Cognitive science for machine learning 3: Models and theories in cognitive science (Part 2)

Nick Chater

OVERVIEW

1. FRAGMENTATION IN COGNITIVE SCIENCE
2. SCALING AND CODING
3. INFERENCE
4. ARCHITECTURE
5. WHERE NEXT?

1. FRAGMENTATION IN COGNITIVE SCIENCE

FRAGMENTION RATHER THAN INTEGRATION…

- ...of theory
  - Language acquisition
  - Perception
  - Memory
  - Reasoning
  - Decision making

  ...are often viewed as independent…

- ...of experiments
  - Focus on increasingly detailed behavioral and/or imaging studies of specific phenomena
  - Extrapolation across tasks or domains is typically secondary

MACHINE LEARNING AND AI AS AN INTEGRATING FORCE

- Identifying and solving abstract structures of problems
- And potentially common tools for their solution
- Just as ML techniques apply across a variety of application domains...
- ...so common ML principles might apply across aspects of cognition (e.g., Bayes in perception, categorization, inference, learning, causal reasoning)
- Key goal of cognitive science: search for general principles

REINTEGRATING COGNITIVE SCIENCE

- The ideal:
  - one game of 20 Questions for cognition
  - not a separate game of 20 Questions for lexical decision, one for short term verbal memory, one for face recognition...

- Which questions?

- Need to be general and empirically tractable

CANDIDATE QUESTIONS AND PRINCIPLES

Principle		Domains
SCALING AND CODING	SCALE INVARIANCE	Much more general than cognitive science
	ABSOLUTE VS RELATIVE CODING	Perception, decision making, valuation, well-being
INFERENCE	SIMPLICITY	Perceptual organization, language acquisition, inductive inference, memory
	GENERATIVE VS DISCRIMINATIVE MODELS	Perception, classification
ARCHITECTURE	MODULARITY VS UNIFIED SYSTEM	Perception, motor control, language, reinforcement learning

2. SCALING AND CODING

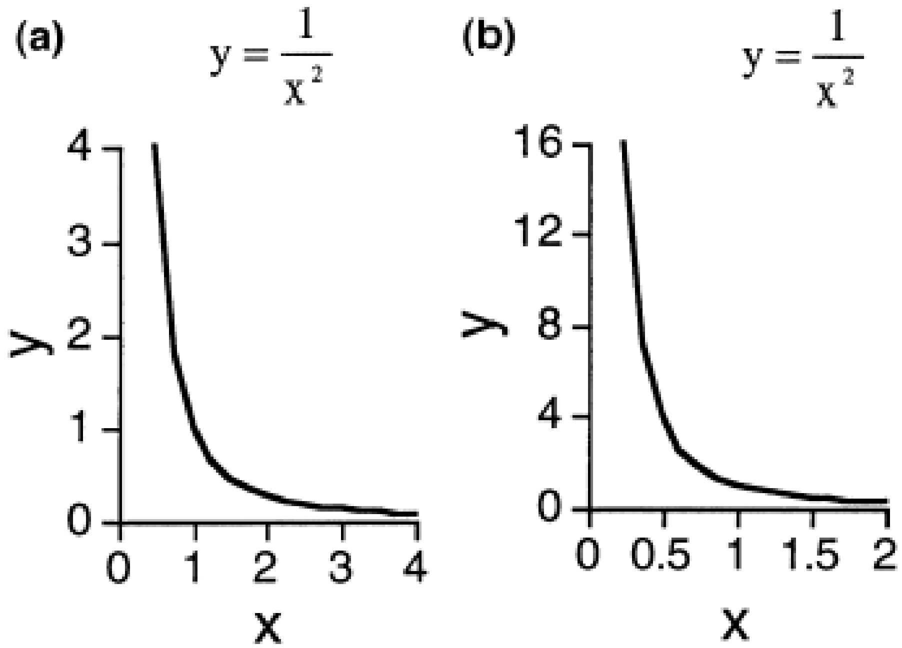
i. SCALE INVARIANCE

ii. ABSOLUTE vs RELATIVE CODING

SCALE-INVARIANCE

- In a nutshell:
  - Throw away "units"
  - Can you reconstruct them from your data?

- If *not*, phenomenon is scale-invariant

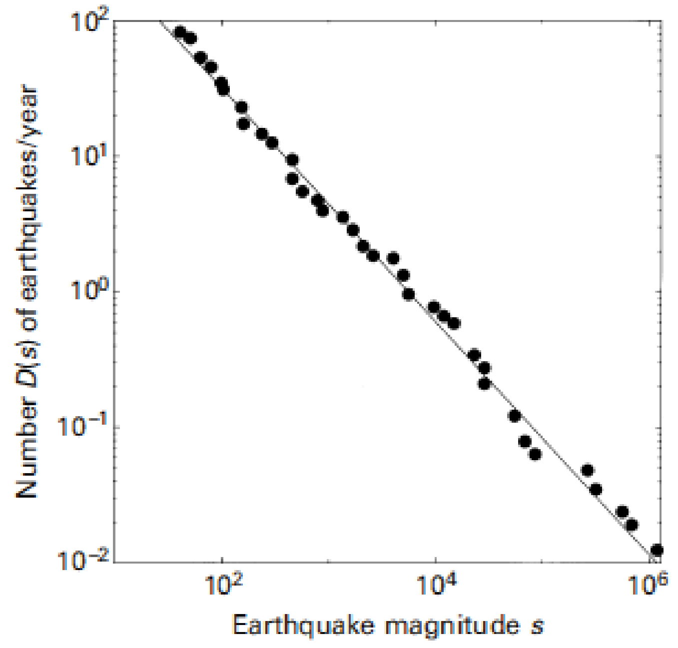


Only power laws $y \propto x^{\alpha}$ are scale invariant

THE UBIQUITY OF SCALE-INVARIANCE

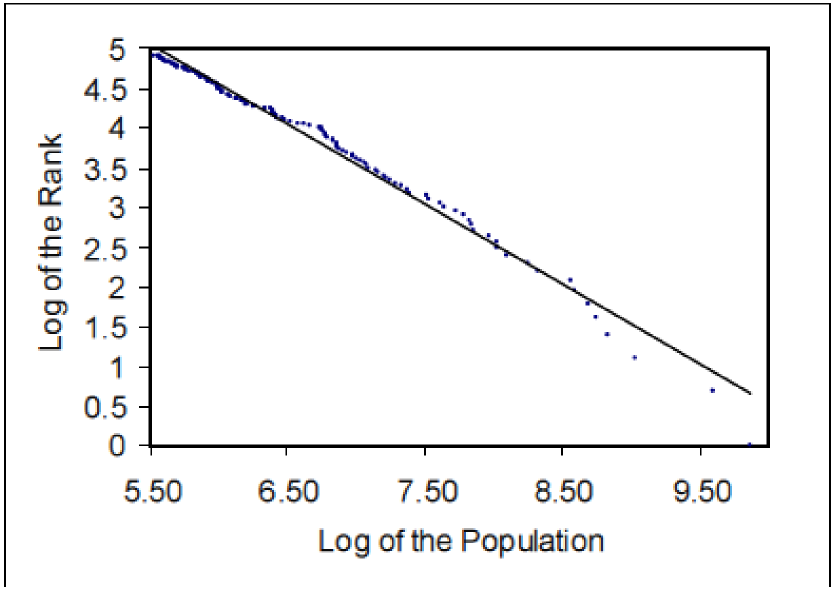
- City sizes
- Size of firms
- River sizes
- Distribution of digits (Benford's Law)
- Word frequencies (Zipf's Law)

- Scale-invariance as a "null hypothesis" which implies many well-known psychological laws…

