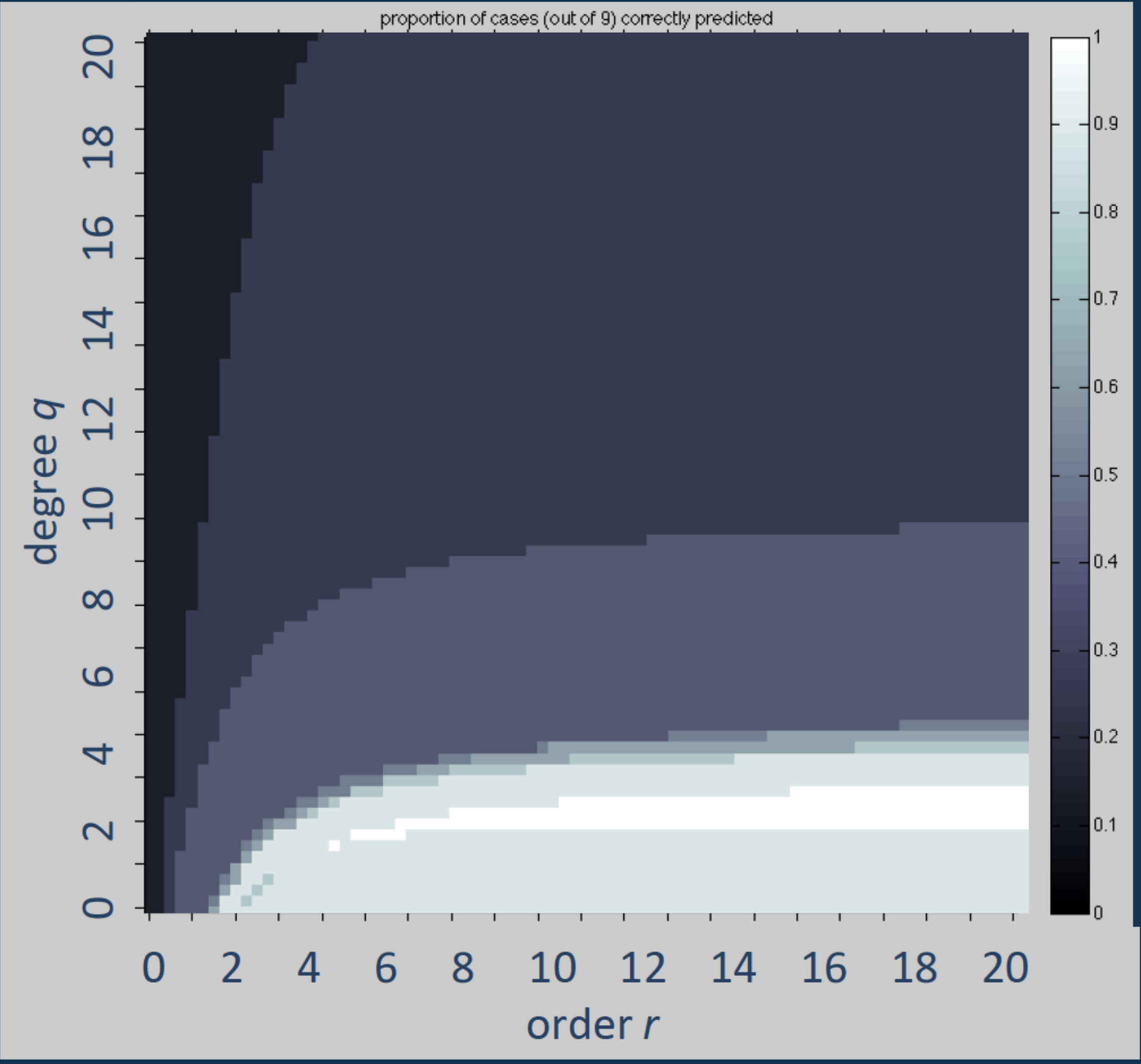


Species A plankton

Species B plankton

What information gain best explains people's choices given experience-based learning of probabilities??

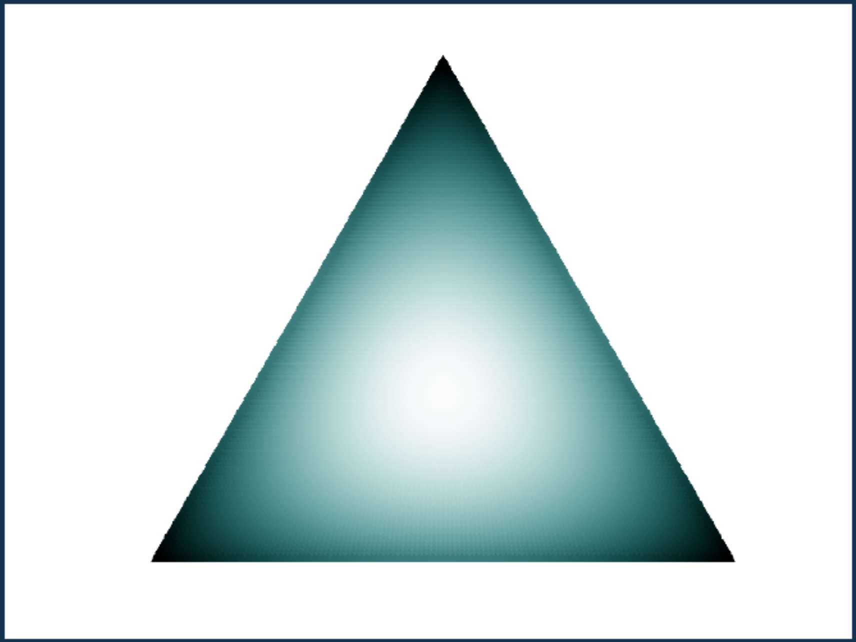
- data from search choices following experience-based learning
  (Nelson et al., *Psych Sci*, 2010)
- white = all experiments correctly predicted;
  black = none correctly predicted
- moderate Arimoto works as well as error entropy

