Unifying principles in cognitive science

Constantin Rezlescu
20 November 2017
c.rezlescu@ucl.ac.uk

Unifying psychological principles

Three suggestions:

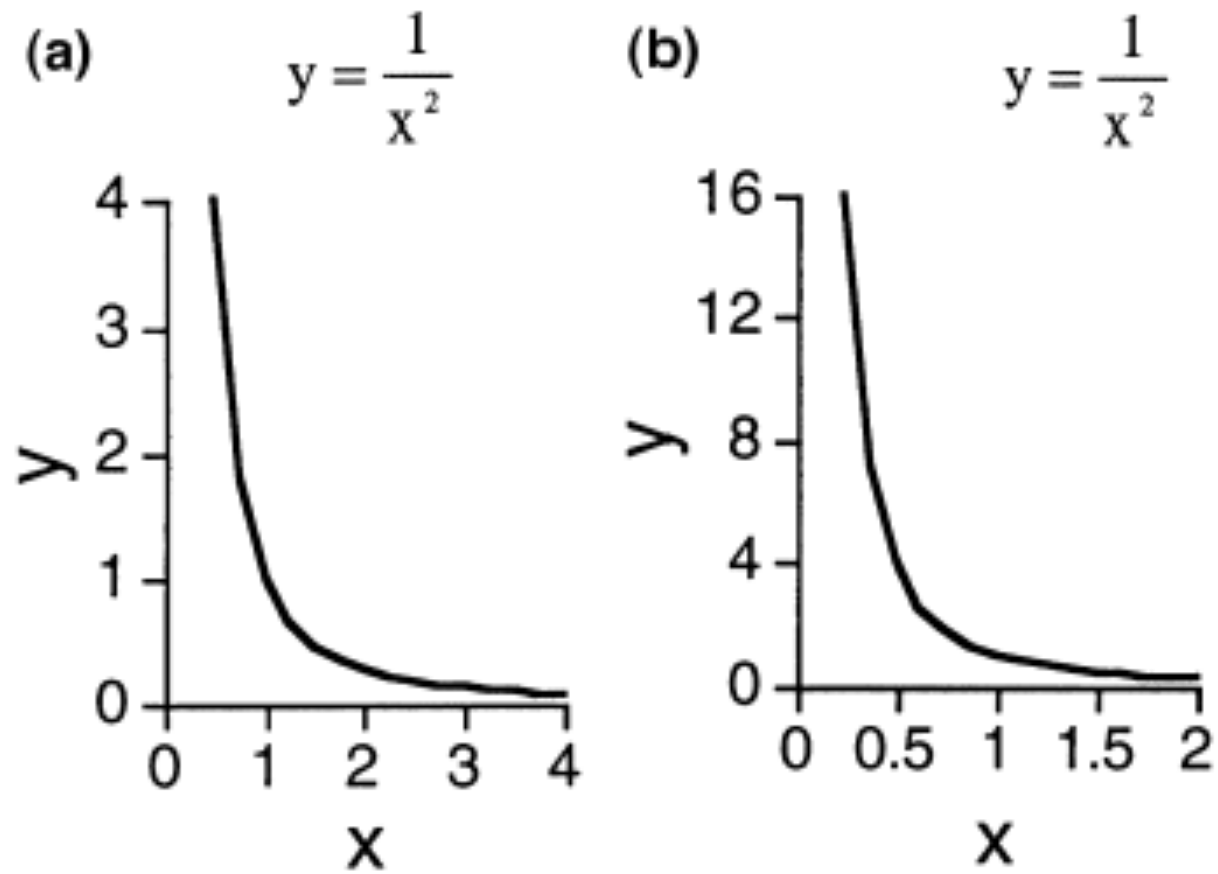
1. Scale invariance
2. No absolute coding of magnitude (Decision by sampling)
3. Common magnitude system for numbers, space and time (ATOM - A theory of magnitude)

1. Scale invariance

How tall is this wave?

What is scale invariance?

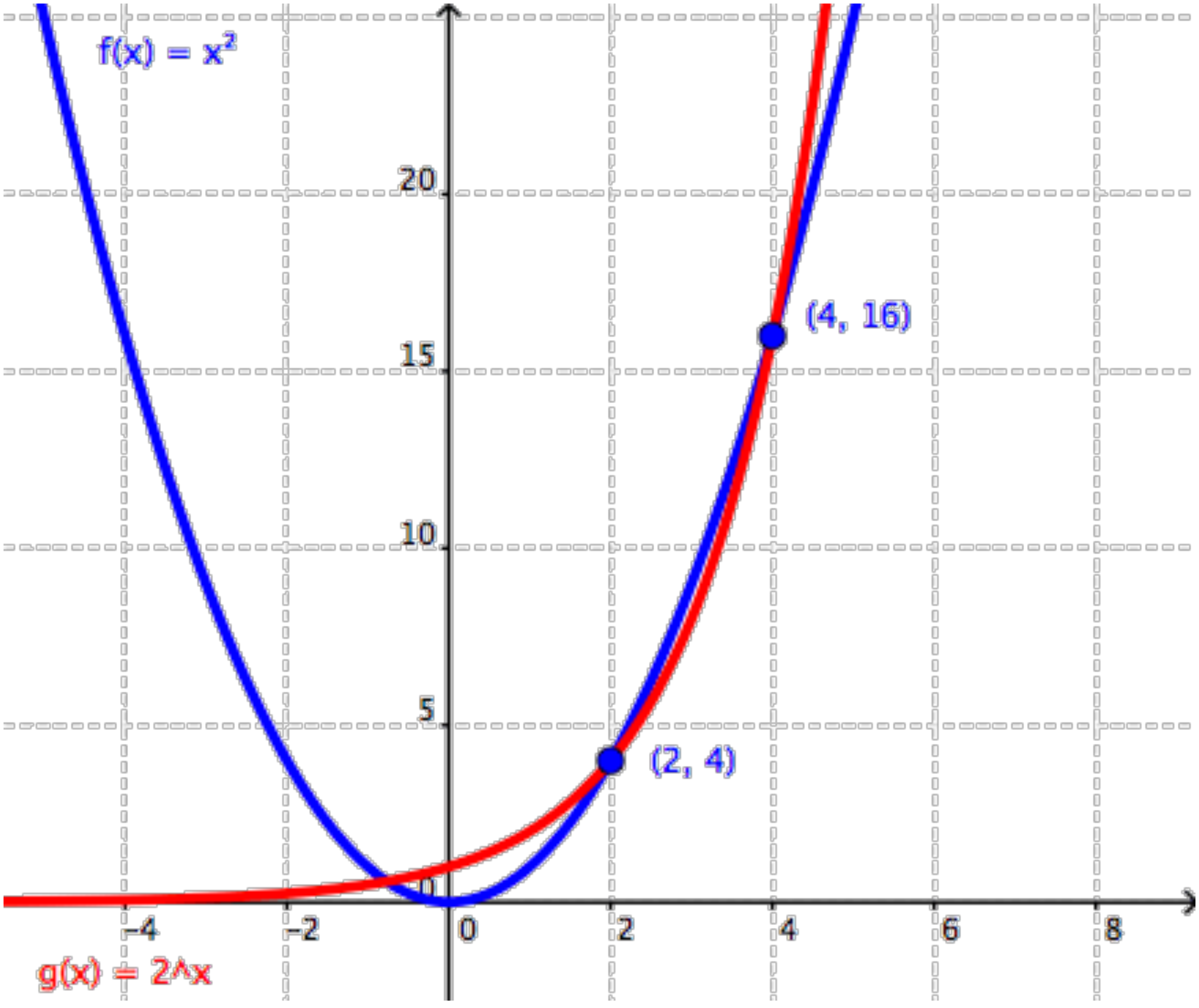
- In a nutshell:
    - throw away “units”
    - can you reconstruct them from your data?
- If not, phenomenon is scale-invariant