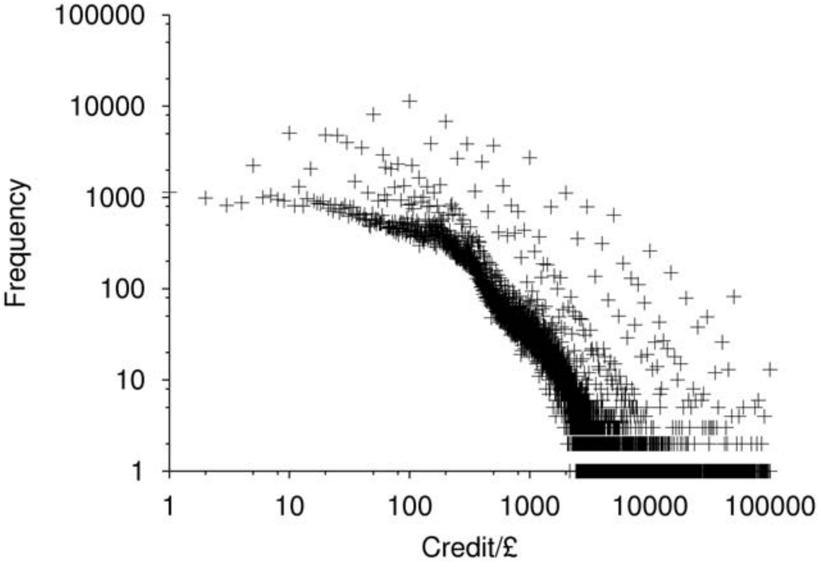


Frequencies of earthquakes of different magnitudes

THE UBIQUITY OF SCALE-INVARIANCE

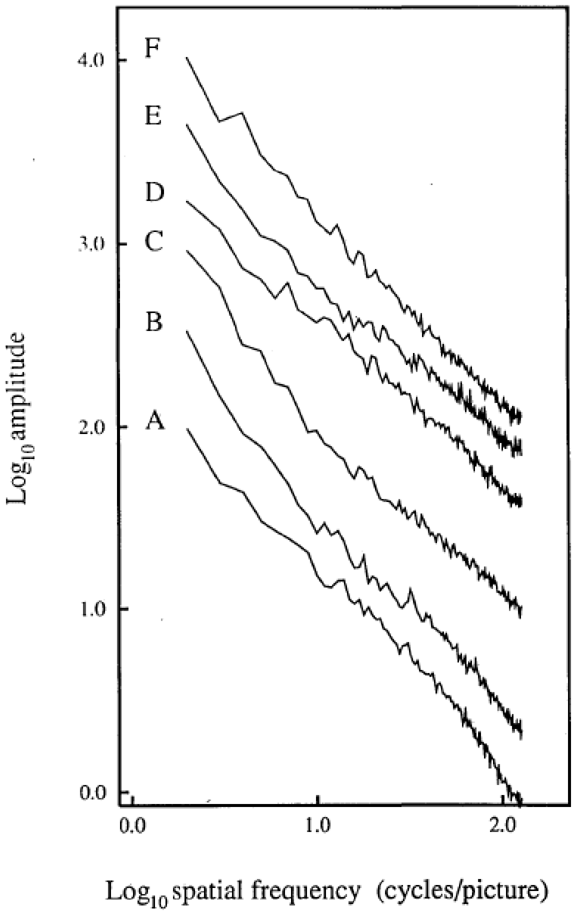


City sizes

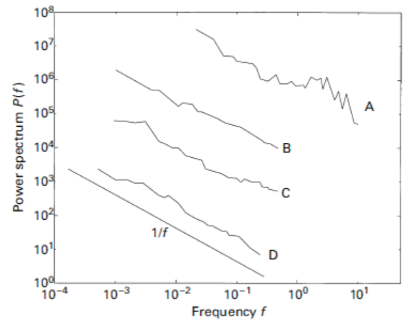


Bank transactions

SCALE INVARIANCE IN THE VISUAL ENVIRONMENT, AND SENSORY SYSTEMS



Amplitude spectrum of natural images Field, 1987



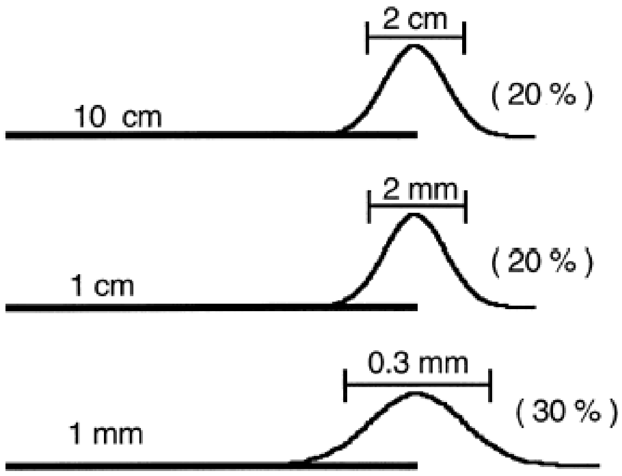
Audition: Voss and Clark 1978

- Scale-invariance in (some) aspects of psychophysics
  - Detection of change in grating amplitude, frequency or orientation (Jamar et al 1983; Kingdom et al 1985)
    - Though detection itself is not scale-invariant
- Self-similar transforms in retinal machinery Teichert et al, 2007
  - and image processing and computer vision (Barnsley)

FROM SCALE-INVARIANCE TO PSYCHOLOGICAL "LAWS"

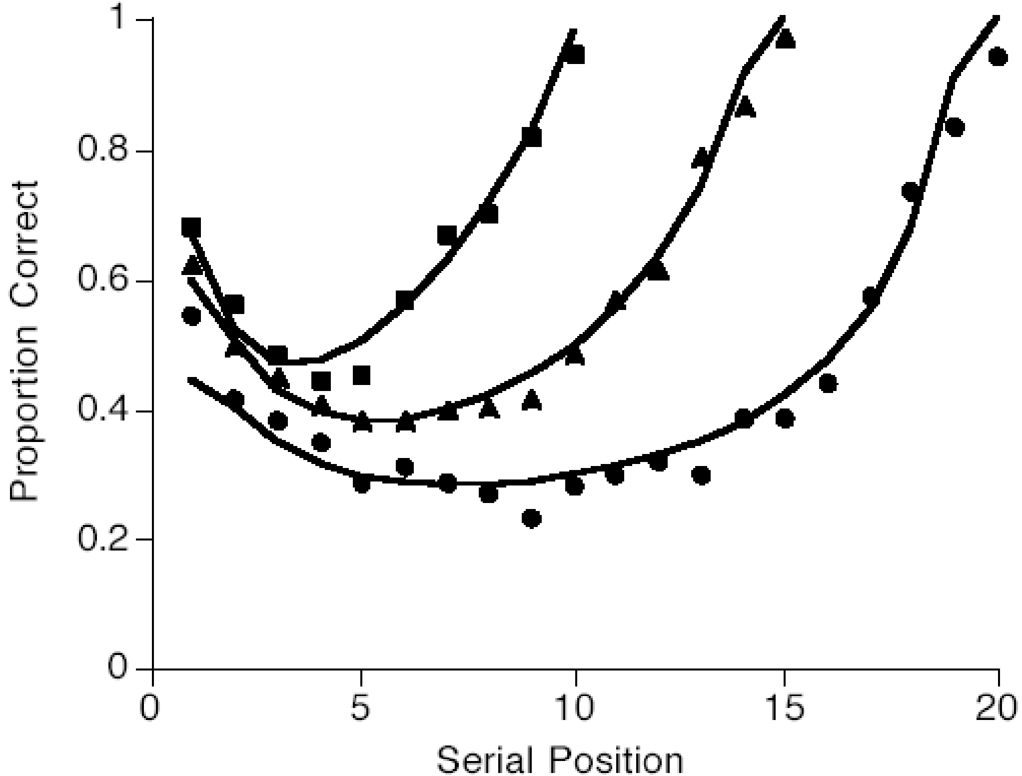
Regularity	*Form*	*Explanation*
Weber's Law	$\Delta I \propto I$	$\Delta I/I$=constant, if independent of units
Stevens' Law	$I^{\alpha} \propto S$ (power law)	$\Delta I/I \propto \Delta S/S$ Ratio preserving: input-output
Power law of forgetting	$m(t) \propto t^{-\alpha}$	Ratio preserving: memory-time
Power law of practice	$\mathrm{RT}(N) \propto t^{-\alpha}$	Ratio preserving: trials-speed
Fitts' Law (revised Kvalseth, 1980)	$T = a(\Delta D/D)^{\alpha}$	Ratio preserving: time-precision
Herrnstein's matching law	$\Pr("R_i") = \frac{\left(Payoff(R_i)\right)^{\alpha}}{\sum_j \left(Payoff(R_j)\right)^{\alpha}}$	Ratio preserving: Prob of choice to mean payoff