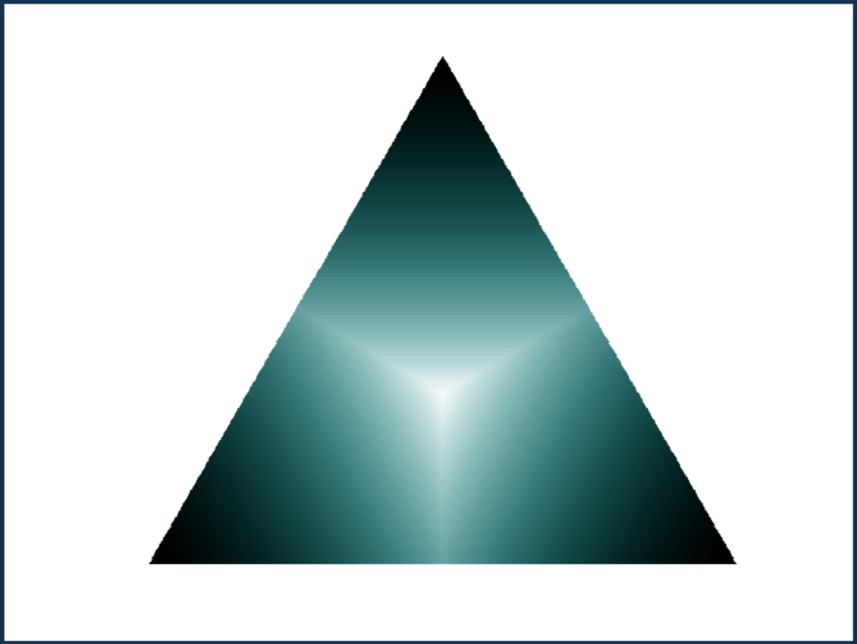


Our conundrum

Shannon is nice theoretically

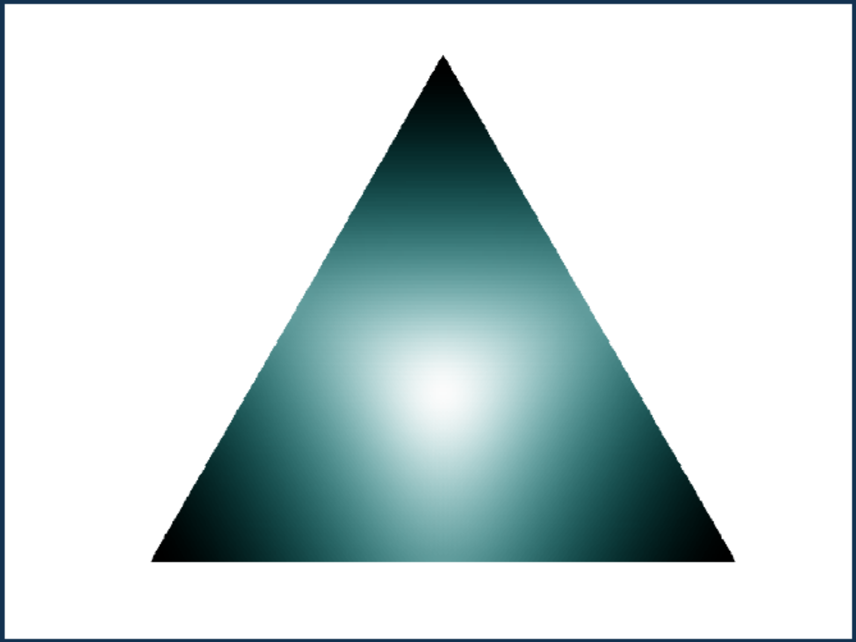
But error entropy explains empirical data better
(Nelson et al., *Psych Sci*, 2010)



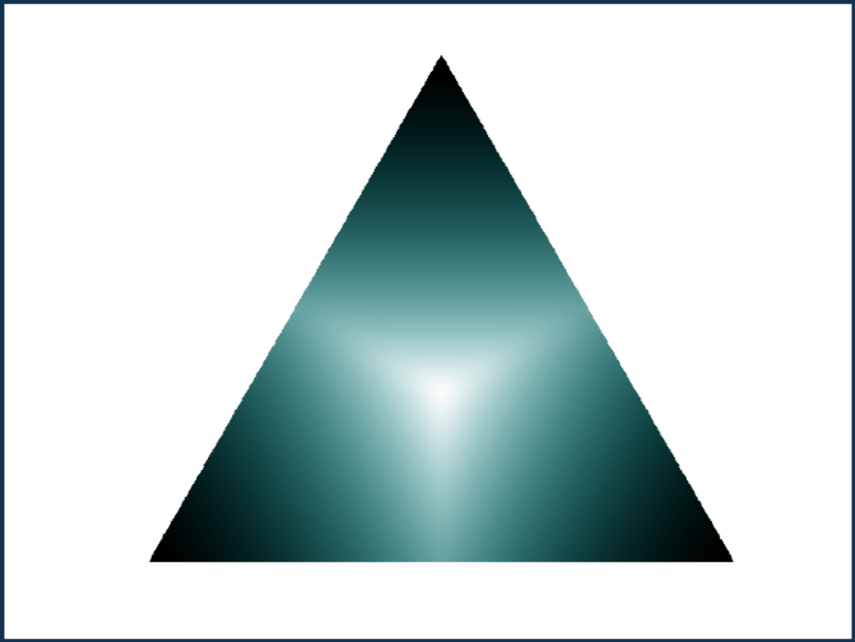


Maybe we can have our cake and eat it too?

Arimoto
(order=5, degree=1.8)



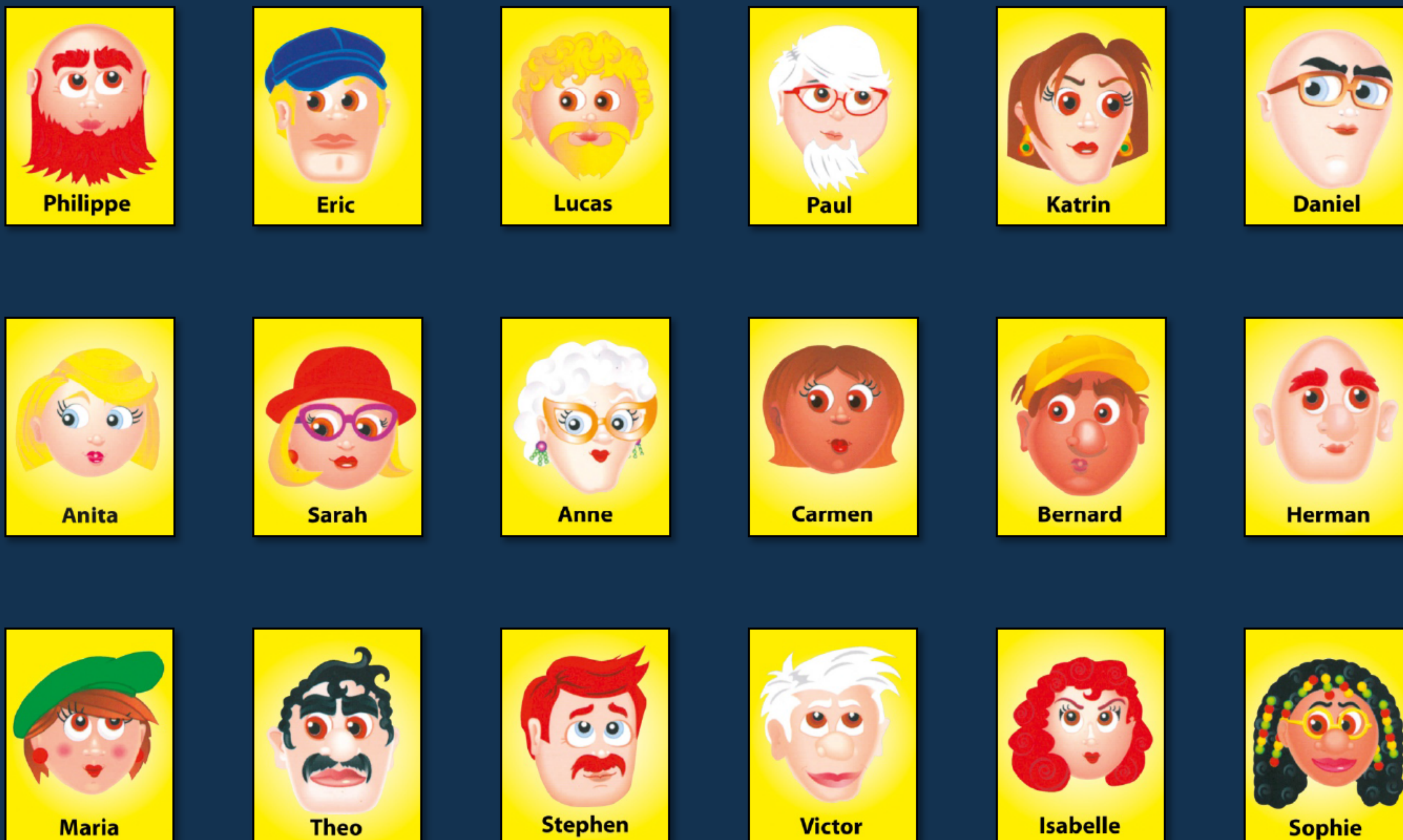
Arimoto
(order=20, degree=1.95)



The Person Game. (non-strategic)

Goal: identify the person, with fewest yes-no questions

from Nelson, Divjak, Gudmundsdo r, Martignon & Meder, *Cognition*, 2014

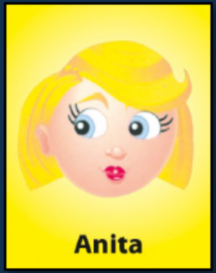


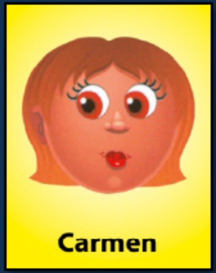
The Person Game.