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

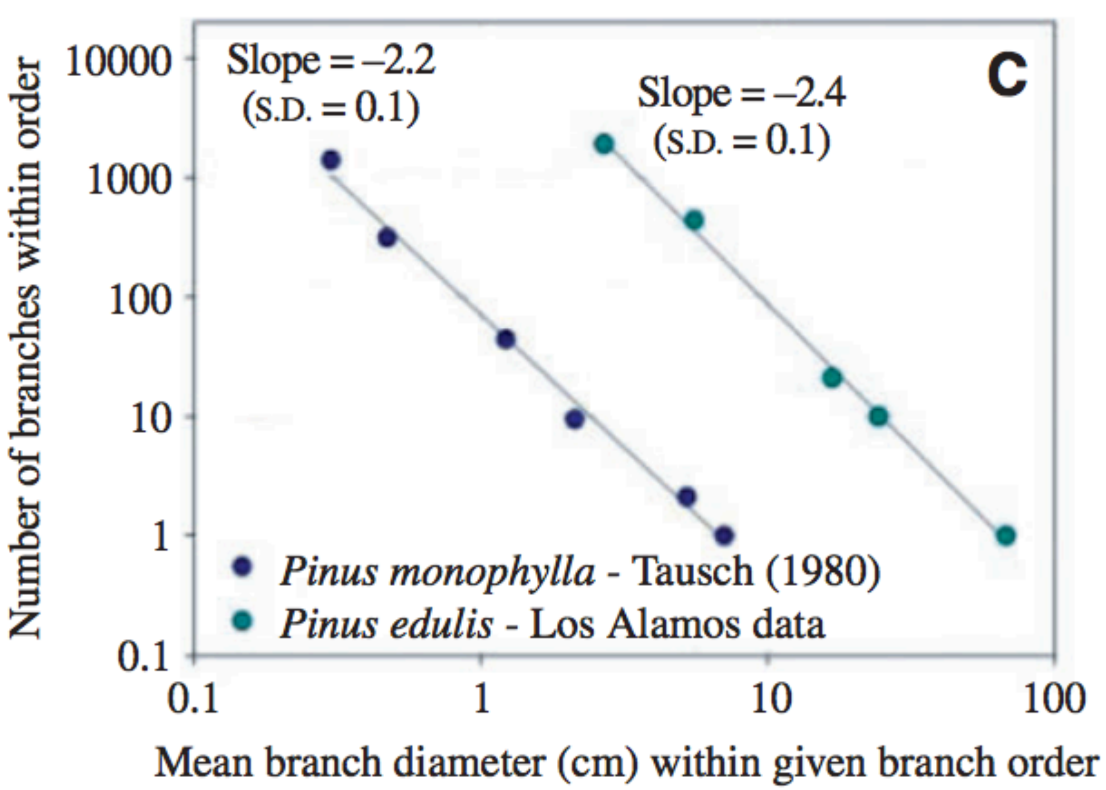
Chater & Brown (1999)