WEBER'S LAW



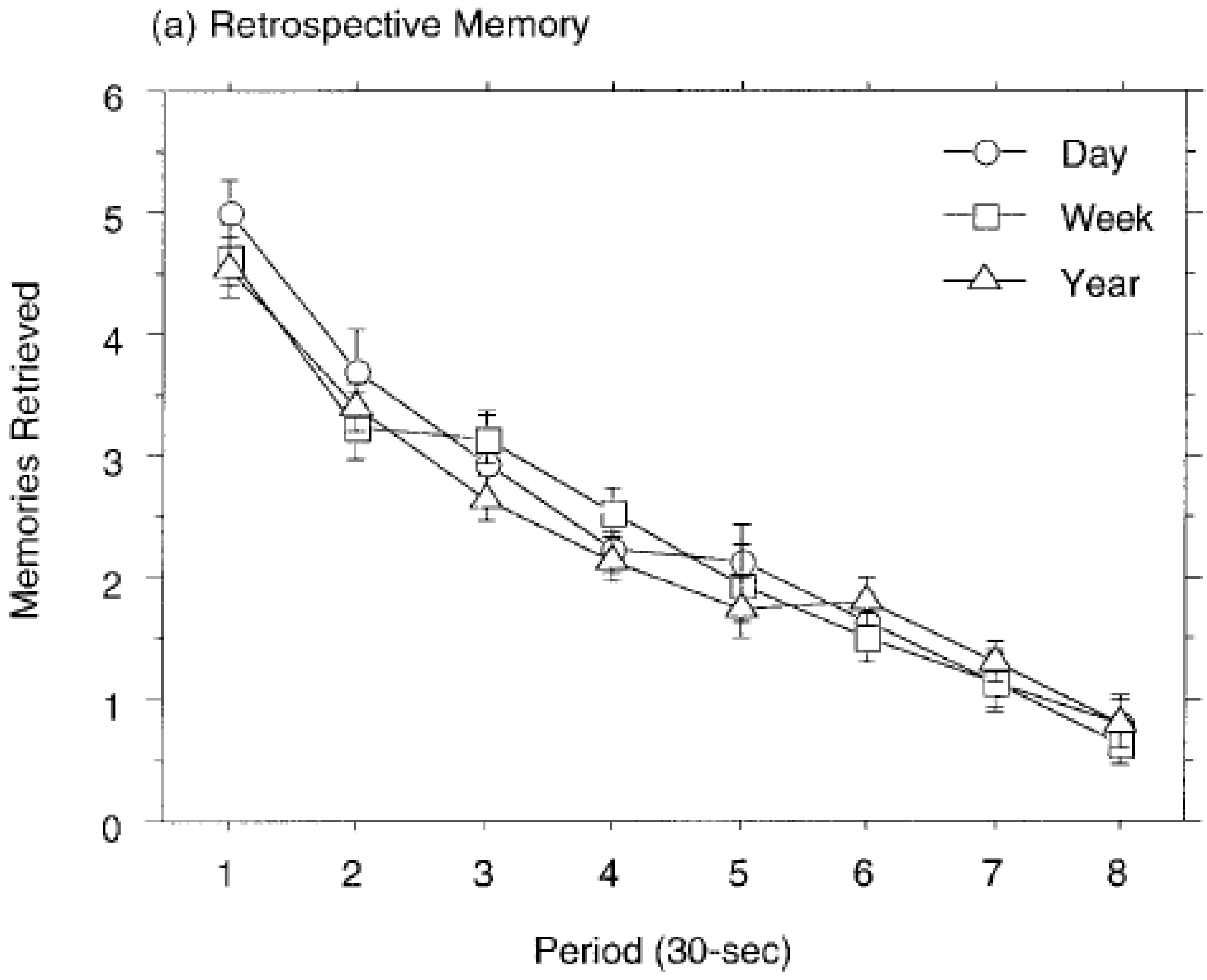
Endless cases of invariance, in perception, motor control, learning and *memory*

SERIAL POSITION IN IMMEDIATE FREE RECALL

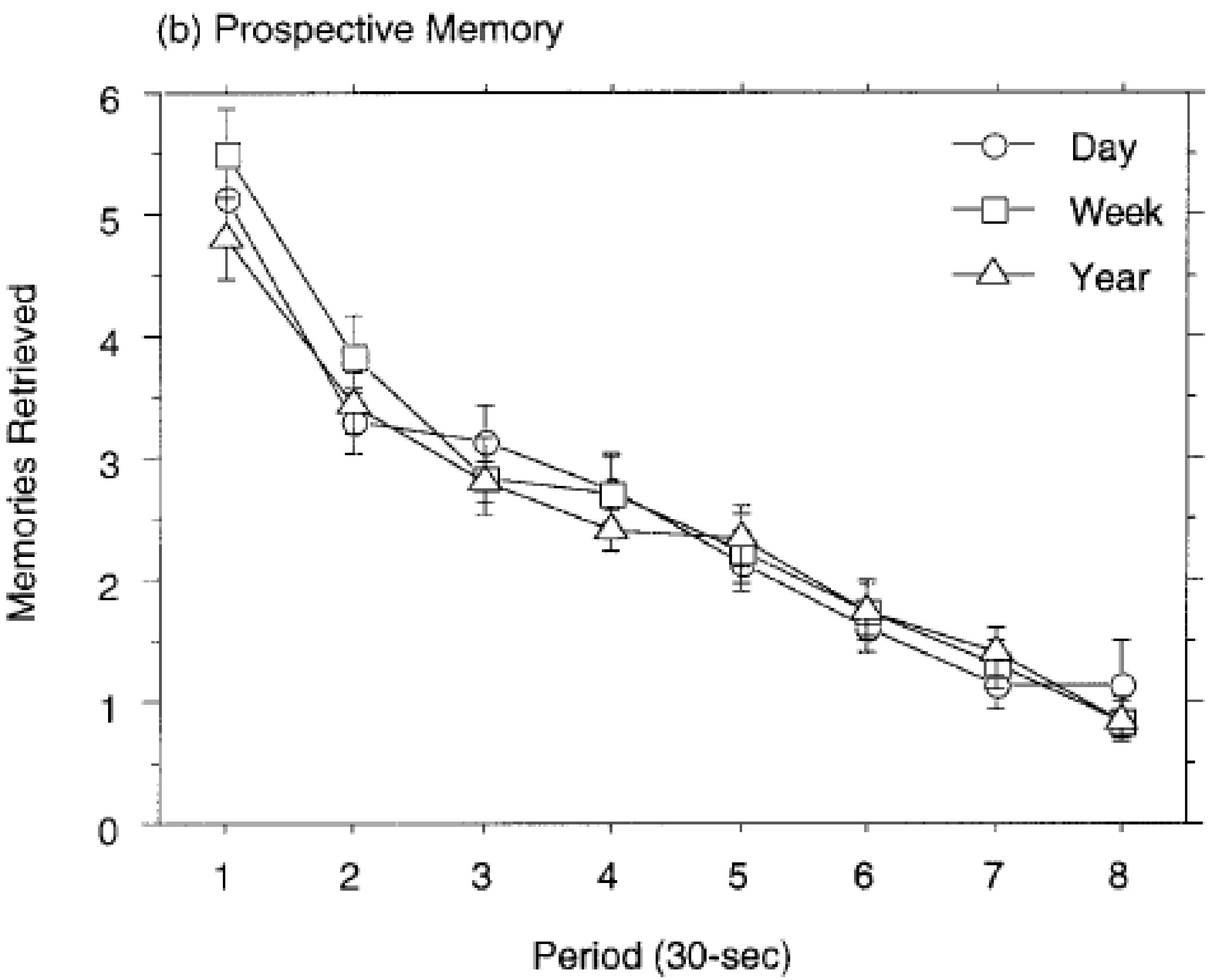


Data from Murdock, 1962; model fits using SIMPLE (Brown, Neath & Chater)

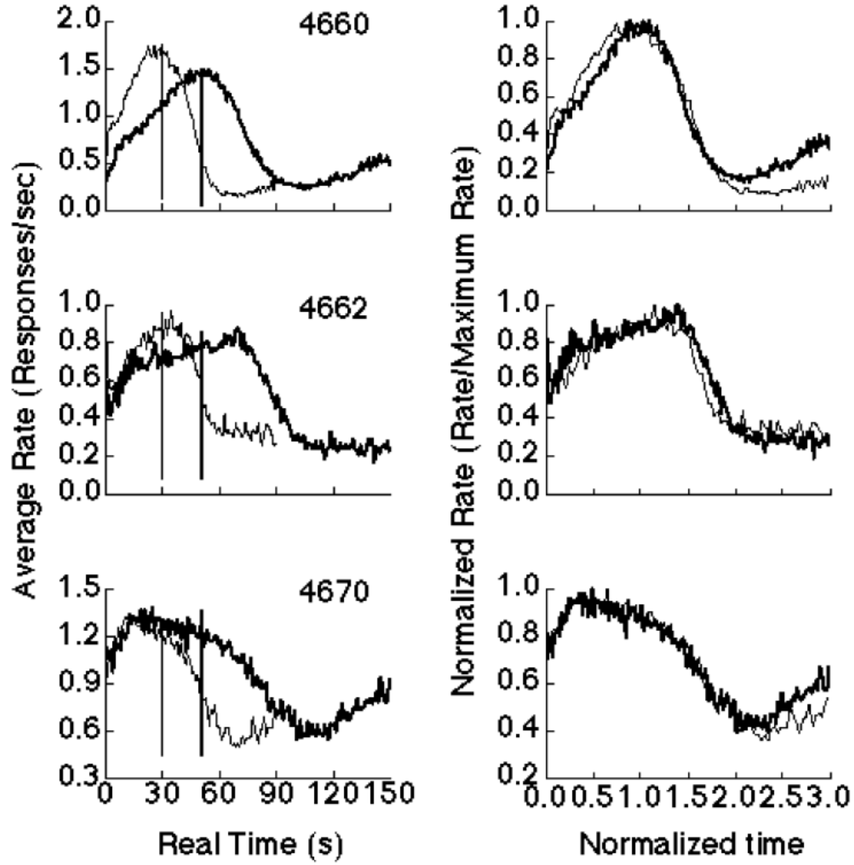
MEMORY RETRIEVAL OVER DIFFERENT TIME PERIODS IN RETROSPECTIVE MEMORY
(Maylor, Chater & Brown, 2001, *PB&R*)



AND PROSPECTIVE MEMORY



TIME-INVARIANCE OF ANIMAL AND HUMAN LEARNING



Gallistel and Gibbon peak procedure pigeon data

IMPLICATIONS

- Lots of quantitative relations can be predicted purely from scaling

- Be careful in introducing scales into a model/theory

- Care linking up scales across levels of analysis
  - e.g., neural long-term potentiation appears to have a distinctive time-scale;
  - learning and memory do not