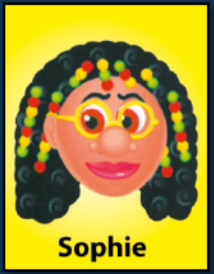
Is it a male face?

The Person Game.

Is it a male face? No

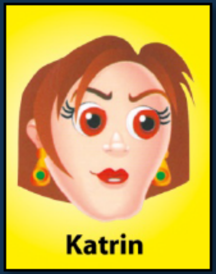


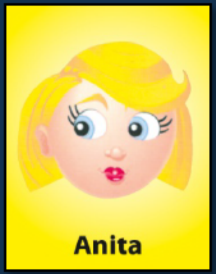


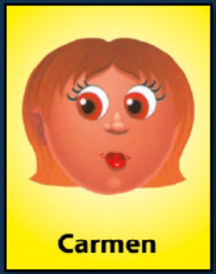


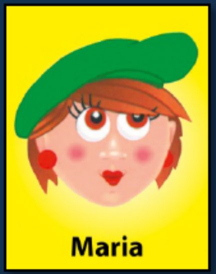
The Person Game.

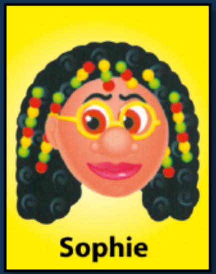
Do they have brown hair?





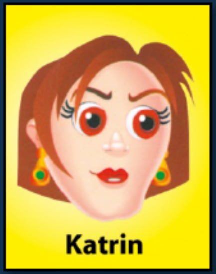


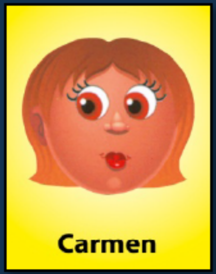




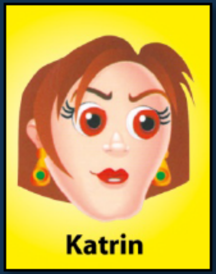
The Person Game.

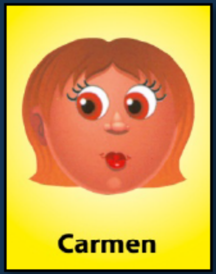
Do they have brown hair? Yes





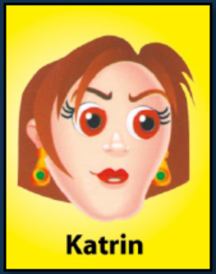
The Person Game.
Do they have a hat?

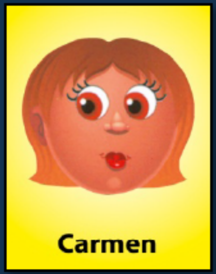




The Person Game.

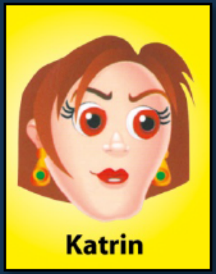
Do they have a hat? No

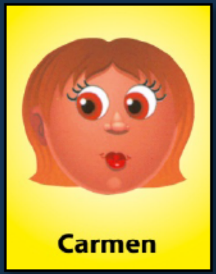




The Person Game.

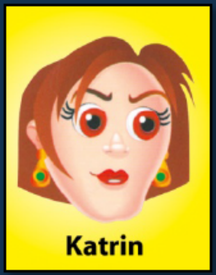
Do they have earrings?



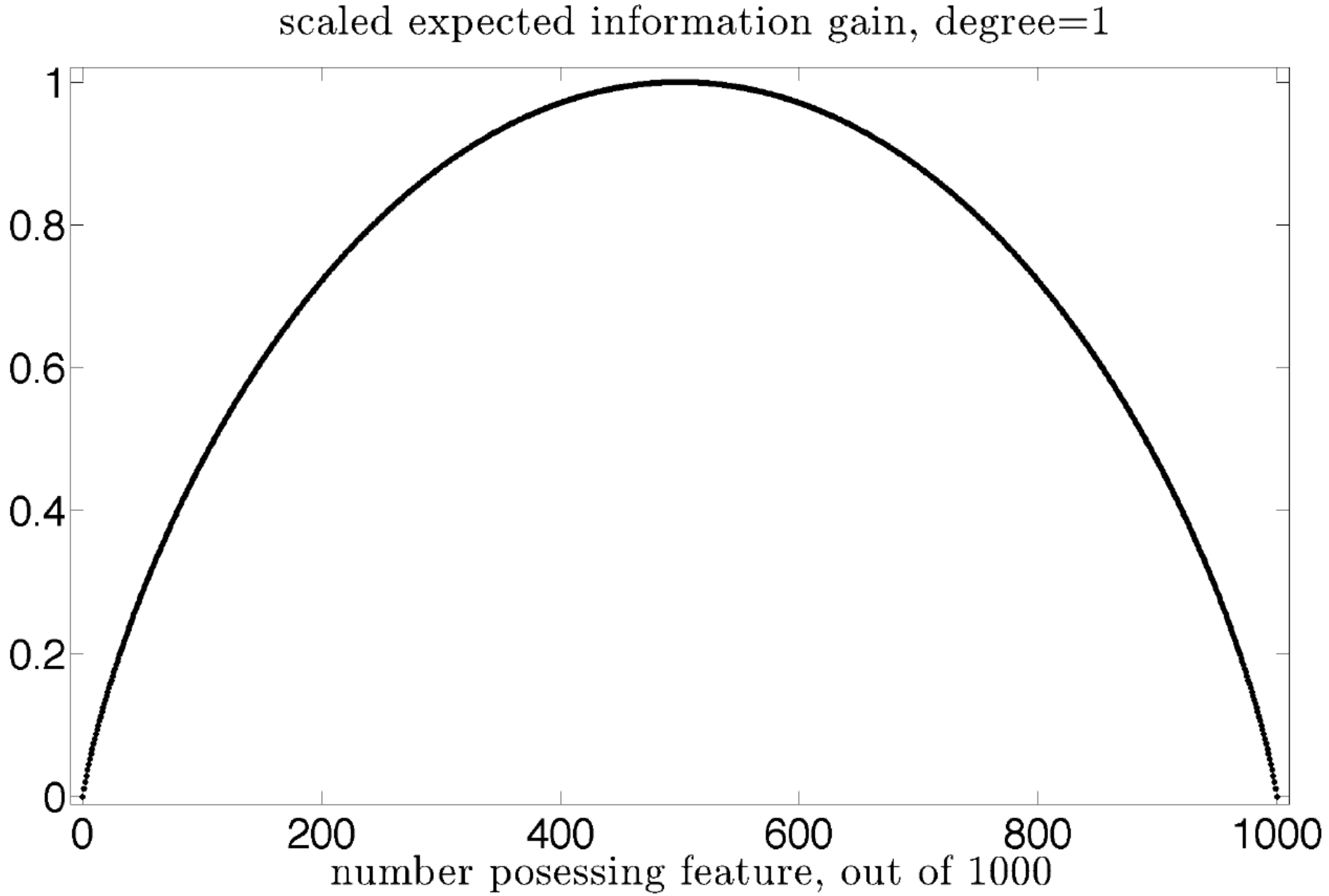


The Person Game.