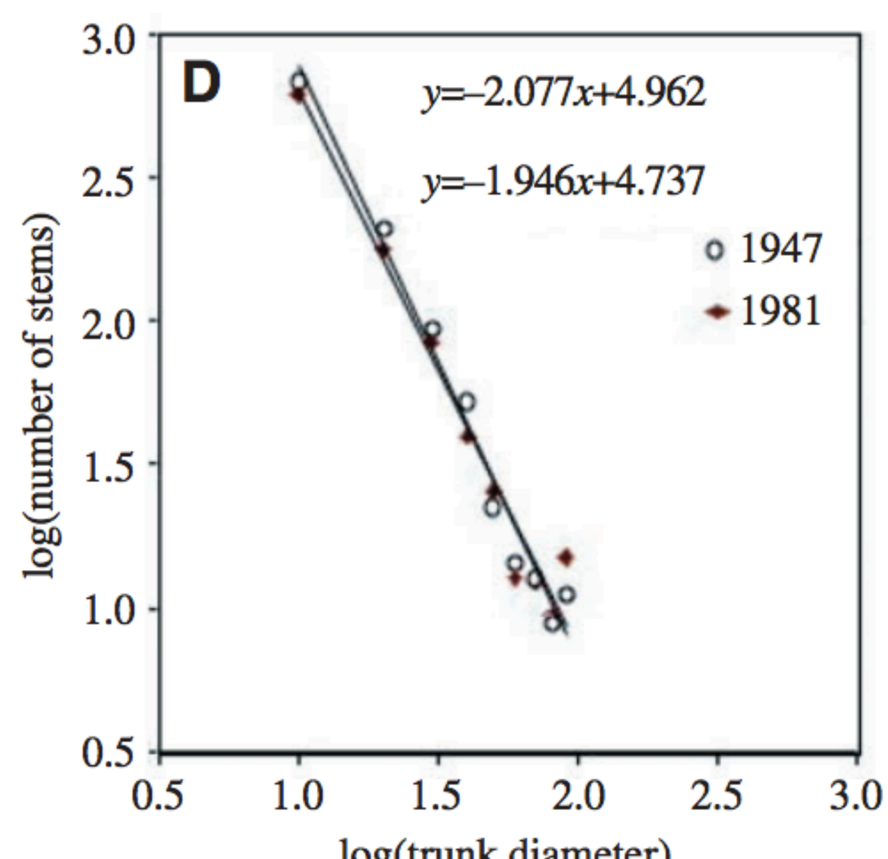
Power vs. exponential functions



Scale invariance is everywhere

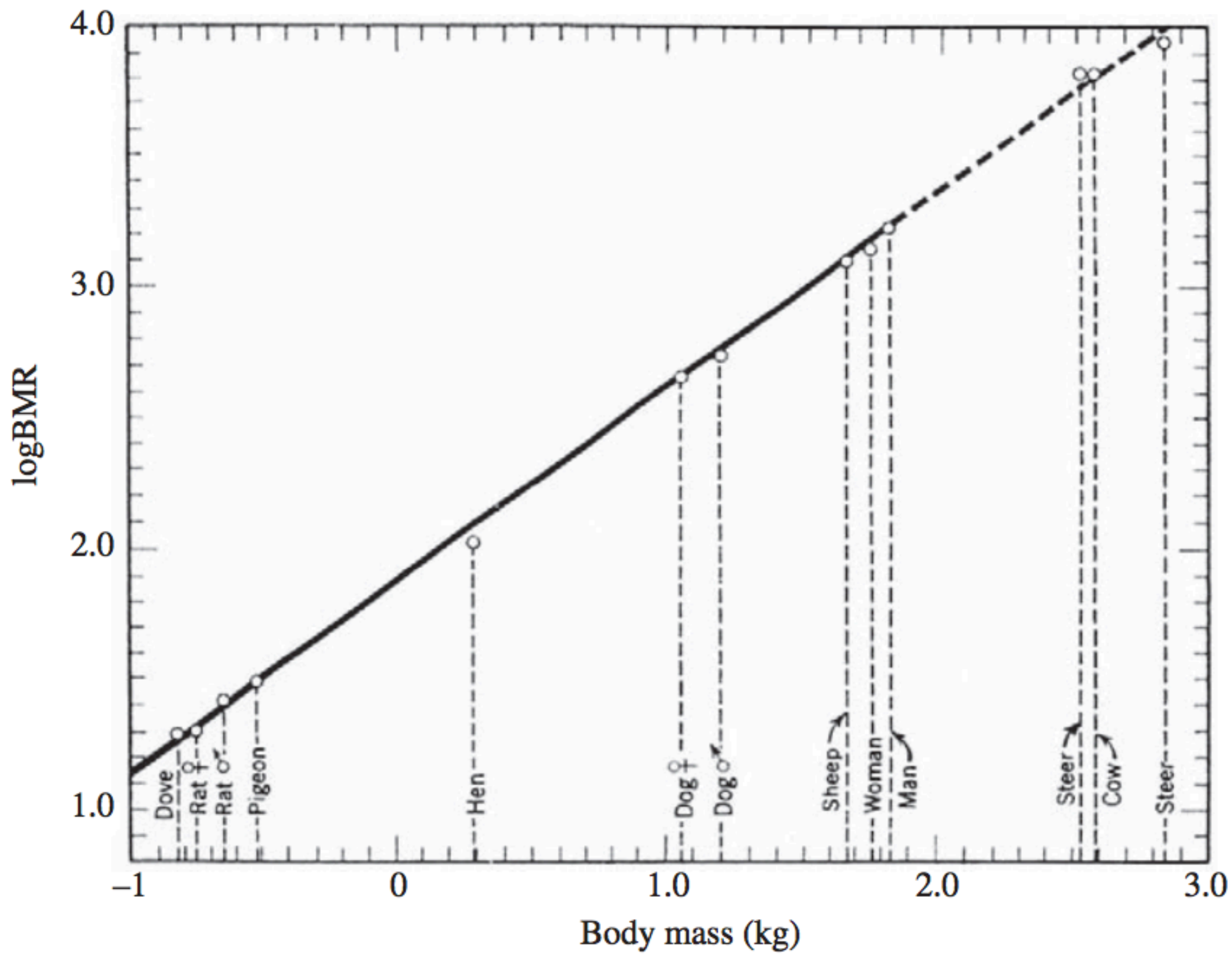
West & Brown, 2005





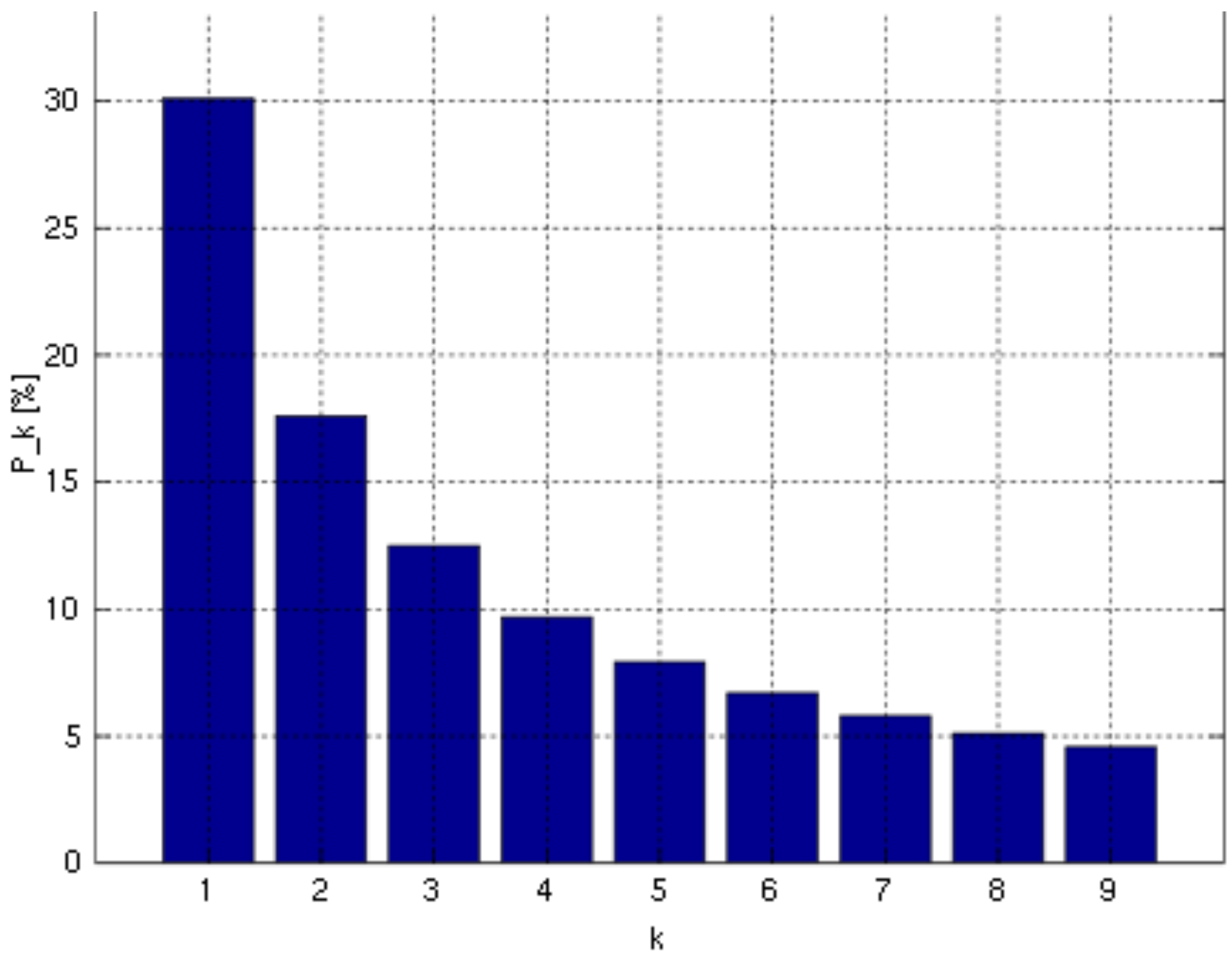
Scale invariance is everywhere

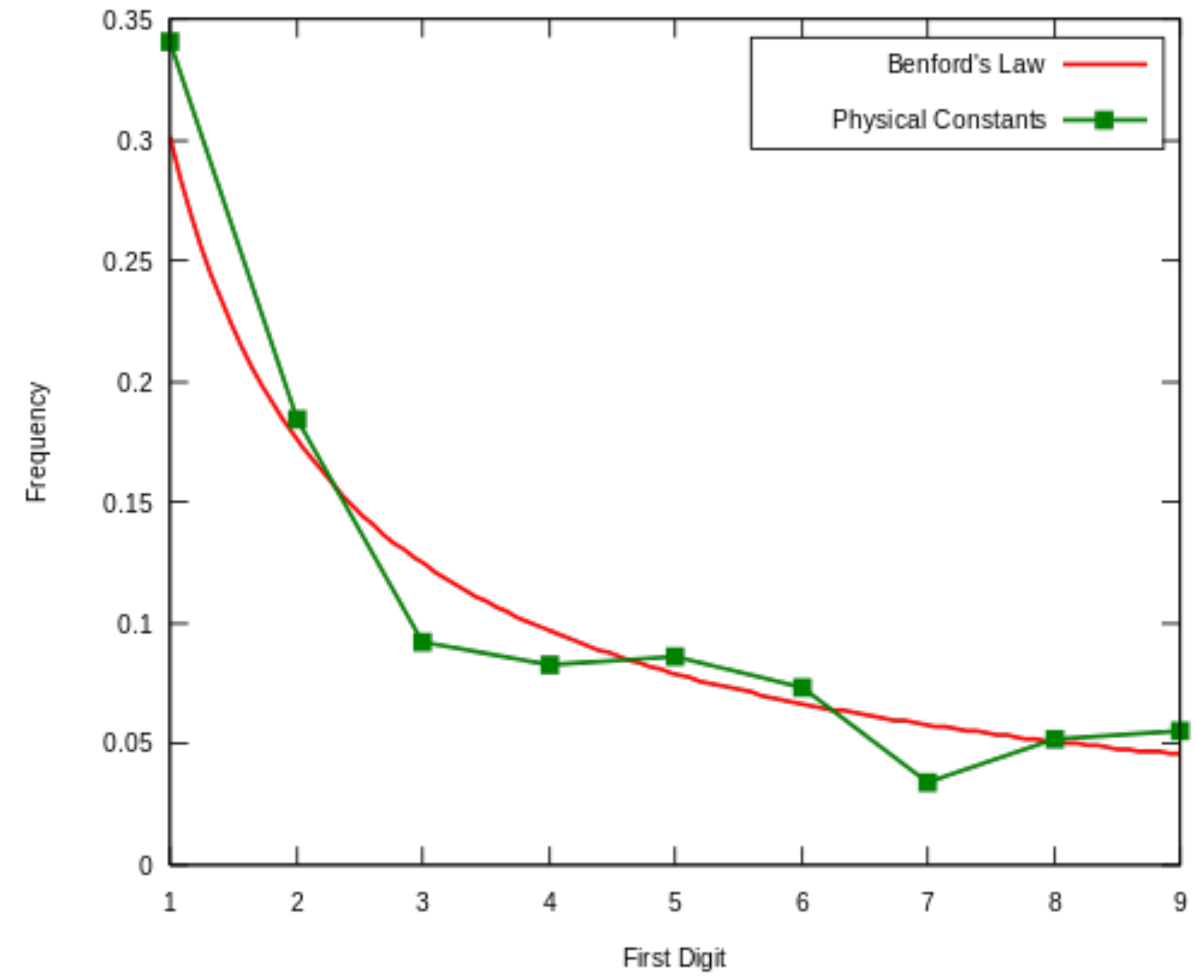
- Metabolic rate of mammals and birds vs. body mass (log-log scale)



Kleiber's Law
(Spence, 2009)

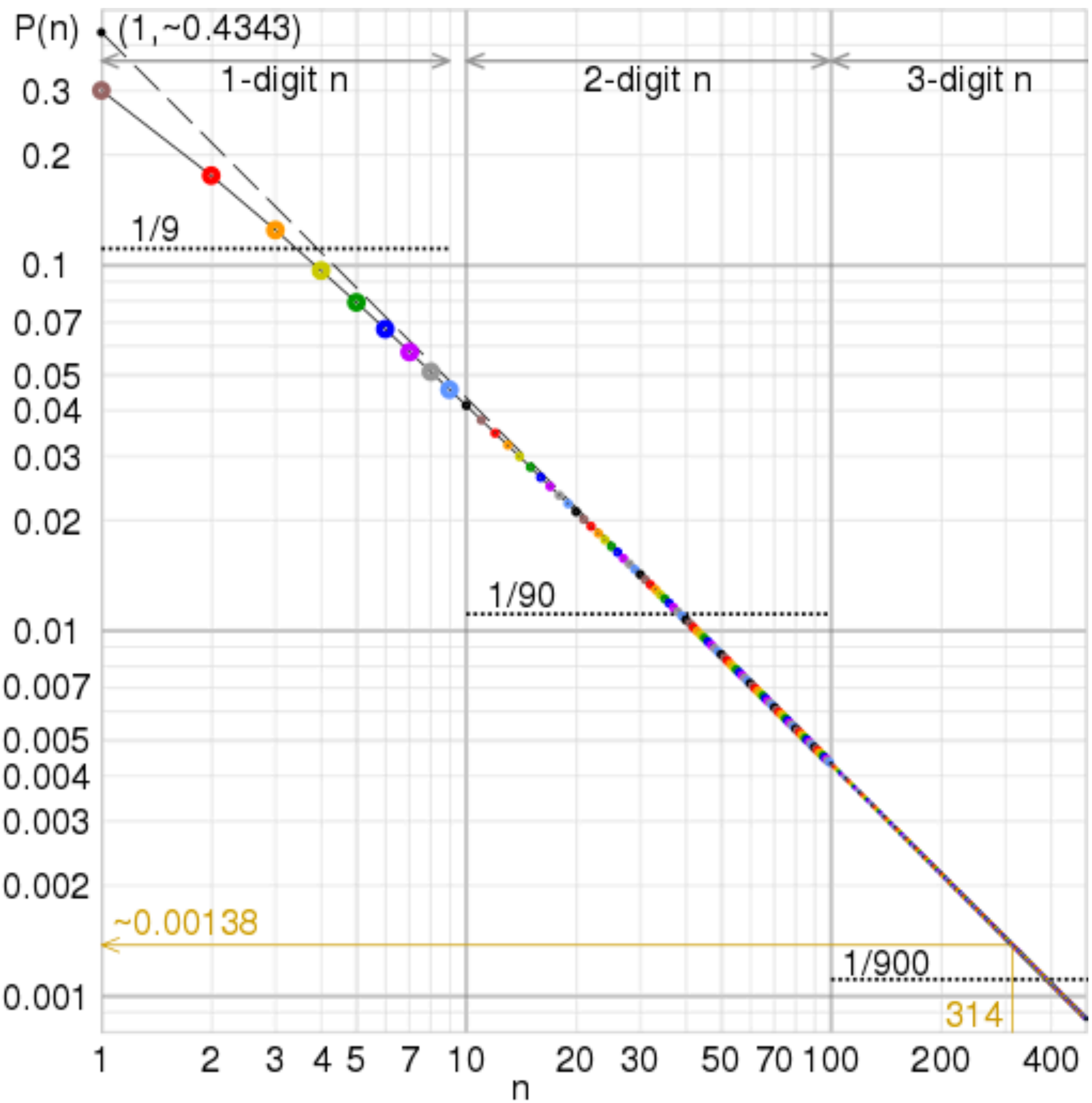
Distribution of digits (Benford's law)