- Merely capture scaling laws is not good evidence for a model

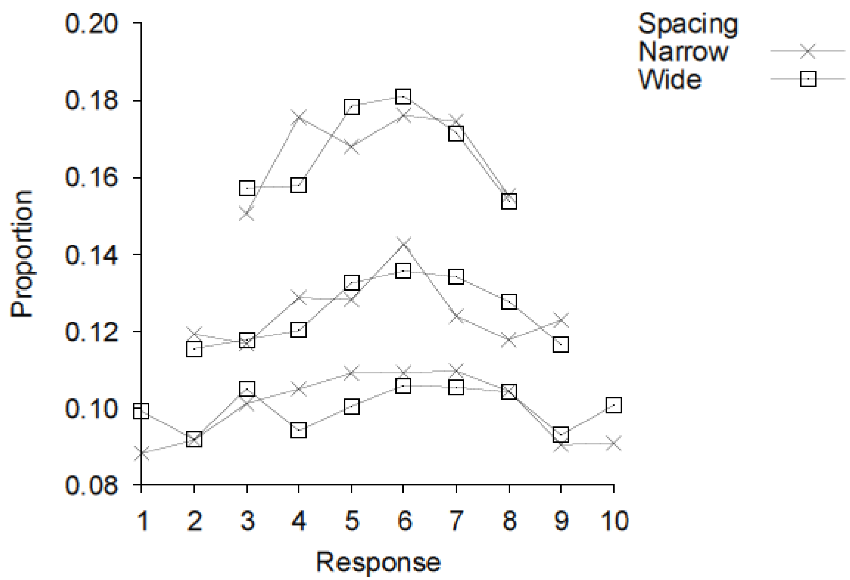
2. SCALING AND CODING

i. SCALE INVARIANCE

ii. ABSOLUTE vs RELATIVE CODING

NO ABSOLUTE CODING OF MAGNITUDES

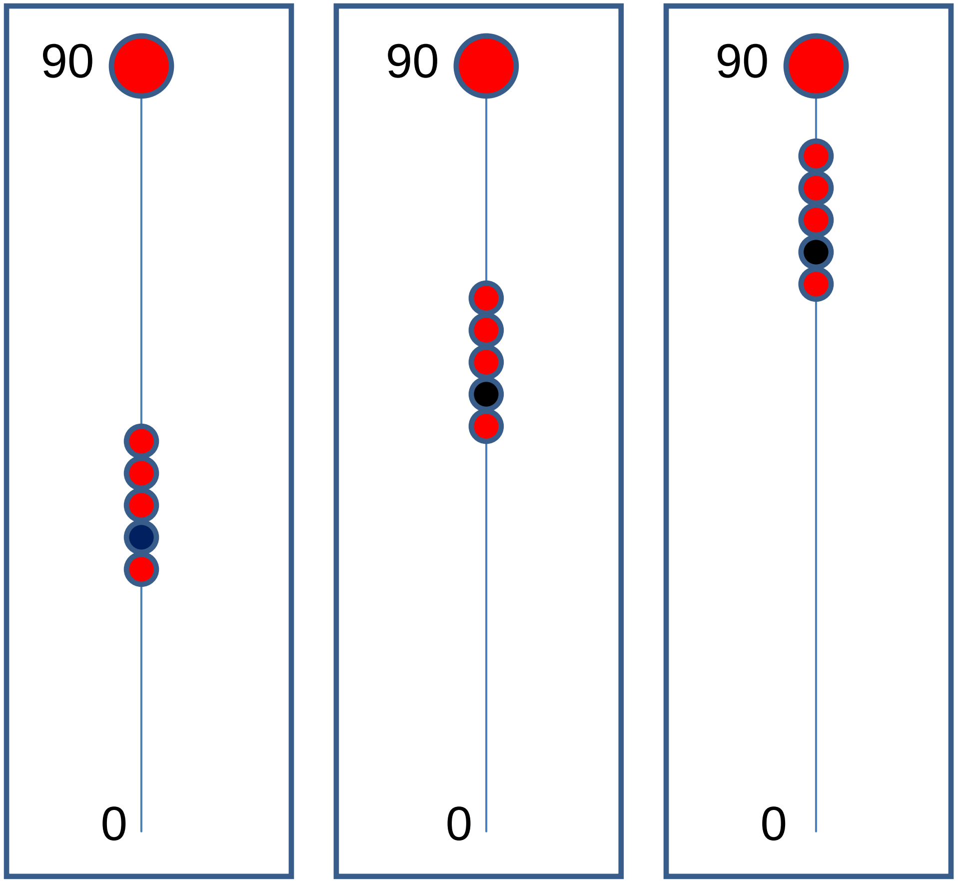
- Absolute identification
  - Limit of 5 items, independent of spacing



Wide vs narrow spacing (X2), pure tones, Stewart, Brown & Chater, 2005, Psych Rev

NO STABLE RATIO JUDGEMENTS

- Garner:
  - Asks people to halve loudness of 90Db auditory input
  - Range options between 50-60, 60-70, 70-80 Db
  - Choose within the *range of options*



PROSPECT RELATIVITY: PEOPLE HAVE NO STABLE RISK-PREFERENCE

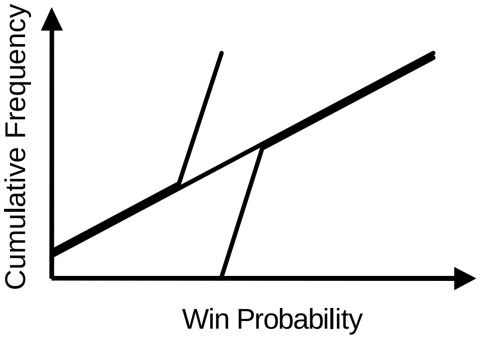
3 experimental conditions

All	*Risky*	*Safe*
.95 chance of £5	.95 chance of £5	
.90 chance of £10	.90 chance of £10	
.85 chance of £15	.85 chance of £15	
.80 chance of £20	.80 chance of £20	
.75 chance of £25	.75 chance of £25	
.70 chance of £30		.70 chance of £30
.65 chance of £35		.65 chance of £35
.60 chance of £40		.60 chance of £40
.55 chance of £45		.55 chance of £45
.50 chance of £50		.50 chance of £50

Stewart, Chater, Stott & Reimers, J Exp Psych: General, 2004

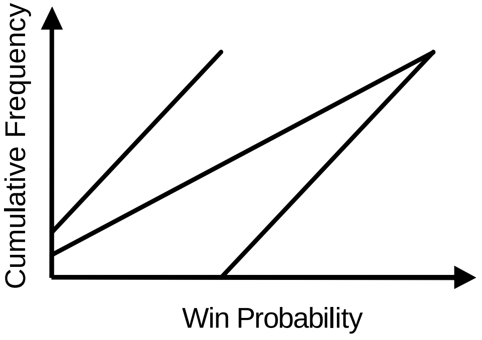
PREDICTIONS

- Stable risk aversion



a

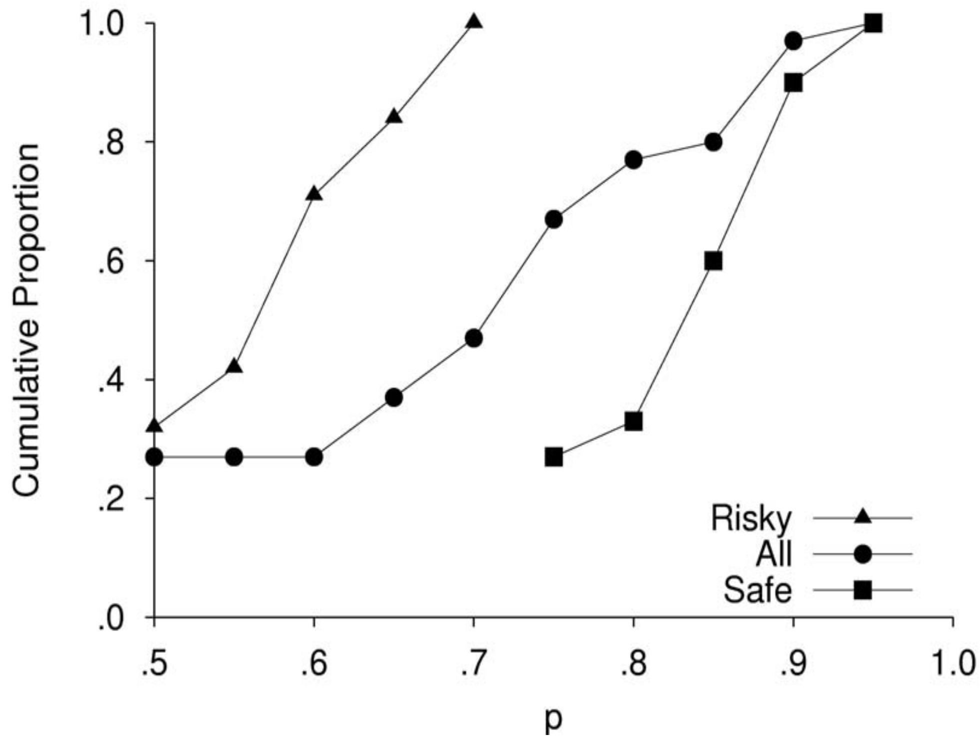
- Unstable risk aversion (DbS)



b

CHOICES STRONGLY INFLUENCED BY RANGE OF OPTIONS AVAILABLE

(CF GARNER ON PSYCHOPHYSICS)



Riskiness of gambles judged relative to other items (i.e., the 'sample')

NO UNDERLYING SCALES → NO INTEGRATION

No underlying "psychoeconomic" scales for

- Utility
- Subjective probability
- Time
- ...