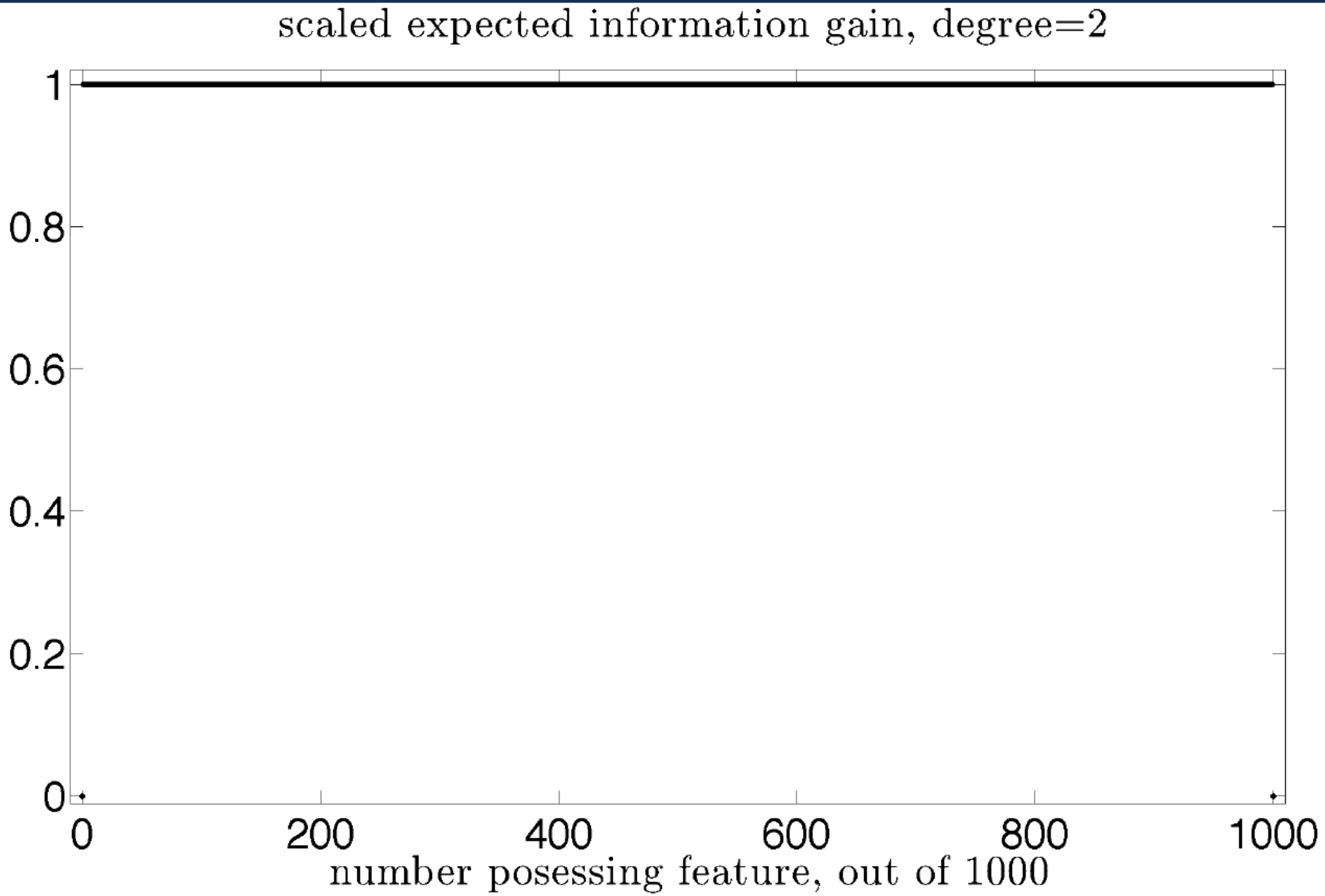
Do they have earrings? Yes



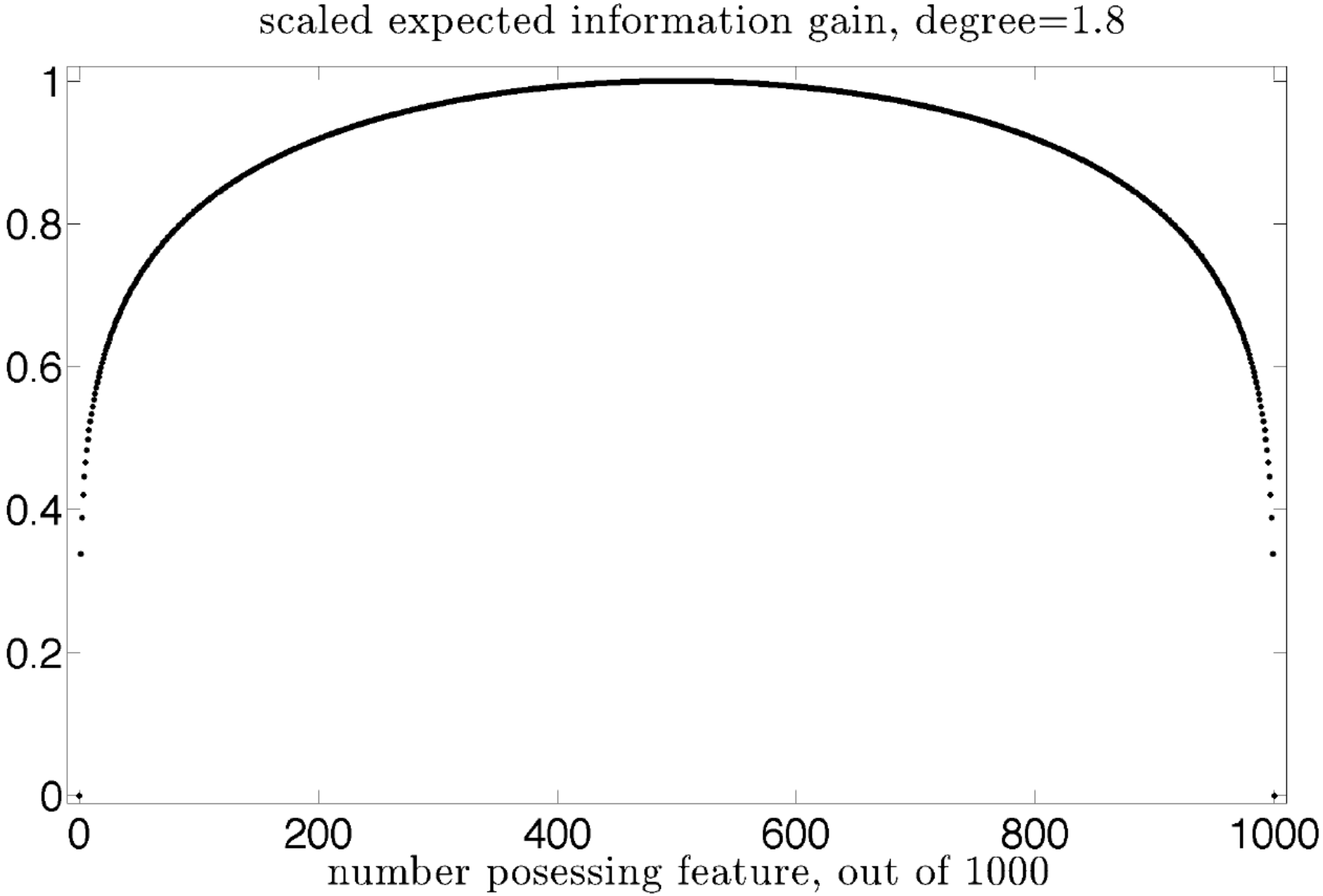
Shannon entropy likes splithalfy questions (splithalfiness)
"Ask about a feature that is possessed by 50% of remaining items"