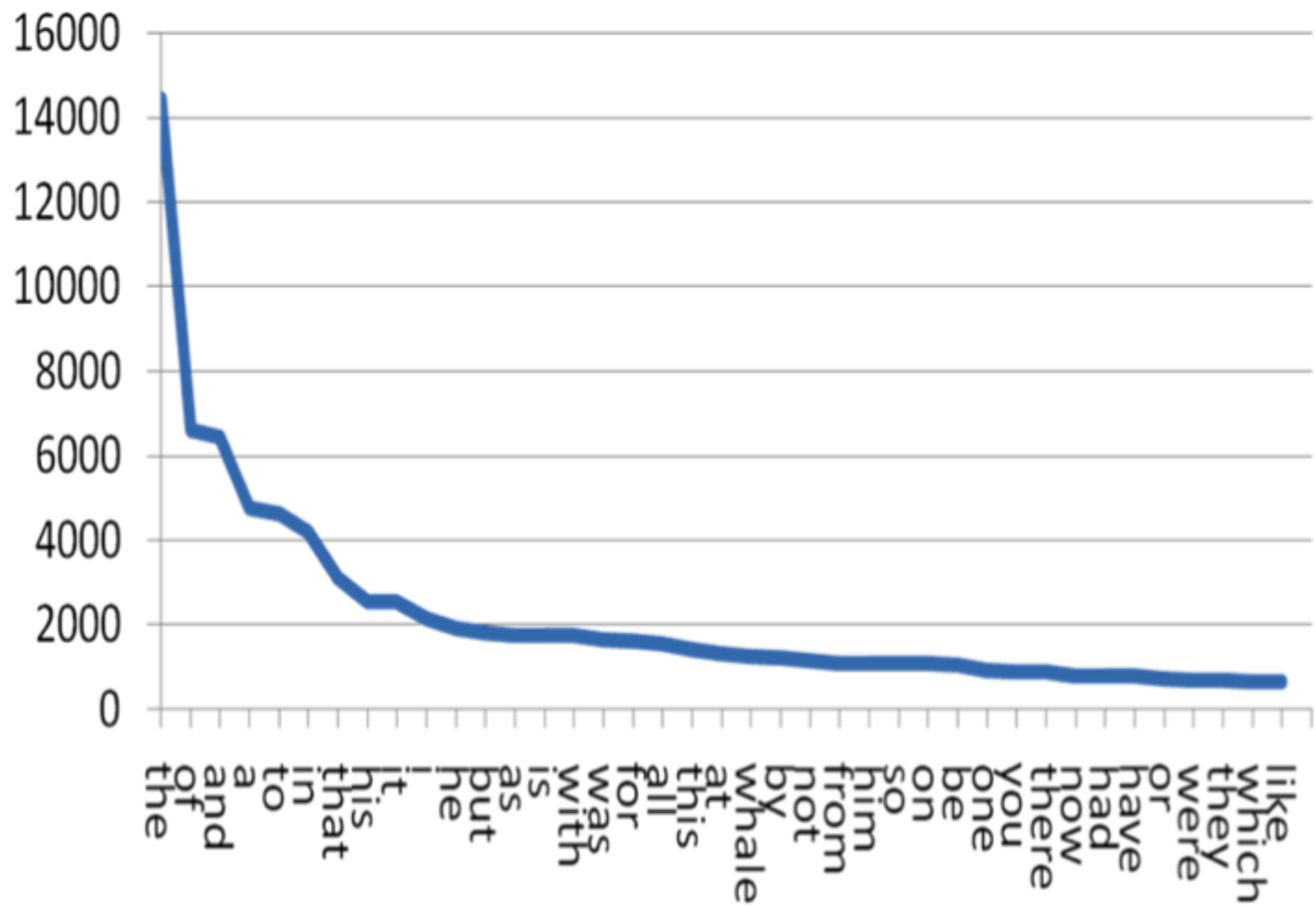


- electricity bills, street addresses, stock prices, house prices, population numbers, death rates, lengths of rivers

Distribution of digits (Benford's law)



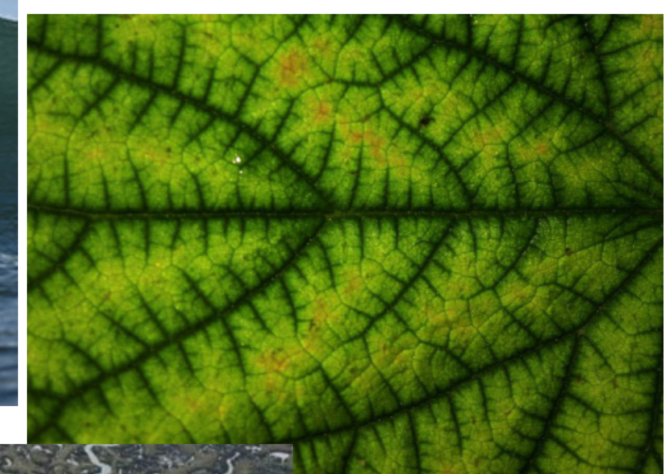
Word frequencies (Zipf's law)



Scale invariance is everywhere

Scale-invariance as a “null hypothesis” for the cognitive and brain sciences

This null hypothesis implies many well-known psychological laws…



- City sizes
- Company sizes
- River sizes
- Earthquakes
- ….

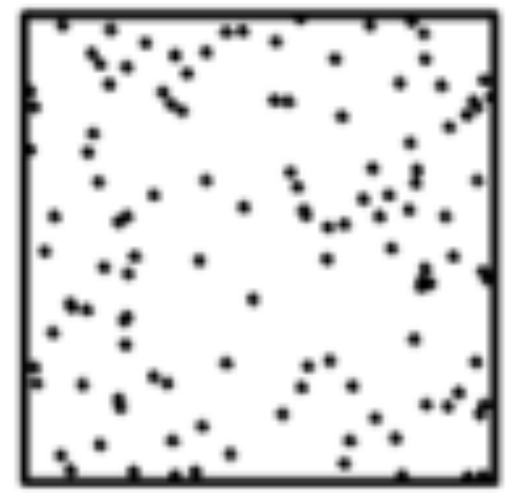
From scale-invariance to psychological "laws"

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)	$\Delta I/I$ µ $\Delta S/S$ 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)	$T=a+ b.log(\Delta D/D)$	Ratio preserving: time-precision
Luce Choice Rule		Ratios in "goodness" (however measured)
Herrnstein's matching law		Ratio preserving: Prob of choice

Chater & Brown (2008)

Which box has more dots?