- No stable trade-offs between different types of good
- No "cost-benefit" analysis
- No stable monetary valuations (e.g., of pains or pleasures)

Relates to Gigerenzer et al.s one-reason decision making;

Shafir et al.s reason-based choice;

Decision by Sampling Stewart, Chater, & Brown (2006) (Day 6)

3. INFERENCE

i. SIMPLICITY

ii. GENERATIVE VS DISCRIMINATIVE

THE SIMPLICITY PRINCIPLE

- Find explanation of “data” that is as simple as possible
  - An ‘explanation’ *reconstructs* the input
  - Simplicity measured in code length
  - Mimicry theorem with Bayesian inference (e.g., Chater, 1996, *Psych Review*; “deep” analysis by Li & Vitányi, 1997, 2009)
  - Some connections to Statistical Learning Theory (Vapnik, 1995), but link not generally well-understood

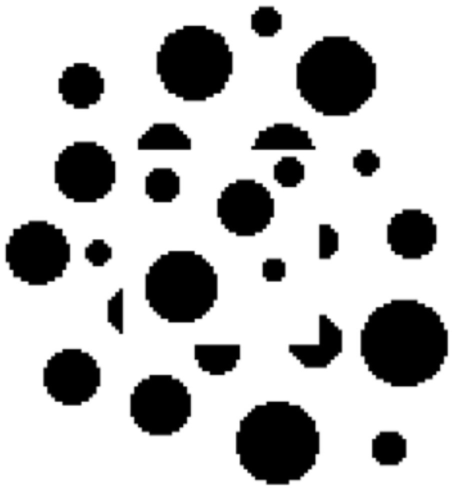
SIMPLICITY AS “IDEAL” INDUCTIVE METHOD

- Deep mathematical theory: Kolmogorov complexity theory
  - Li & Vitányi, 1993, 1997, 2009
- Predicting using simplicity converges on correct predictions
  - Solomonoff, 1978
- Scaled-down to generate a non-standard statistical theory
  - minimum message length, e.g., Wallace & Boulton, 1968;
  - minimum description length, e.g., Rissanen, 1989
- And applicable to Machine Learning
  - Grunwald, 2007

SIMPLICITY HAS BROAD SCOPE

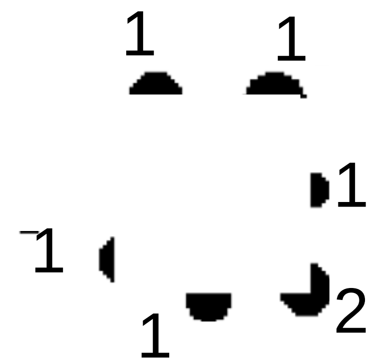
Domain	*Principle*	*References*
Perceptual organization	Favour simplest interpretation	Koffka, 1935; Leeuwenberg, 1971; Attneave & Frost, 1969;
Early vision	Efficient coding & transmission	Blakemore, 1990; Barlow, 1974; Srivinisan, Laughlin
Causal reasoning	Find minimal belief network	Wedelind
Similarity	Similarity as transformational complexity	Chater & Vitányi, 2003; Hahn, Chater & Richardson, 2003
Categorization	Categorize items to find shortest code	Feldman, 2000; Pothos & Chater, 2002
Memory storage	Shorter codes easier to store	Chater, 1999
Memory retrieval	Explain interference by cue→trace complexity	Rational foundation SIMPLE (Brown, Neath & Chater, 2005)
Language acquisition	Find grammar that best explains child's input (NB Day 6)	Chomsky, 1955; J. D. Fodor & Crain; Chater, 2004; Chater & Vitányi, 2005; Hsu et al.

OBSERVATIONS MAY SUGGEST GENERAL PRINCIPLES – E.G., FAVOUR THE SIMPLEST EXPLANATION

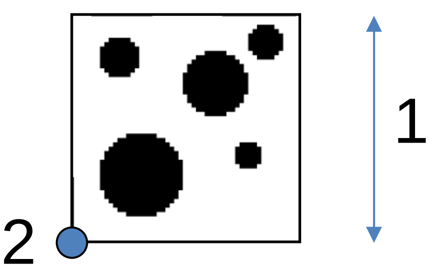


Kanizsa

7

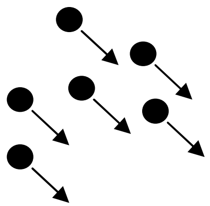


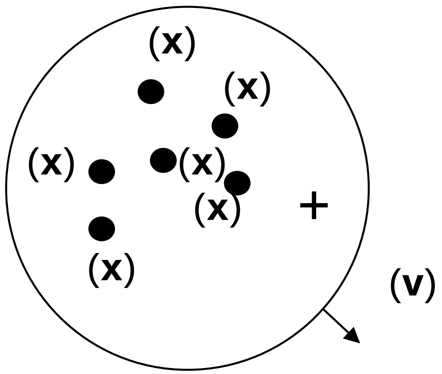
3



Find simple abstract patterns… e.g., postulating a square needs 3 parameters; simpler than 7 parameters for accounting for ‘cuts’ in circles separately

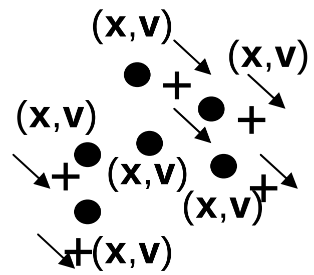
LONG TRADITION OF SIMPLICITY IN PERCEPTION
(MACH, KOFFKA, LEEUWENBERG) E.G., GESTALT LAWS





Grouped

6 + 1 vectors



Ungrouped

6 x 2 vectors

COMMON FATE – THINGS THAT MOVE TOGETHER ARE GROUPED TOGETHER

3. INFERENCE

i. SIMPLICITY

ii. GENERATIVE VS DISCRIMINATIVE

HOW MUCH OF COGNITION IS REVERSIBLE?

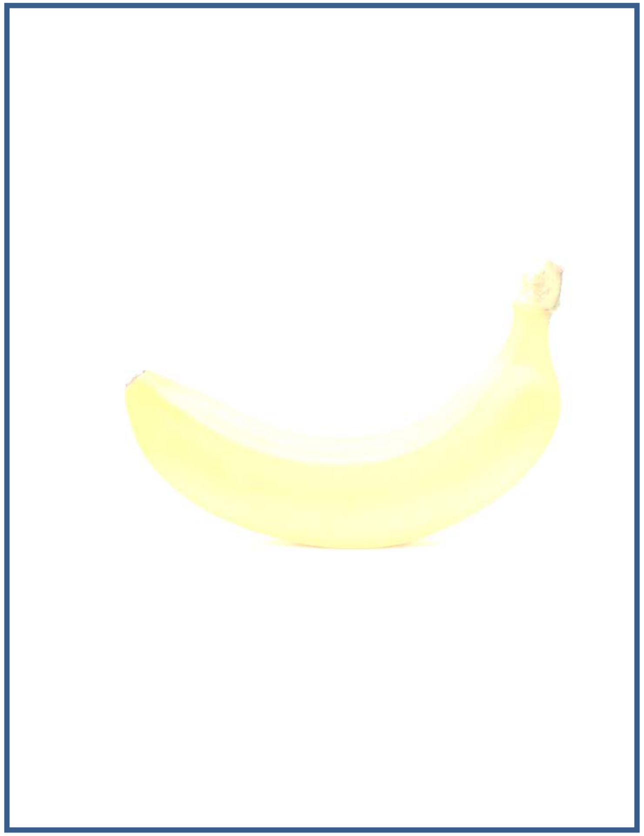
For each cognitive mapping from A to B,
there often is a corresponding mapping from B to A

- Perception
- Language production
- Memory encoding

- Imagery
- Language comprehension
- Memory retrieval

EVIDENCE

- A terribly designed but fun uncontrolled experiment!
- Perky (1910) projected patch of colour onto the back of a translucent projection screen, while asking people to image, e.g., a banana