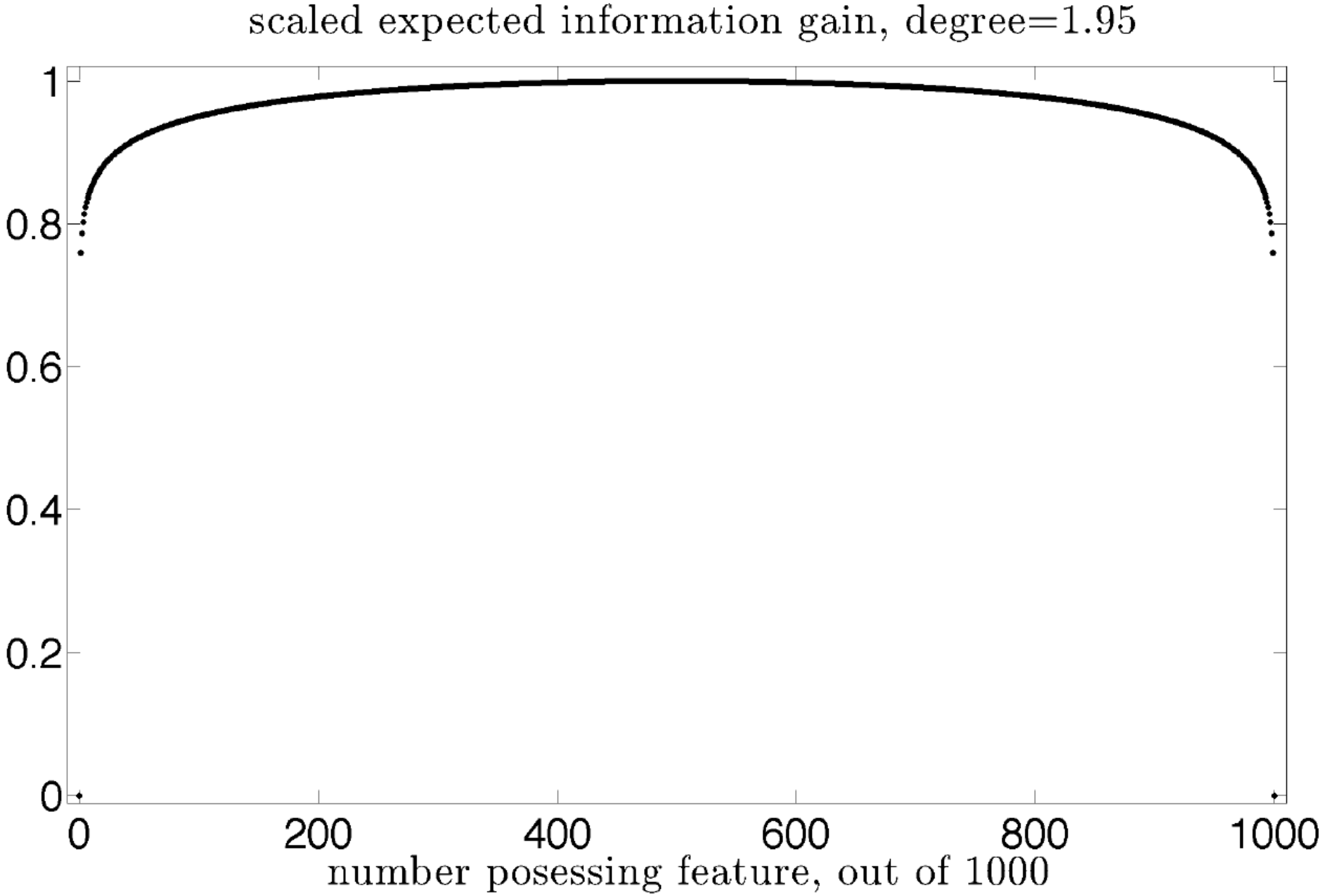
Probability gain is indifferent to splithalfiness
"All questions are equally useful"



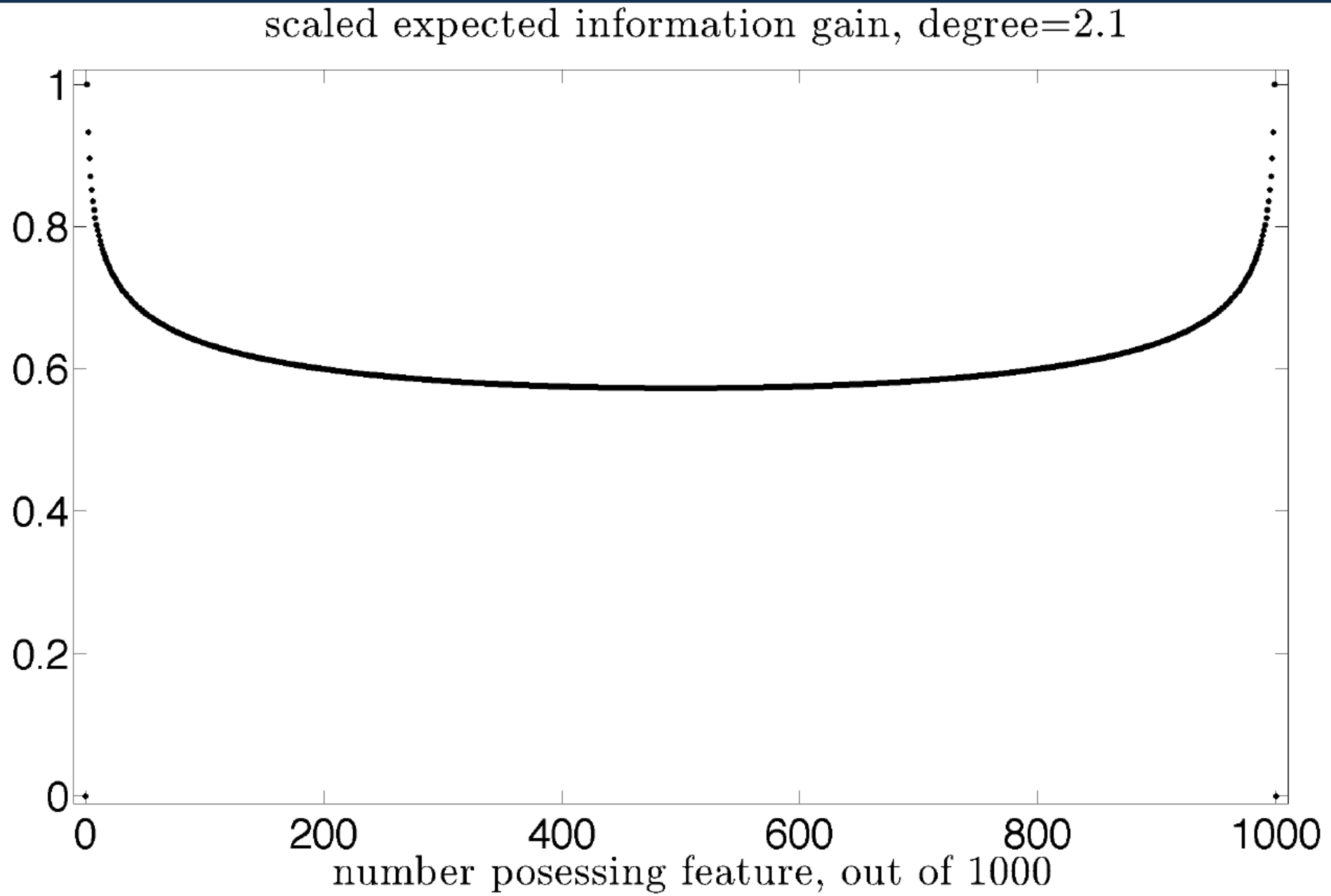
Arimoto (order=5, degree=1.8) entropy likes splithalfiness
"Have your splithalfiness and explain your data too!"



Arimoto (order=20, degree=1.95) entropy likes splithalfiness
"Have your splithalfiness and explain your data too!"