110



120

10

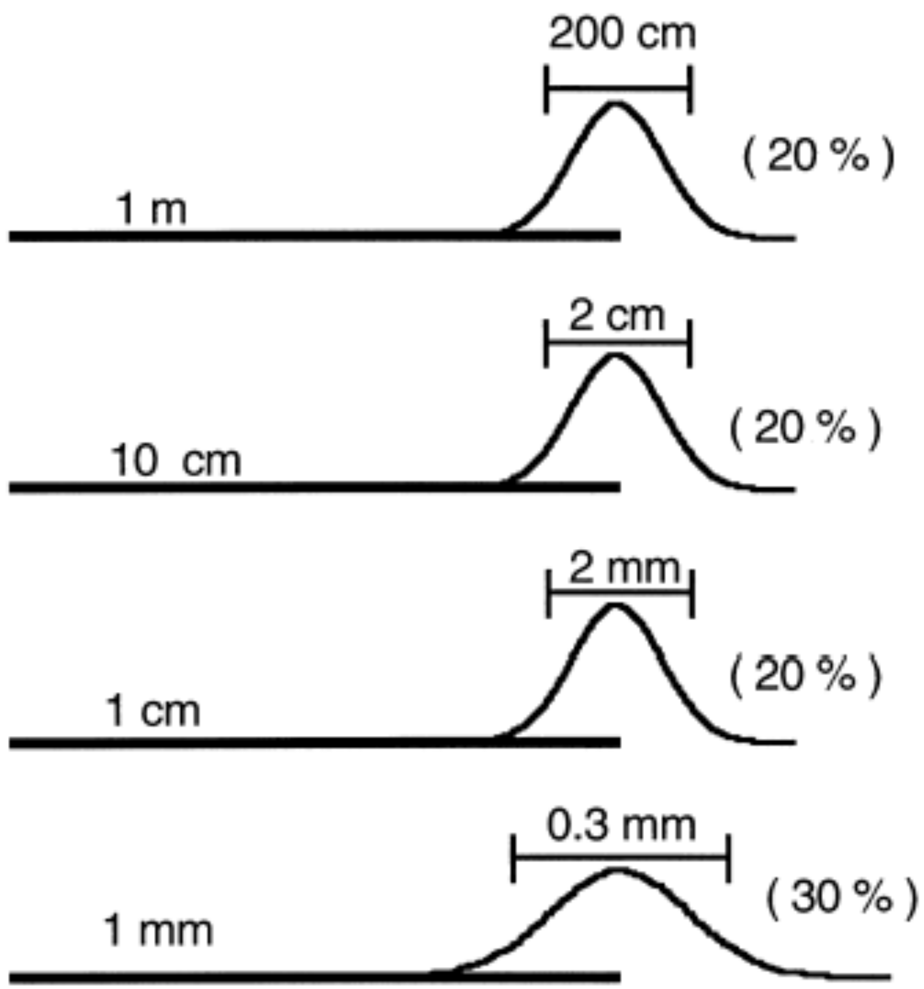
20

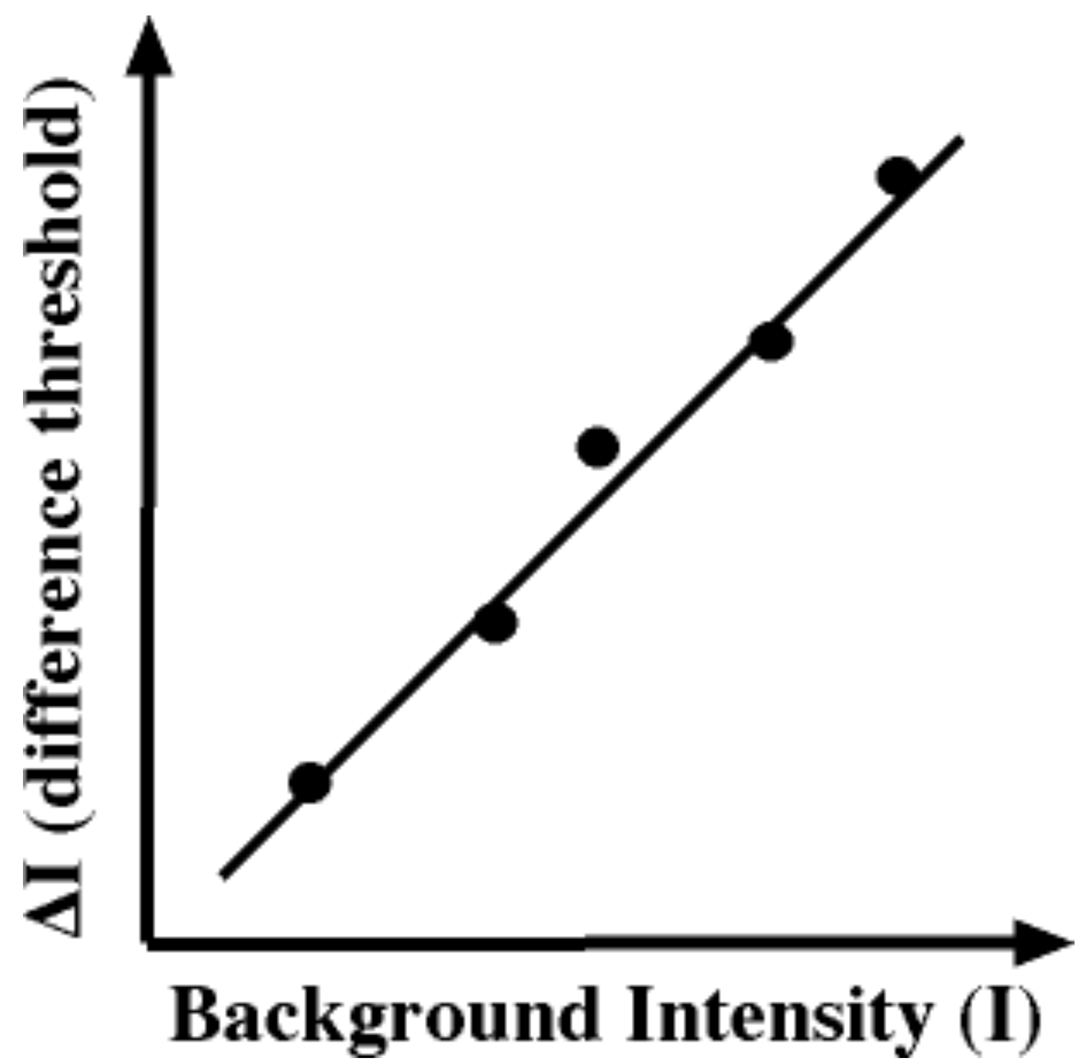
Weber's Law

- psychological law quantifying the perception of change in a given stimulus
- JND = just noticeable difference = ΔI
- Weber's law: $\Delta I \propto I$
- The change in a stimulus that will be just noticeable is a constant ratio of the original stimulus
- It doesn't hold for extremes!

Weber's Law

$\Delta I \propto I$



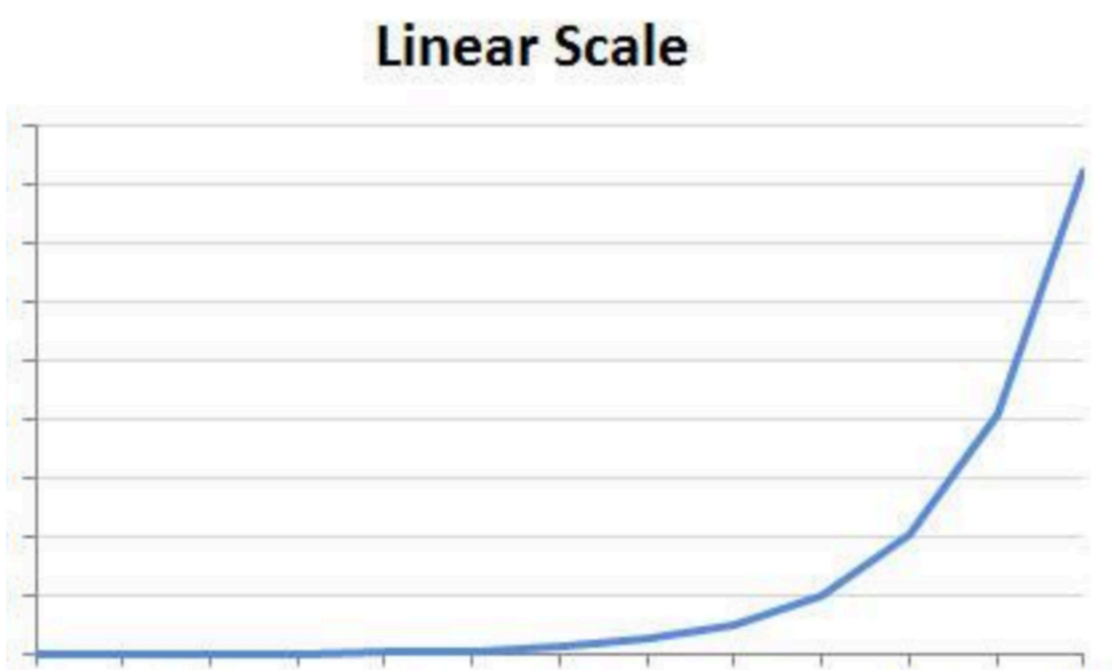


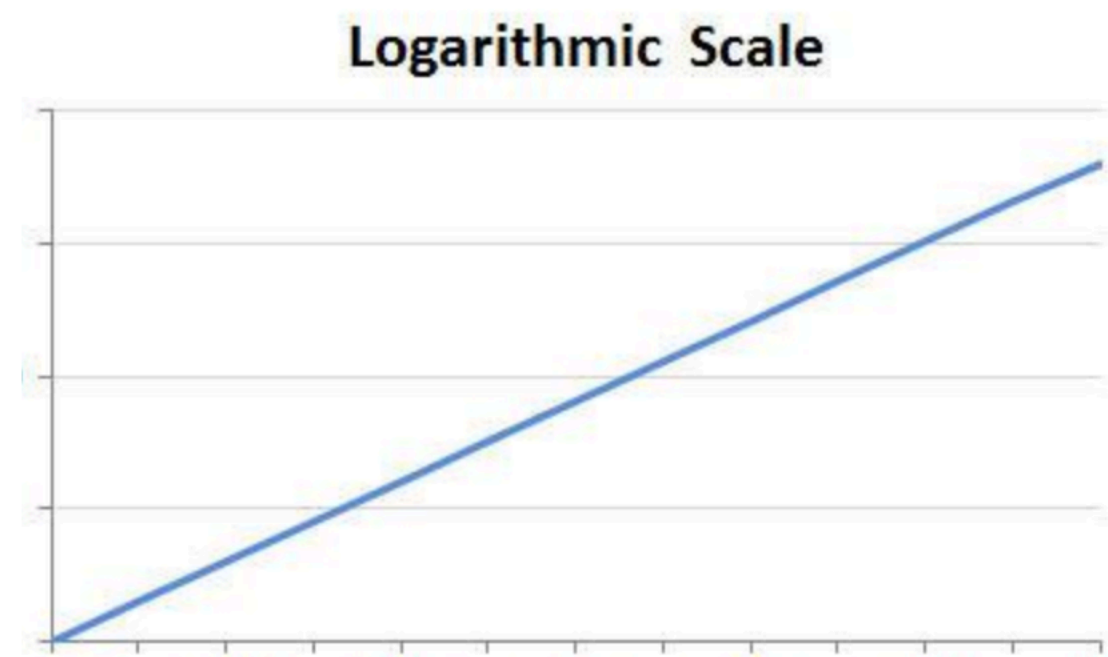
Chater & Brown (1999)