EVIDENCE

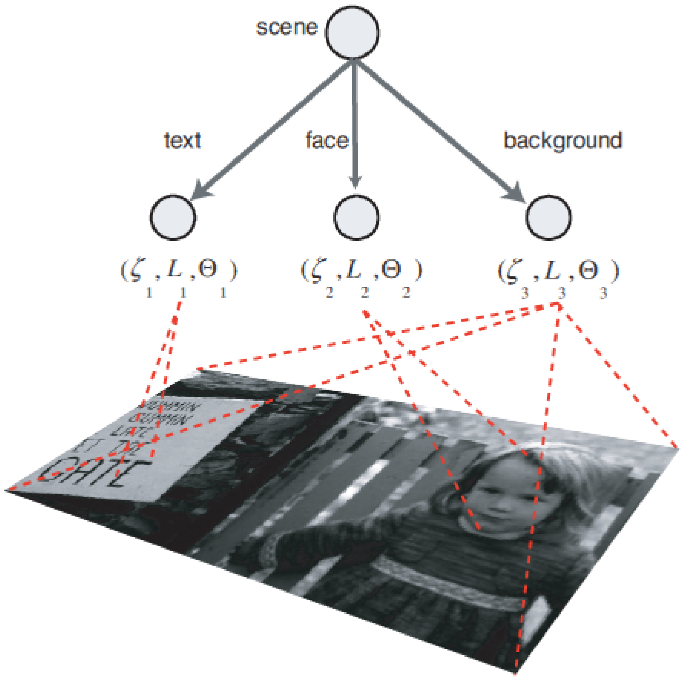
- Neuroscience evidence
  - Brain imaging---same areas for perception-imagery etc
  - Impact of brain injury
    - Lose e.g., colour vision and colour imagery in tandem

- Cognitive evidence
  - Learning seems to transfer
    - learning to understand a new word;
    - learning to produce it
  - Interference between imagery and perception
  - Subtle perceptual effects replicate in imagery

e.g, Ganis, Thompson, Mast & Kosslyn (2004). Chapter 67, The Cognitive Neurosciences III. MIT Press

EXPLANATION?

- Mappings via *models* of the world
- E.g., Bayesian generative models of perception
  - Pr(*World*|*Image*)
    - from
  - Pr(*Image*|*World*)
    - Via Bayes theorem

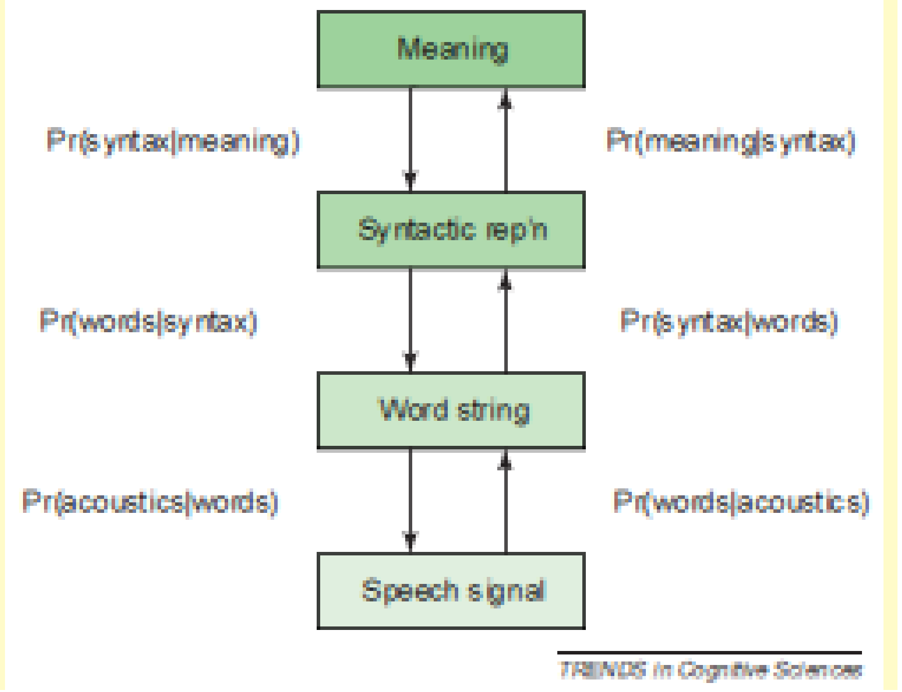


Cf. Generative vs discriminative statistical/perceptual models
(Griffiths & Yuille, 2006; picture from Yuille & Kersten, 2006)

SIMILARLY FOR LANGUAGE

- Mappings via *models* of the language
- E.g., Bayesian generative models of perception
  - Pr(*Meaning*|*Speech*)
    - Via
  - Pr(*Speech*|*Meaning*)

(Chater & Manning, TICS, 2006)



AND GENERATIVE VS DISCRIMINATIVE *INSTRUCTIONS* MAY CHANGE PEOPLE'S CATEGORIZATIONS

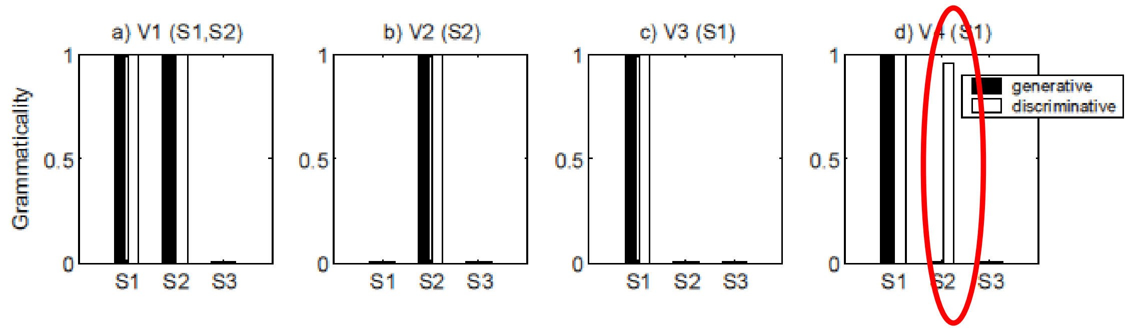


Figure 2: Predicted grammaticality judgments from generative and discriminative models. In parentheses next to the verb index in the title of each plot is the sentence structure(s) that were shown to be grammatical for that verb in the training corpus.

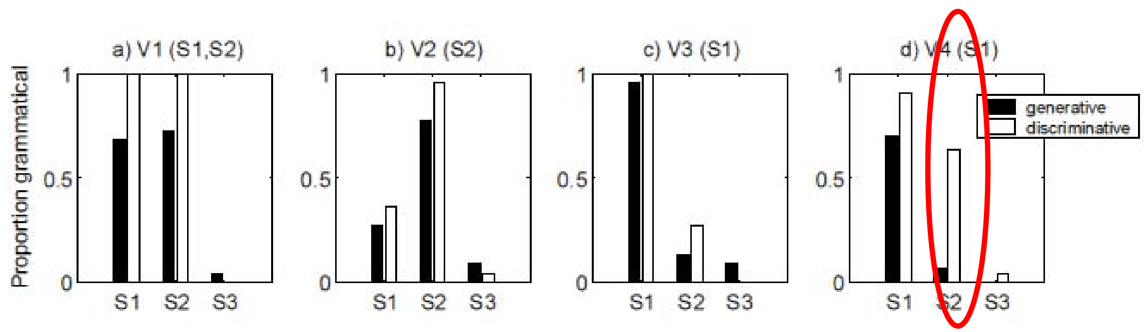


Figure 3: Human grammar judgments, showing proportion grammatical for each sentence structure.

Hsu & Griffiths, NIPS, 2009,

4. ARCHITECTURE

MODULARITY VS UNIFIED SYSTEM

MODULARITY OF MIND
VS. UNIFIED SINGLE SYSTEM?

- Fodor (1983)
- Module = System which is informationally encapsulated from general cognition, and other modules
  - Perceptual processes?
  - Motor control?
  - Language processing?
  - Learning processes

- Associated with
  - Special neural hardware/brain localization
  - Computational autonomy
  - Little attentional control
  - Genetic basis

CONVERSELY, “CENTRAL” PROCESSES CANNOT BE ISOLATED…

- The realm of central processes is typically assumed to be the realm of belief-desire explanation
  - Any thought or behaviour can potentially be ‘countermanded’ by new information
  - And this new information may be arbitrarily ‘distant’ (outside the module)