Higher-degree measures *dislike* splithalfiness: "Better to ask a 1:999 question than a 500:500 question"



Interim Conclusions: Entropy and Information

- Sharma-Mittal unifies many measures
- probability gain explained some data best, but had undesirable properties, and failed to explain other data
- Sharma-Mittal helped us find normatively desirable measures with better descriptive psychological adequacy than Shannon or probability gain
- Sharma-Mittal generates novel, testable, predictions for psychology (and neuroscience, applied domains, …)

contact: jonathan.d.nelson@gmail.com,
http://jonathandnelson.com/

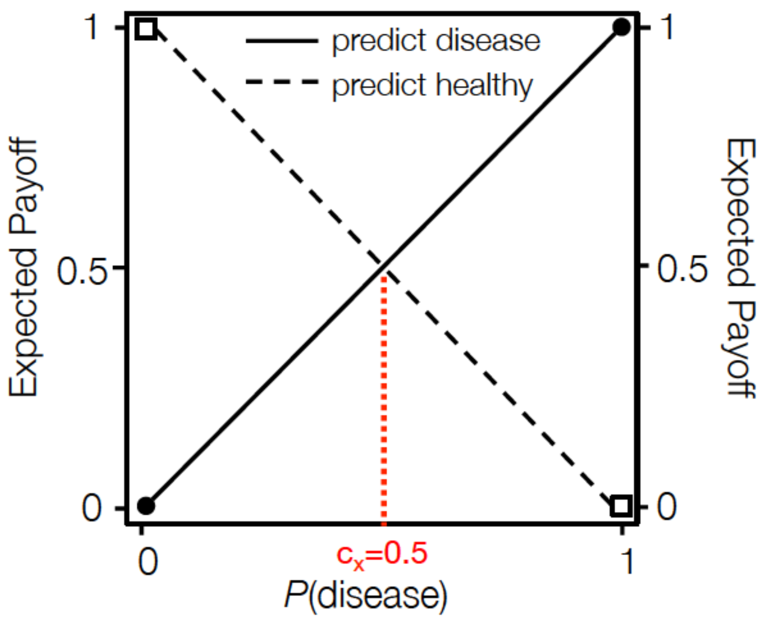
Part 3/3: brainstorming future research

What if asymmetric payoffs apply?

Meder & Nelson (2012), *Judgment and Decision Making*

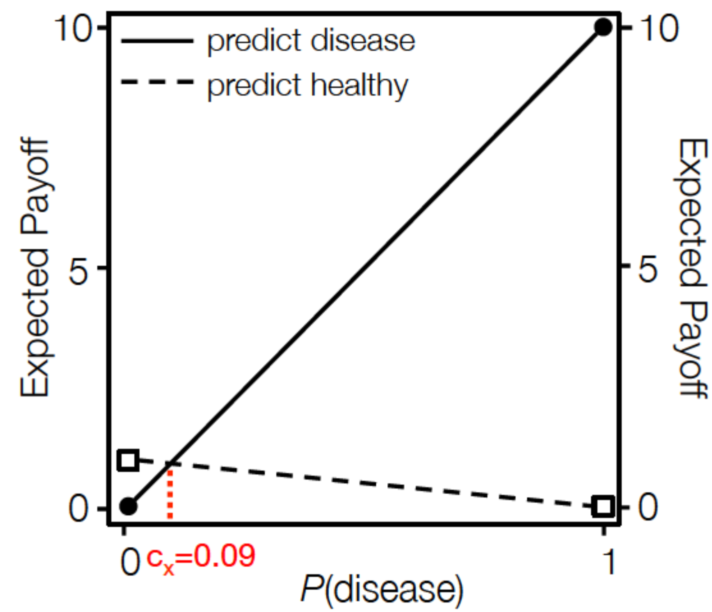
Symmetric rewards

	Disease	Healthy
Predict disease	1	0
Predict healthy	0	1

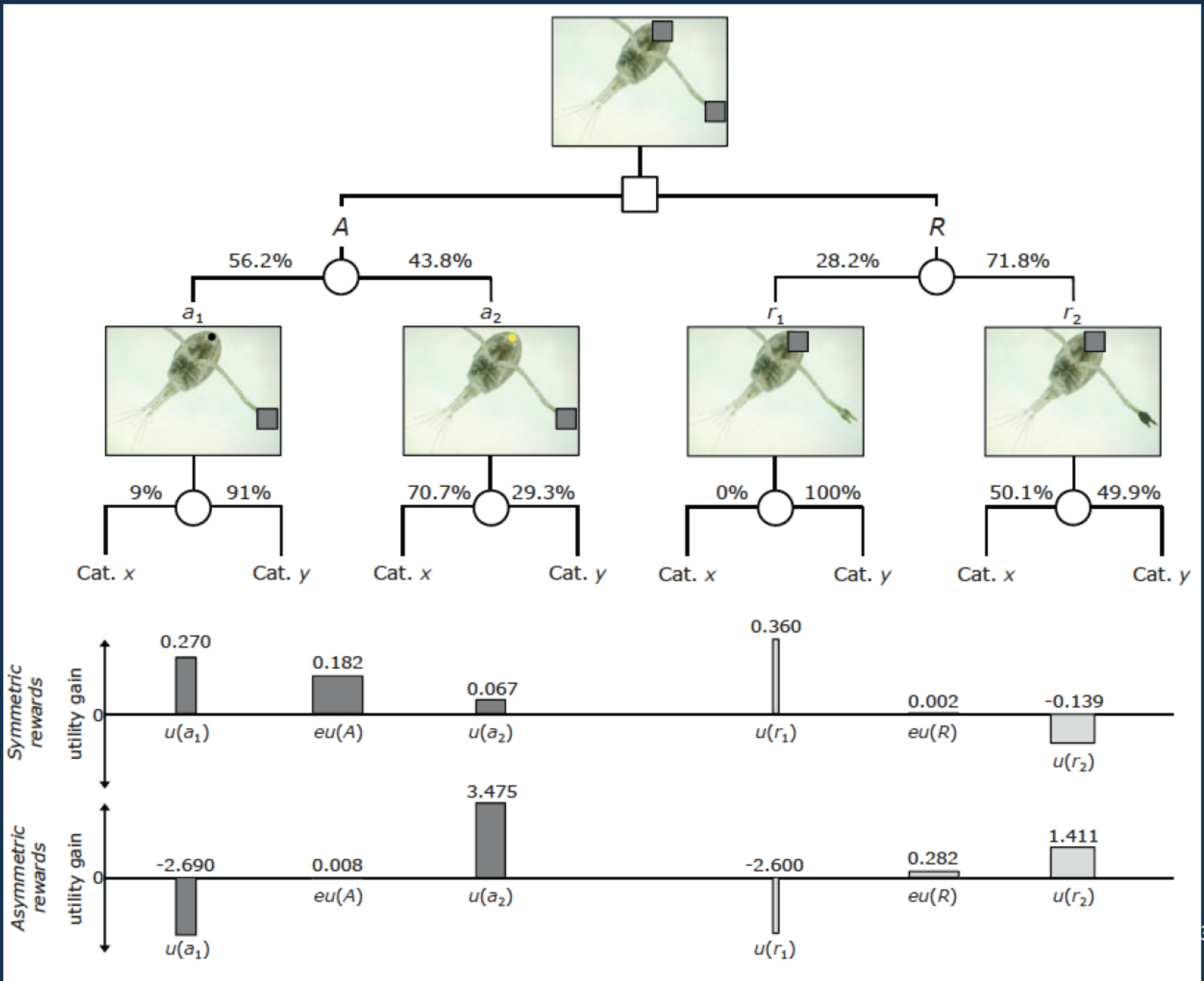


Asymmetric rewards

	Disease	Healthy
Predict disease	10	0
Predict healthy	0	1



What if asymmetric payoffs apply?