Fechner's Law

- sensitivity to stimuli depends on: individuals & senses
- subjective sensation is proportional to the **logarithm** of the stimulus intensity

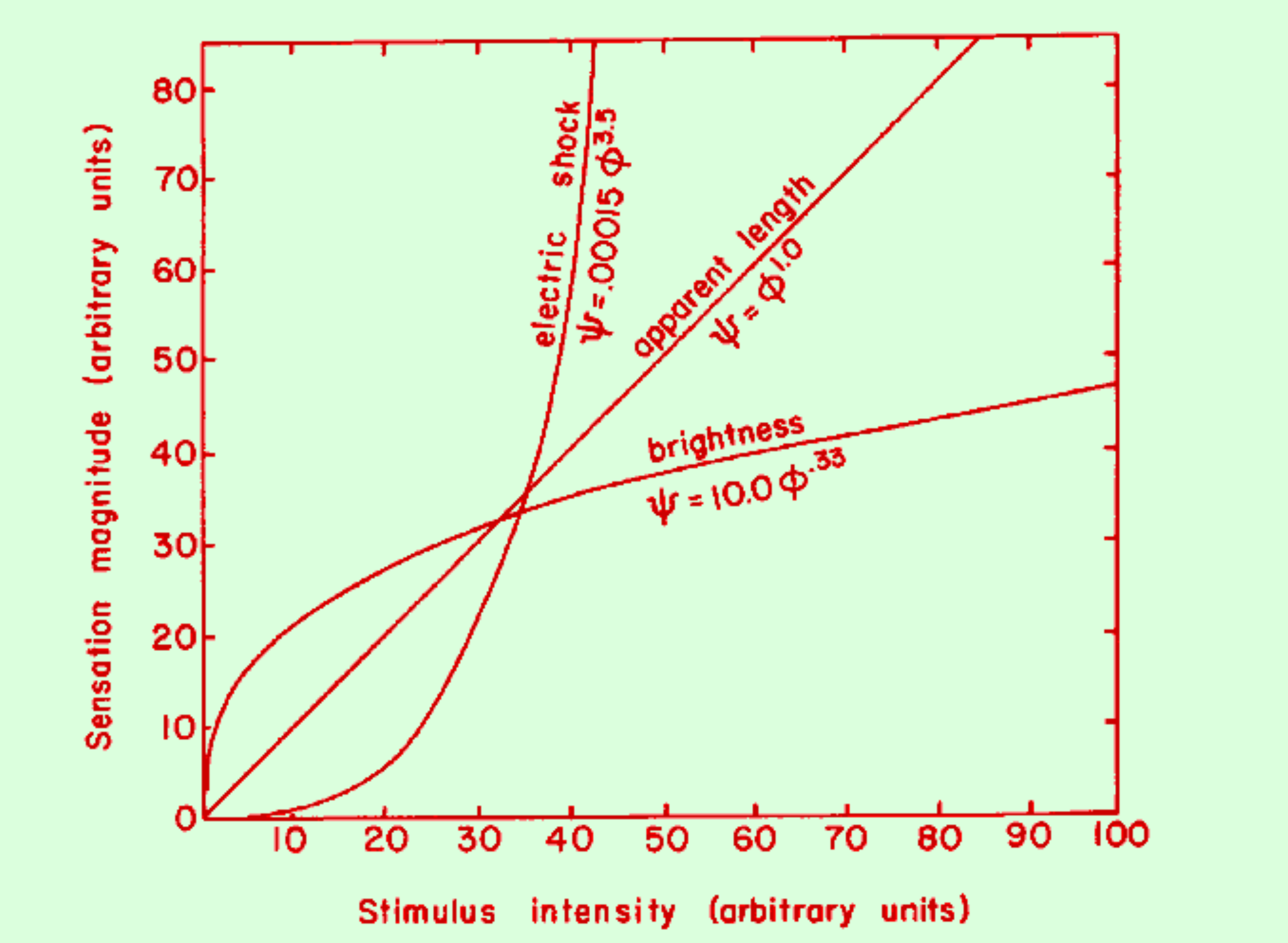
$$p = k \ln \frac{S}{S_0}$$



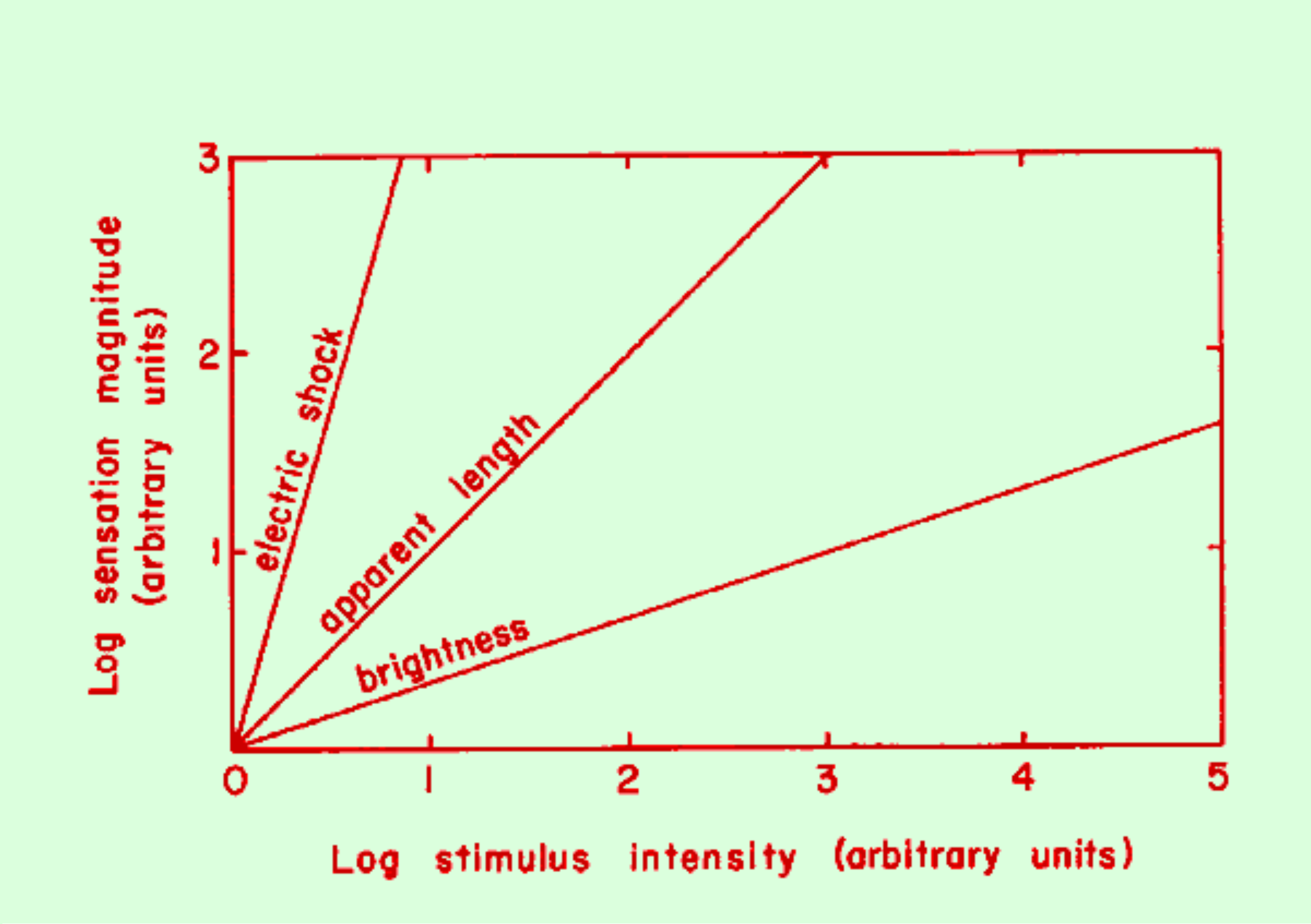


- e.g. vision, sound, weight perception

Stevens' Law

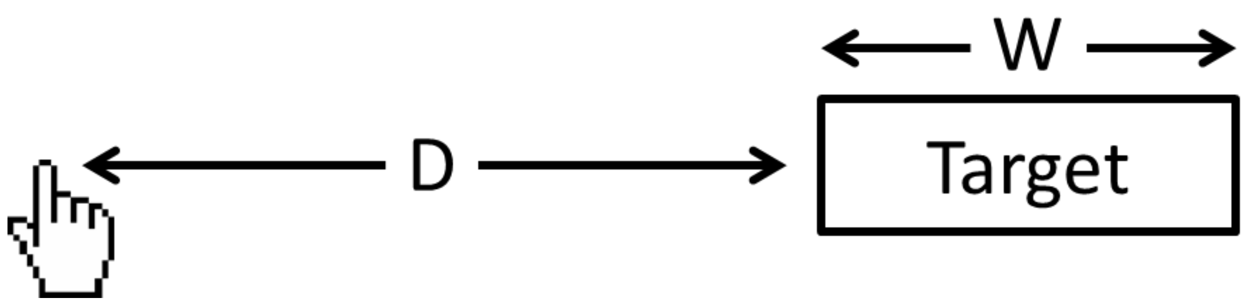


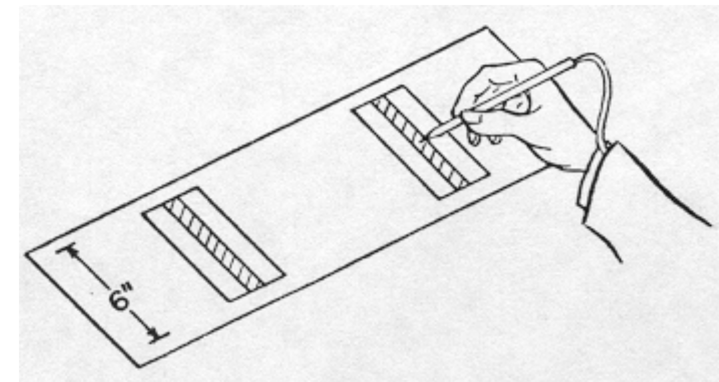
electric shock
$\psi = .00015\,\phi^{3.5}$
apparent length
$\psi = \phi^{1.0}$
brightness
$\psi = 10.0\,\phi^{.33}$
Sensation magnitude (arbitrary units)
Stimulus intensity (arbitrary units)