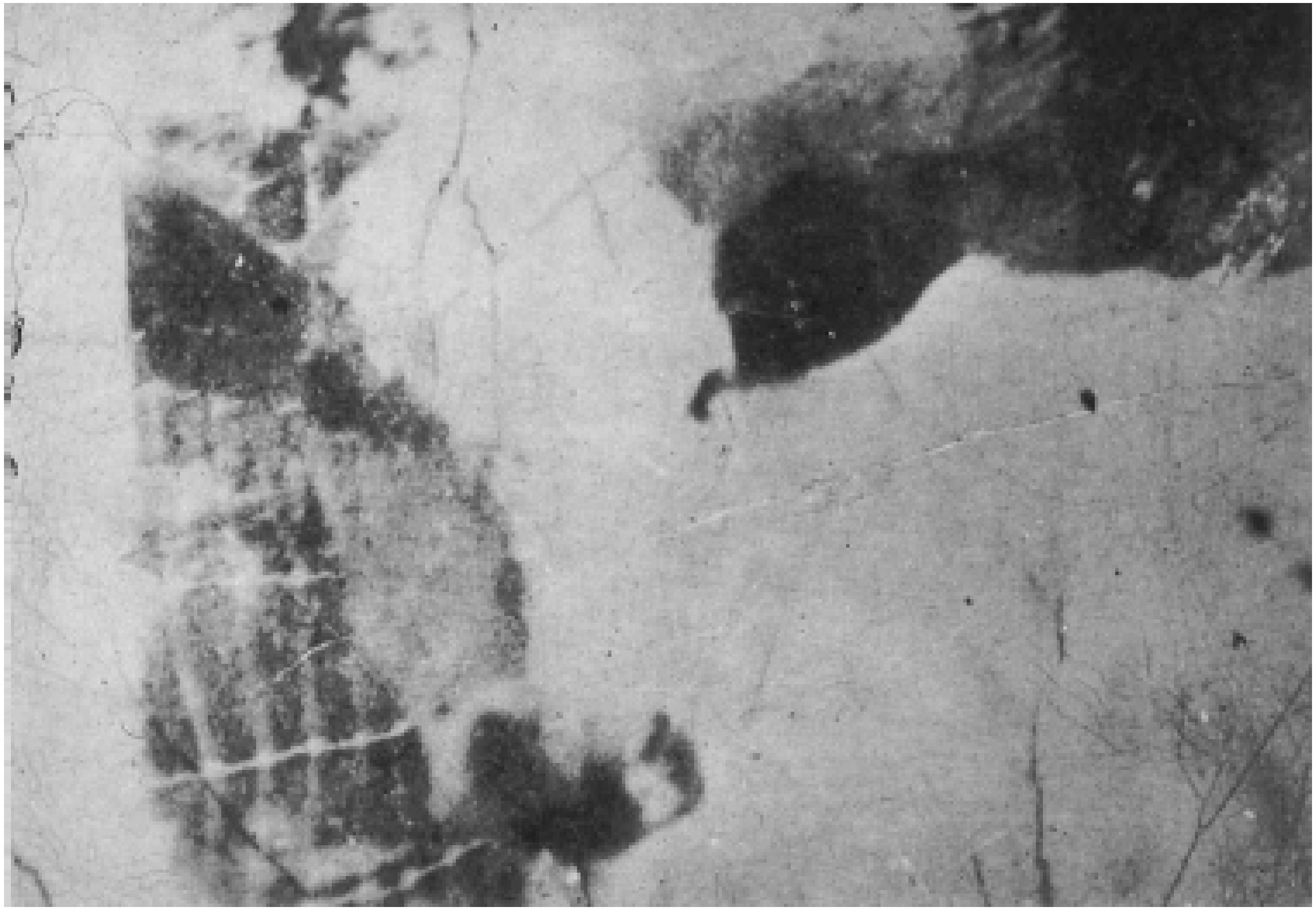
COGNITIVE PENETRABILITY (PYLYSHYN, 1984) AS A KEY TEST: SENSITIVITY TO ARBITRARY INFORMATION

MODULARITY IS THEORETICALLY CENTRAL

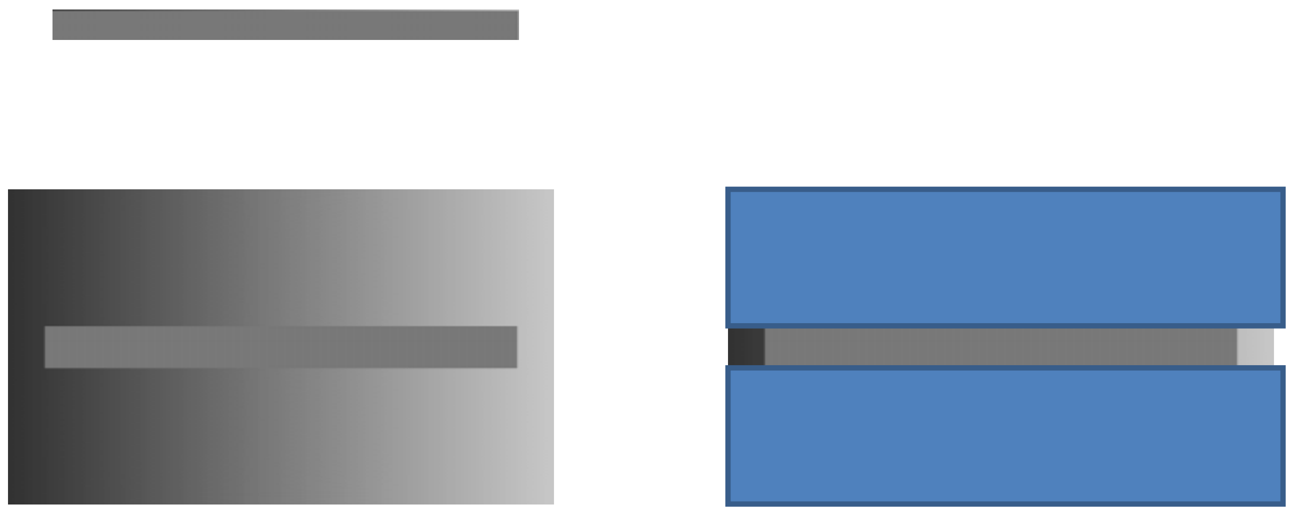
- Decomposing a complex system into its parts is central to reductionist explanation
- Which parts?
- Is cognition decomposable at all?
- If not, is cognitive science feasible at all?
  - Fodor, 1983

HOW MUCH CAN HIGH-LEVEL INFORMATION AFFECT PERCEPTION?