1:1:0:0

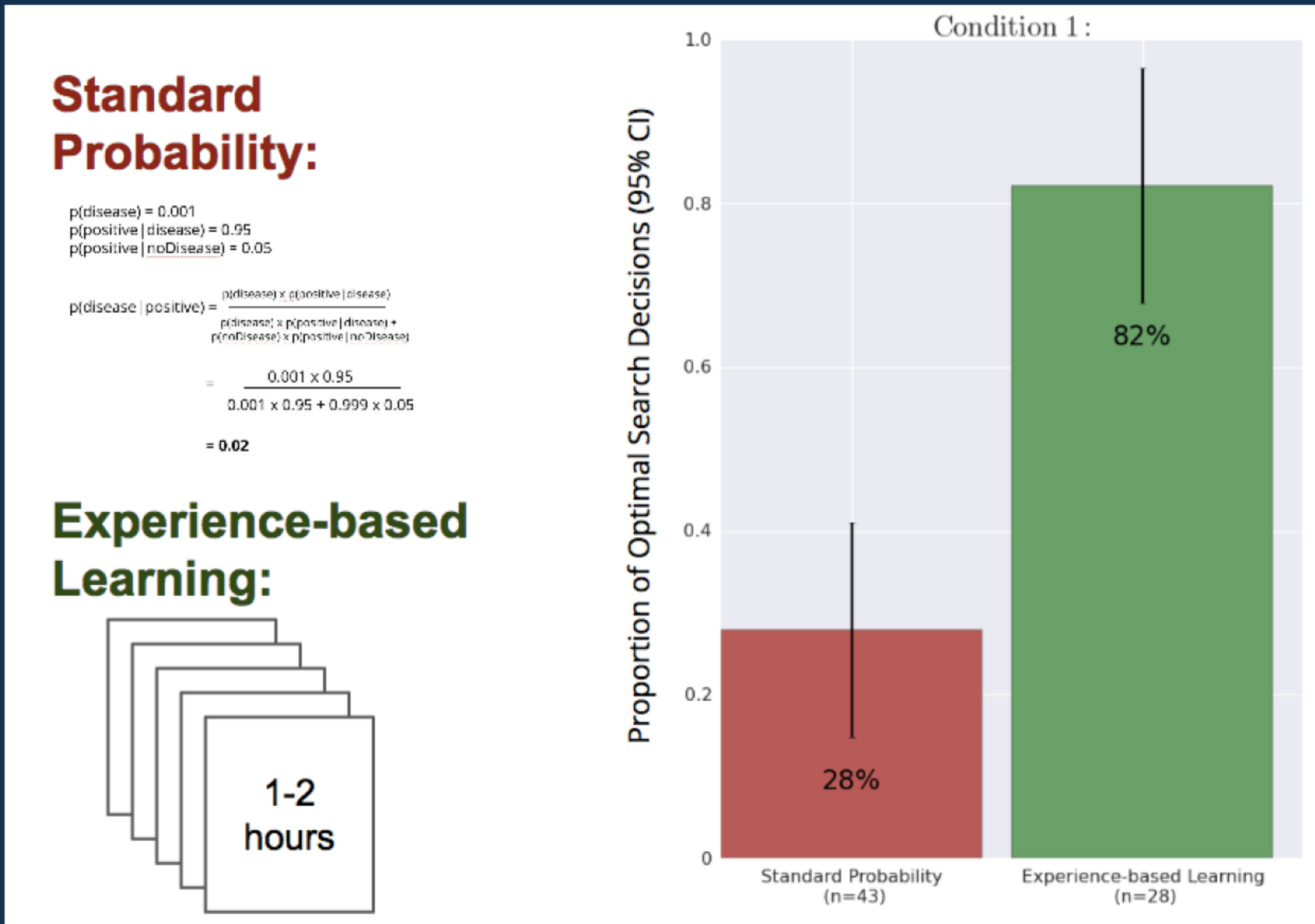
payoffs: 10:1:0:0

What if asymmetric payoffs apply? → Future collaborative research point

- Payoffs matter for test usefulness, and not only for action taken
- People have a hard time taking situation-specific usefulness functions into account
- Maybe an intuitive cover story would help?

Facilitating good information selection decisions

Wu, Meder, Filimon, & Nelson (in press). *Journal of Experimental Psychology: Learning, Memory, and Cognition.*



Facilitating good information selection decisions

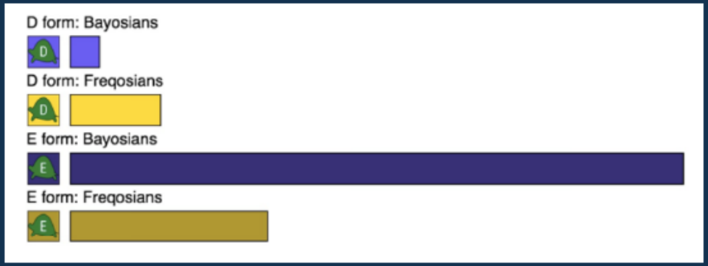
Wu, Meder, Filimon, & Nelson (in press). *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

- Standard probability format not good for Bayesian reasoning: Why use it for information search?
- Planet Vuma-type scenario
- Goal to choose test to maximize classification accuracy
- Also queried various probabilities
- 14 formats: probability, natural frequency, and visual

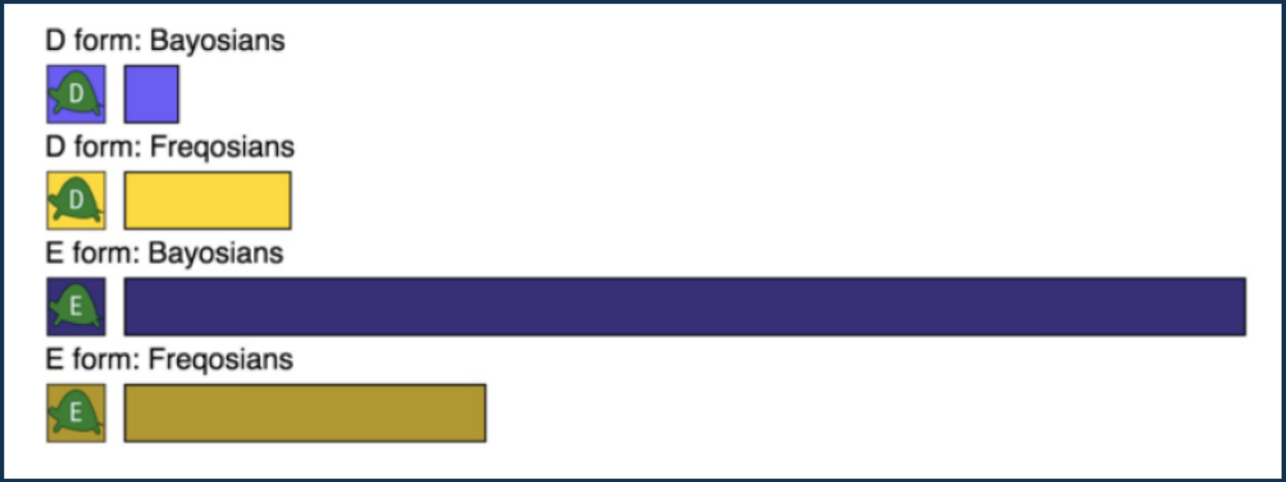
Facilitating good information selection decisions: Results