0
10
20
30
40
50
60
70
80
90
100



electric shock
apparent length
brightness
Log sensation magnitude (arbitrary units)
Log stimulus intensity (arbitrary units)
0
1
2
3
4
5

Fitts' Law





$$\mathrm{ID} = \log_2\left(\frac{2D}{W}\right)$$

$$\mathrm{MT} = a + b \cdot \mathrm{ID} = a + b \cdot \log_2\left(\frac{2D}{W}\right)$$

Learning and forgetting

<u>Power law of forgetting</u>

$m(t) \propto t^{-a}$

t for time

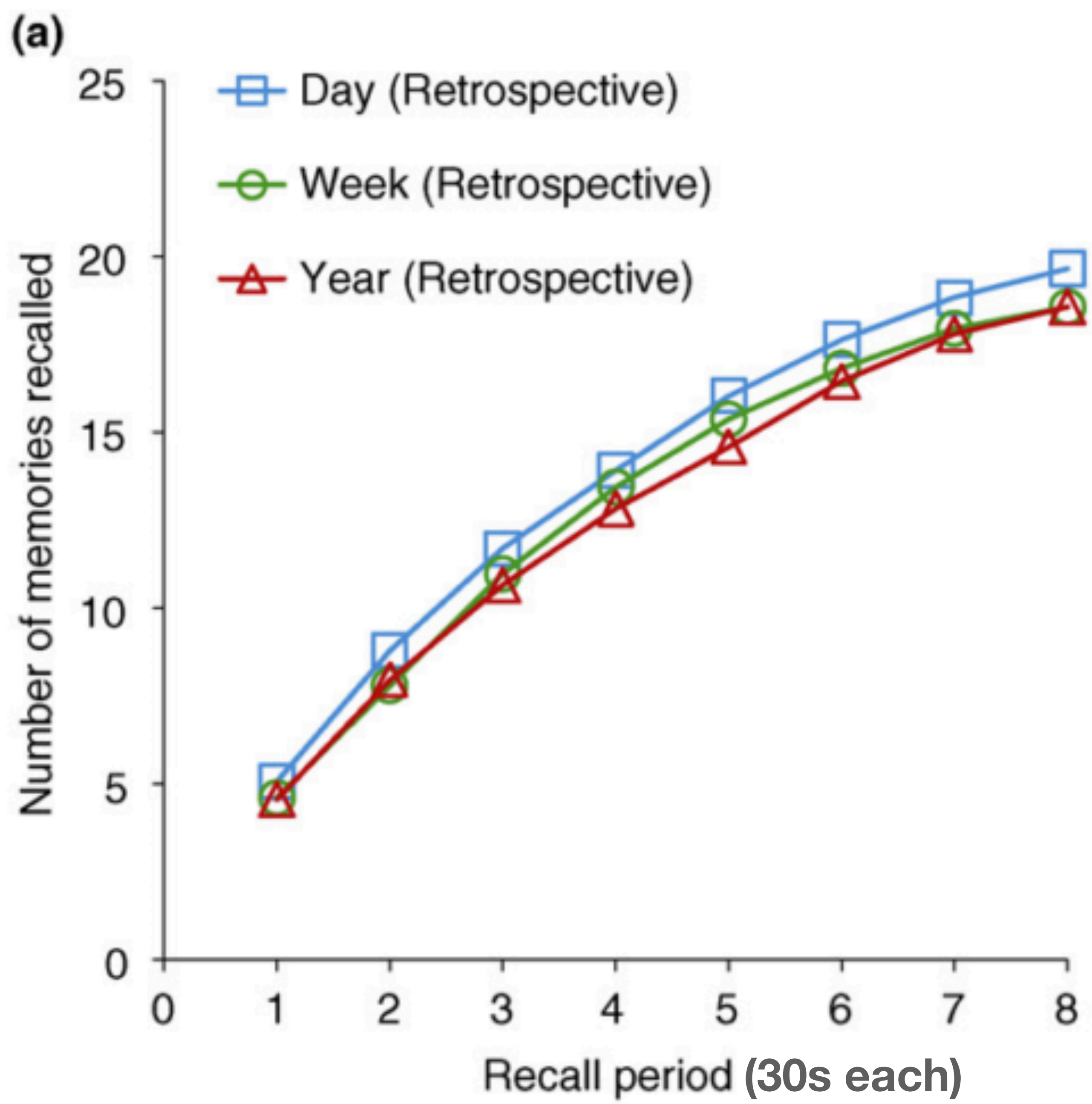
<u>Power law of practice</u>

$RT(N) \propto t^{-a}$

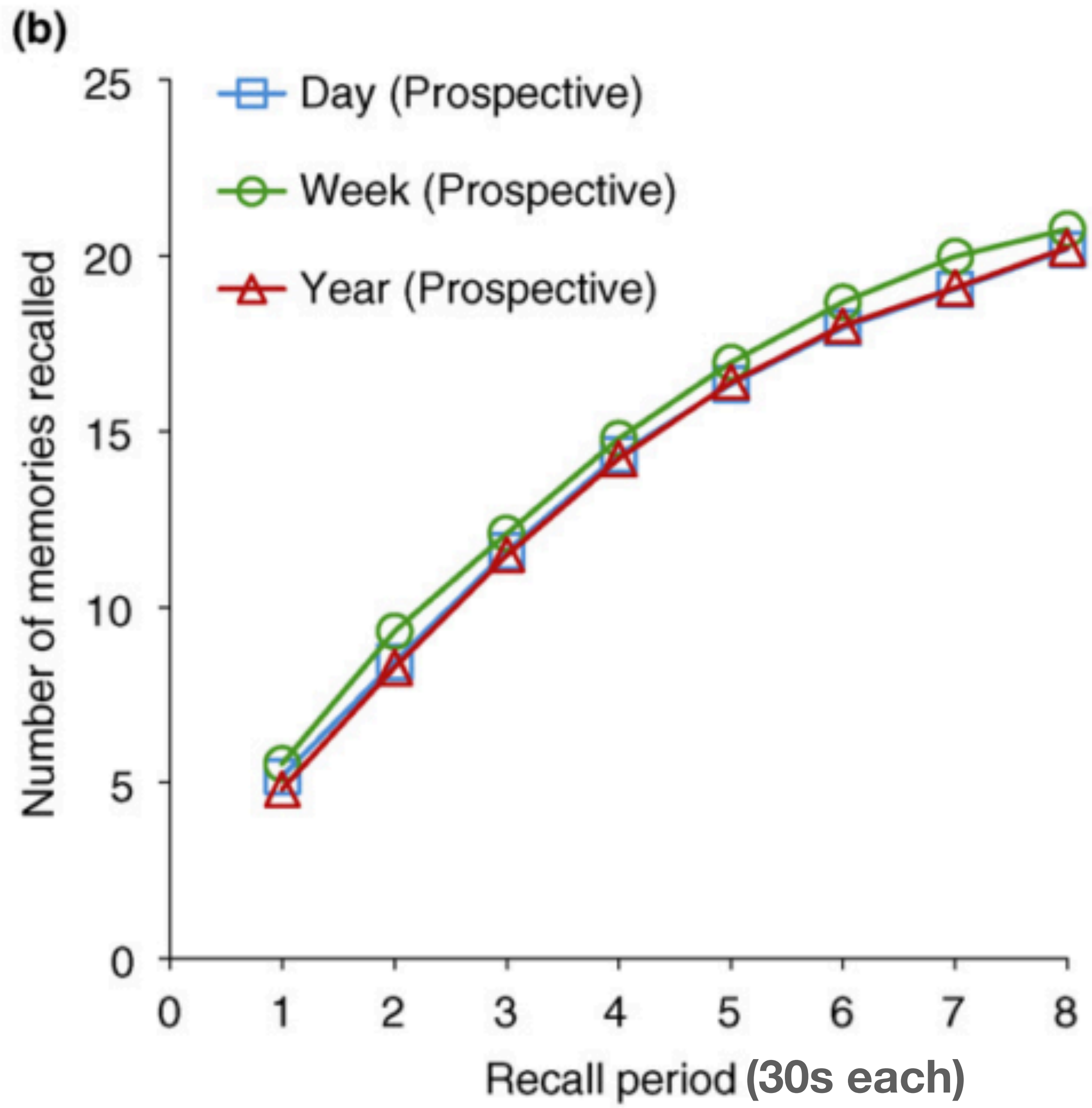
t for trials (and close to time, typically)