DALLENBACH’S COW



BUT MANY ASPECTS OF VISION ARE NOT COGNITIVELY PENETRABLE

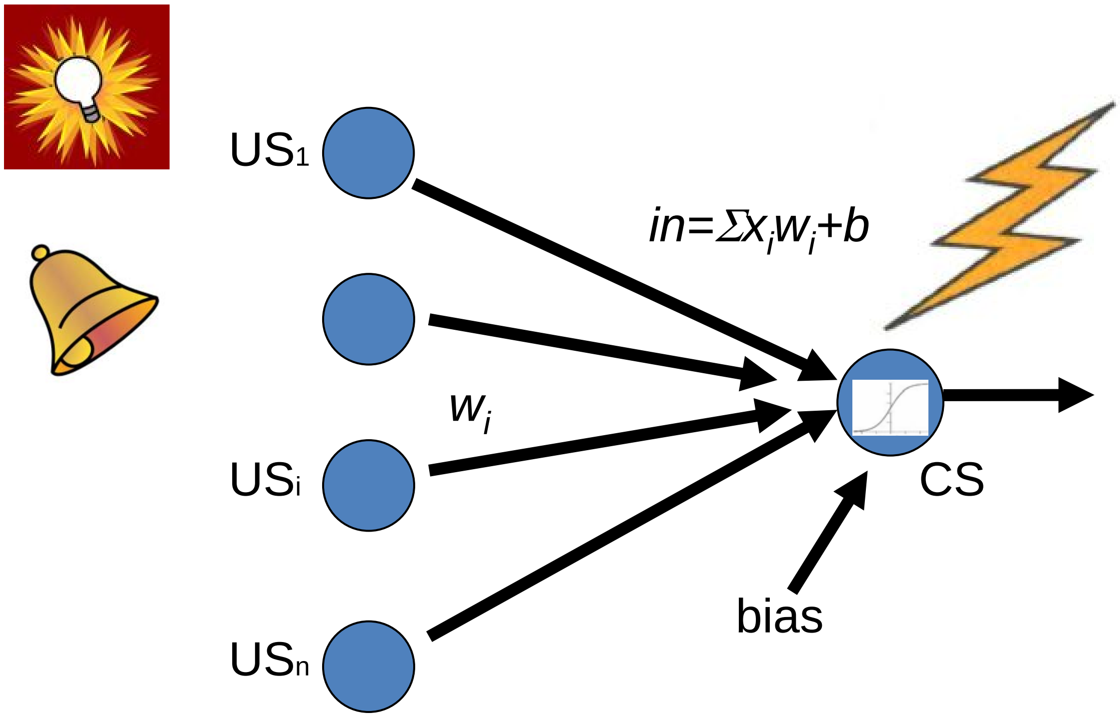


NO AMOUNT OF “EVIDENCE” OR ARGUMENT ELIMINATES THE ILLUSION

FOCUS ON LEARNING: THE CASE OF CONDITIONING IN HUMANS

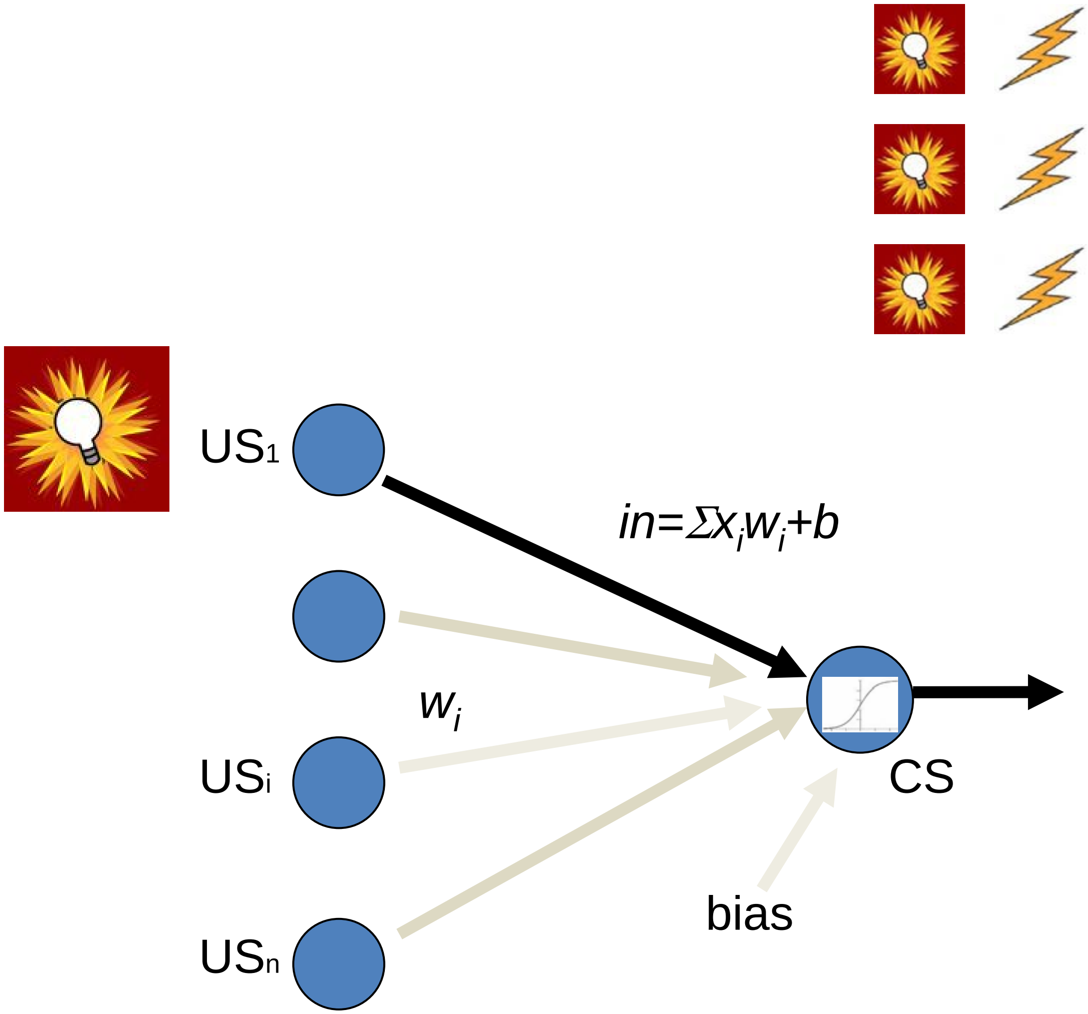
- Conditioning often viewed as resulting from a basic learning mechanism or module
- Rats and pigeons condition
- Potentially drastic implications for viability of cognitive science (see next time…)
- Only possible for modular processes (Fodor)
- Cognitively impenetrable processes (Pylyshyn)

CLASSICAL CONDITIONING IN COMPUTATIONAL TERMS

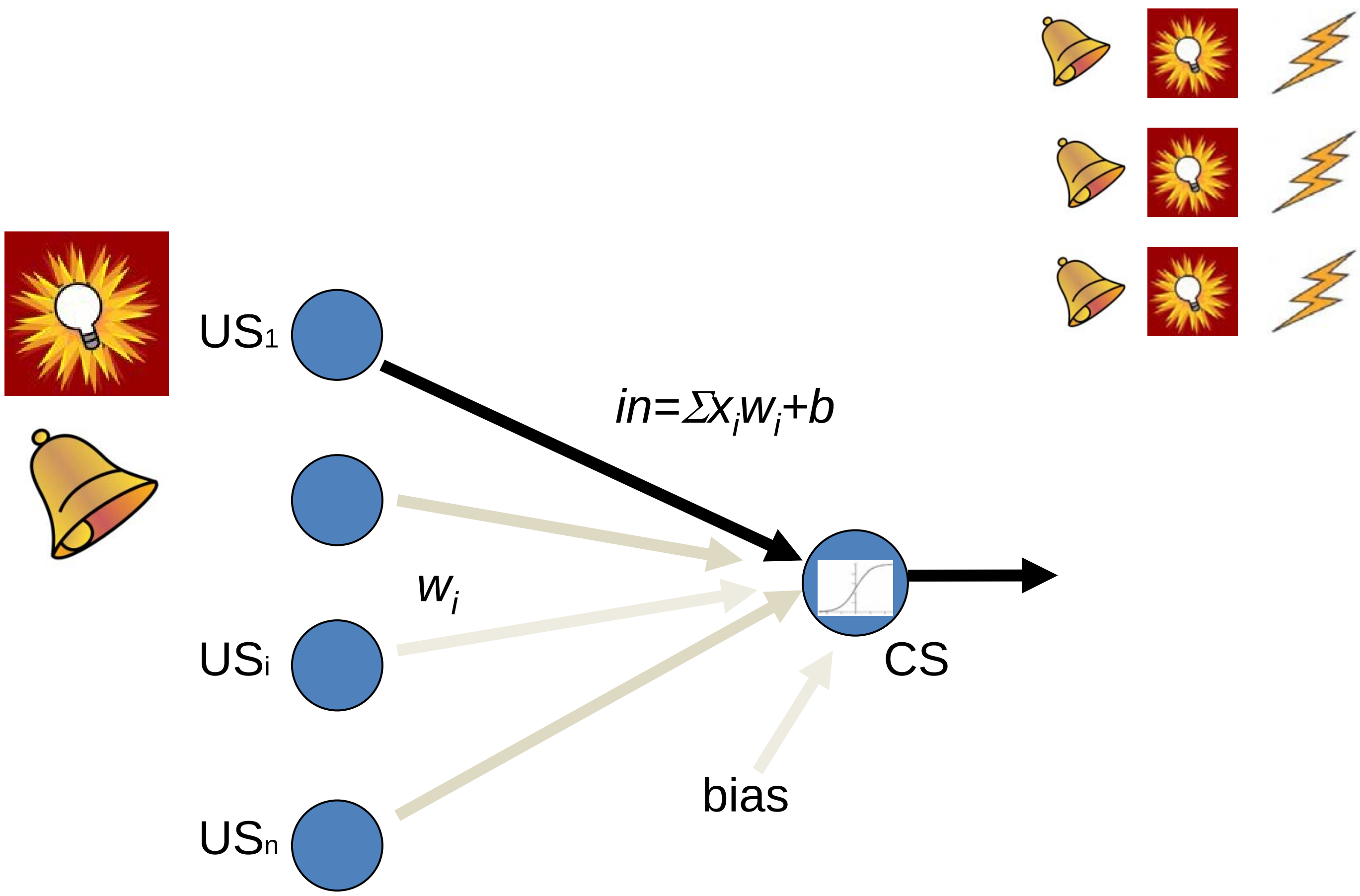


HEBBIAN OR ERROR-DRIVEN LEARNING?

REMINDER: KAMIN BLOCKING : TRAINING PHASE 1



REMINDER: KAMIN BLOCKING: TRAINING PHASE 2



NO ERROR, SO NO FURTHER LEARNING

ARE EXPECTATIONS REALLY A PRODUCT OF GENERAL COGNITION, NOT A SPECIFIC LEARNING RULE?

- Train stimulus → shock
- Measure GSR
  - Now tell people "I've disconnected the electrodes"
- 1. GSR immediately reduces sharply
- 2. and in proportion to the degree that they belief you!
  - Bridger & Mandel, JEP, 1965

- Conditioning requires attention
- Shock when hear, say, 'animal' words
- Dichotic listening; Attend to one channel
- Conditioning only when animal words are in the attended channel

See Brewer, 1974; Shanks & Lovibond, 2002; Mitchell, de Houwer & Lovibond, 2009

SO TWO VERY DIFFERENT VIEWS OF CONFLICT, ADDICTION, WEAKNESS OF WILL

- Conditioning system a module parallel to, and sometimes in opposition to, the conscious, explicit system

- Clash of mechanisms

- Unitary system
- Different 'probes'/'tasks' will get different outputs

- Clash of reasons

Chater (2009) *Cognition*

3. SUMMARY AND IMPLICATIONS

PRINCIPLE-BASED COGNITIVE MODELLING

AIM: A PRINCIPLE-BASED APPROACH TO REVERSE ENGINEERING COGNITION

- Specify general principles
- Embody principles in a model of a specific cognitive domain
- Review and collect experimental data
- Evaluate/revise model
- Evaluate/revise general principles

THE MODELLING CYCLE

AIM: A PRINCIPLE-BASED APPROACH TO REVERSE ENGINEERING COGNITION

- Machine learning has some powerful candidate principles, arising from functional considerations
  - Bayes
  - Kernel machines
  - Reinforcement learning

- Which need to be mapped into cognition using principles capturing empirical regularities
  - Scaling
  - Magnitude coding
  - Simplicity in perception
  - Generative vs discriminative models
  - Modularity

- To assess how and when various ML functional principles apply