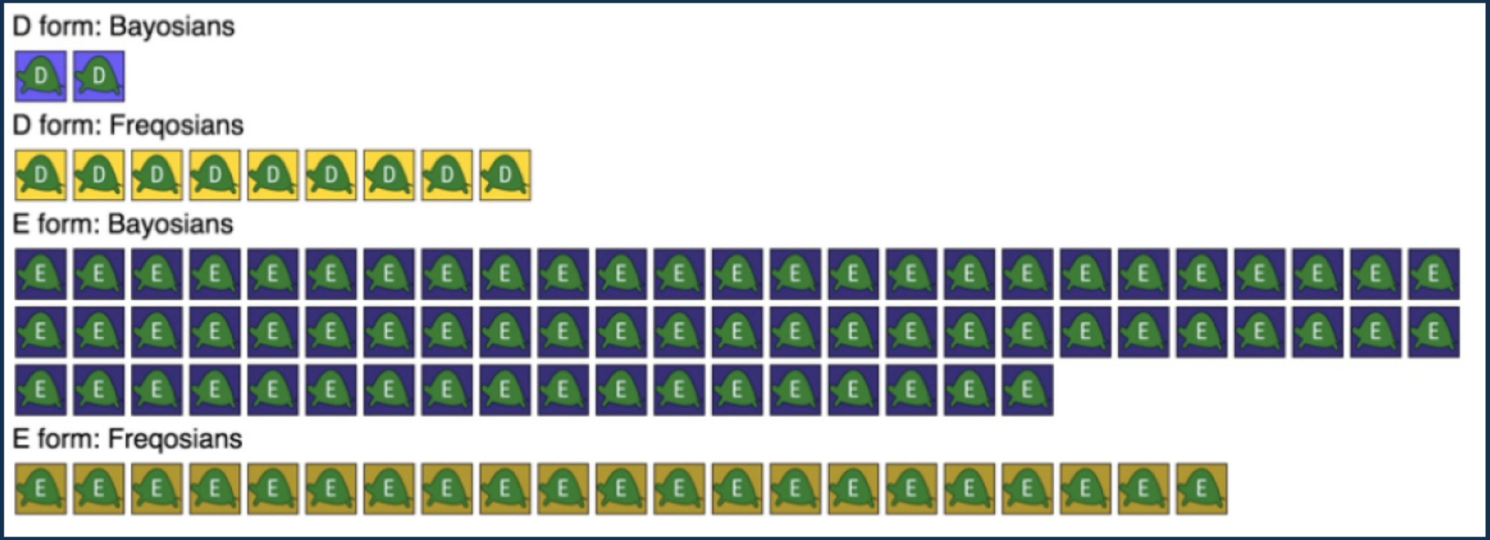
Wu, Meder, Filimon, & Nelson (in press). *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

- Judgment accuracy not related to search-task performance
- Numeracy slightly related to search-task performance
- Worst format was standard probability format
- Best format was posterior bar graph (not countable)
- Posterior icon array, posterior probability formats also good
- No natural frequency format was very good



Using helpful formats for Bayesian inference and search tasks → Future collaborative research point



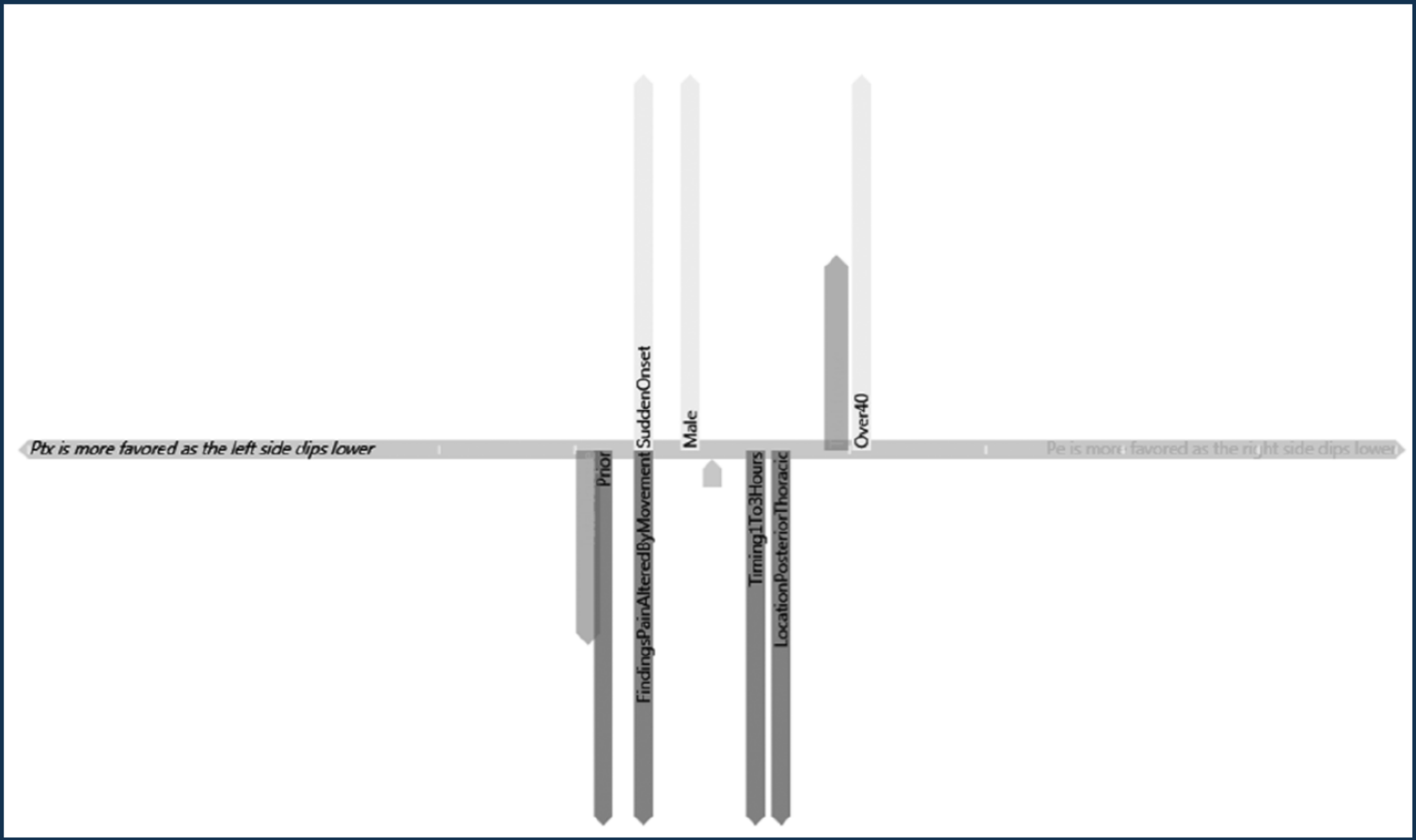


Combining evidence:
→ Mathematical / future collaborative research point

- Suppose:
  - P(J is a terrorist) = 0.01
  - P(J is not a terrorist) = 0.99
  - P(J researched travel to Syria | J is a terrorist) = 0.8
  - P(J researched travel to Syria | J is not a terrorist) = 0.1
  - P(J has been to Turkey | J is a terrorist) = 0.5
  - P(J has been to Turkey | J is not a terrorist) = 0.3
- J has researched travel to Syria, and has been to Turkey. What is the new probability that J is a terrorist?
- Correct answer: we have no idea whatsoever.
- If experience-based learning, people presume class-conditional independence
  Jarecki, Meder, & Nelson (in press), *Cognitive Science*

Balance beam metaphor and class-conditional independence

Hamm, Beasley, Johnson (2012). *Medical Decision Making*



"Nothing drives basic science better than a good applied problem"
(Newell & Card, 1985, p. 238)

- Generalized uncertainty measures that
  - apply if probabilities aren't quite known (cf Jonas's work)
  - take payoffs into account
- Representing probabilities helpfully, to facilitate inference and search decisions
- Combining different sources of evidence: how to take dependencies among sources into account
- Figuring out when (and how) to get people to take payoffs into account when evaluating evidence
- Bayesian and information-theoretic analysis of SAT, like ACH
  - no justification for excluding positive info; info combination rules; etc.

contact: jonathan.d.nelson@gmail.com,
http://jonathandnelson.com/