Memory retrieval over different time periods in retrospective memory

(Maylor, Chater & Brown, 2001, PB&R)



And prospective memory



Maylor, Chater & Brown (2001)

Luce's choice rule

- Ratio of probabilities of choosing R_i over R_j is invariant
- i.e., does not depend on the other R_k, R_m …
- w = utility; j = pool of items

$$P(i) = \frac{w_i}{\sum_j w_j}$$

Luce's choice *axiom*
(nb. R_i, R_j may be entirely qualitatively different)

- If each option, i, is characterized by a scalar, S_i
- Attractiveness of an option must be a power function

$$a_i \propto S_i^{\alpha}$$

- Choice probability can only depend on ratios of attractiveness:

$$\Pr(R_i) = \frac{S_i^{\alpha}}{\sum_j S_j^{\alpha}}$$

Luce's choice *rule*
(nb. R_i, R_j differ by a scalar S_i, S_j)

Herrnstein's matching law

- Probability of choosing an option, R_i based on $Payoff(R_i)$
- Attractiveness of an option must be a power function $(Payoff(R_i))^\alpha$
- Choice probability can only depend on ratios of attractiveness:

$$\Pr("R_i") = \frac{\left(Payoff(R_i)\right)^\alpha}{\sum_j \left(Payoff(R_j)\right)^\alpha}$$

"choice is nothing but behavior set into the context of other behavior"

A very widely observed law of behavior (and cf Luce's Choice Rule)

Conclusion

- Scale-invariance as a null hypothesis for cognitive science
- Explains a high proportion of psychological “laws” (!)
- Violations may be informative
- Scale-invariance as a building block for building “default” cognitive models
    - And might be especially interesting, when combined with other building general principles

- Discussion: Scale invariance as:
    - Deep law?
    - Null hypothesis?
    - Inaccurate approximation?

2. No absolute coding of magnitude

Decision by sampling
Stewart, Chater, Brown (2006)

Cognitive system **doesn’t** have:

- underlying “psychoeconomic” scales for:
    - utility
    - subjective probability
    - time
- there are no relationships (e.g. trade-offs) between scales

Decision by sampling

Cognitive system **does** have:

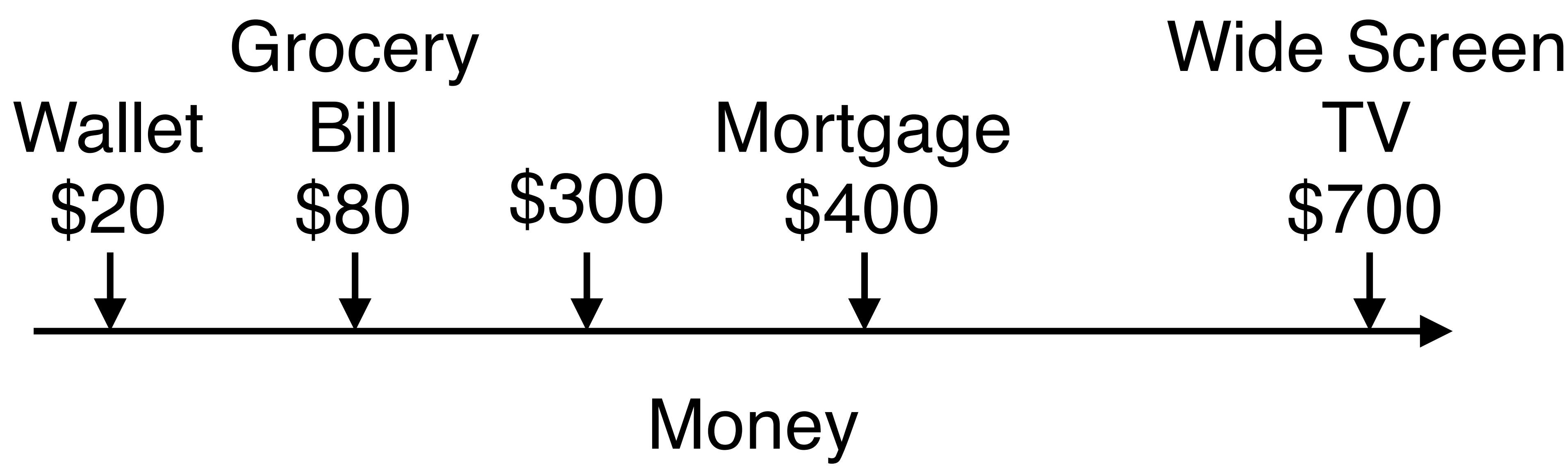
- Only binary judgments
    - "less than", "equal to", and "greater than"
- Values are compared with a small sample of “anchors”
    - from memory
    - from context
- All dimensions (gains, losses, delay, probability, quality, etc) are equal, despite different roles in “rational” model
- Preferences are constructed, depending on sampled anchor values (e.g., Slovic, 1995)
- No stable “psychoeconomic” scales

Decision by sampling

- accounts for a wide range of phenomena
  - concave utility functions
  - losses looming larger than gains;
  - hyperbolic temporal discounting;
  - the overestimation of small probabilities and the underestimation of large probabilities
- **without an underlying psychoeconomic scale!**
- **only relative judgments**

Only rank matters

What is the utility of £300? Here its 3rd of 5 items



Key issue: How do people sample comparison anchors?

From *memory*

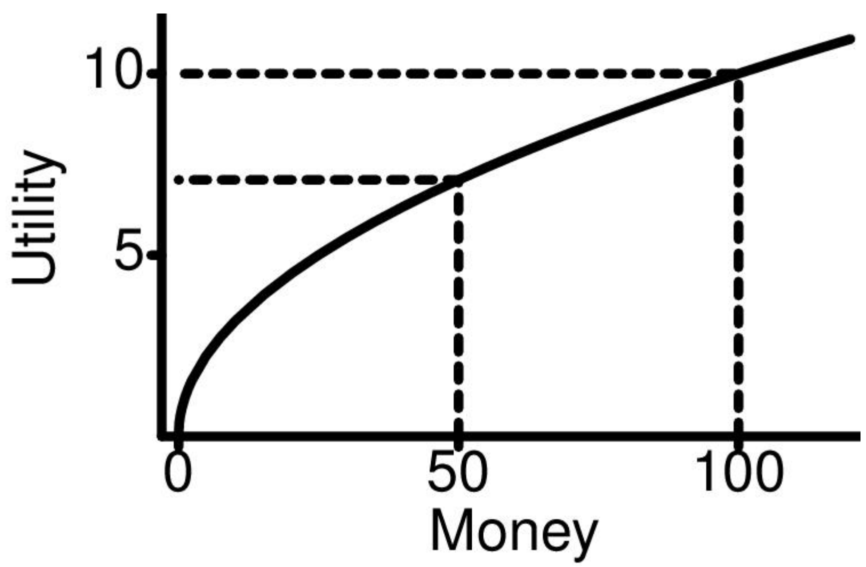
- Assume that samples from memory mirror distribution in the “world” (Anderson)
- Estimate using external “proxies” (e.g., via Google)

From task *context*

- e.g., choice can be affected by “irrelevant” options
- Explore experimentally

Diminishing “utility” of money

- The rational choice approach:



- Implies risk aversion:
- £50 for certain preferred to 50% chance of £100?

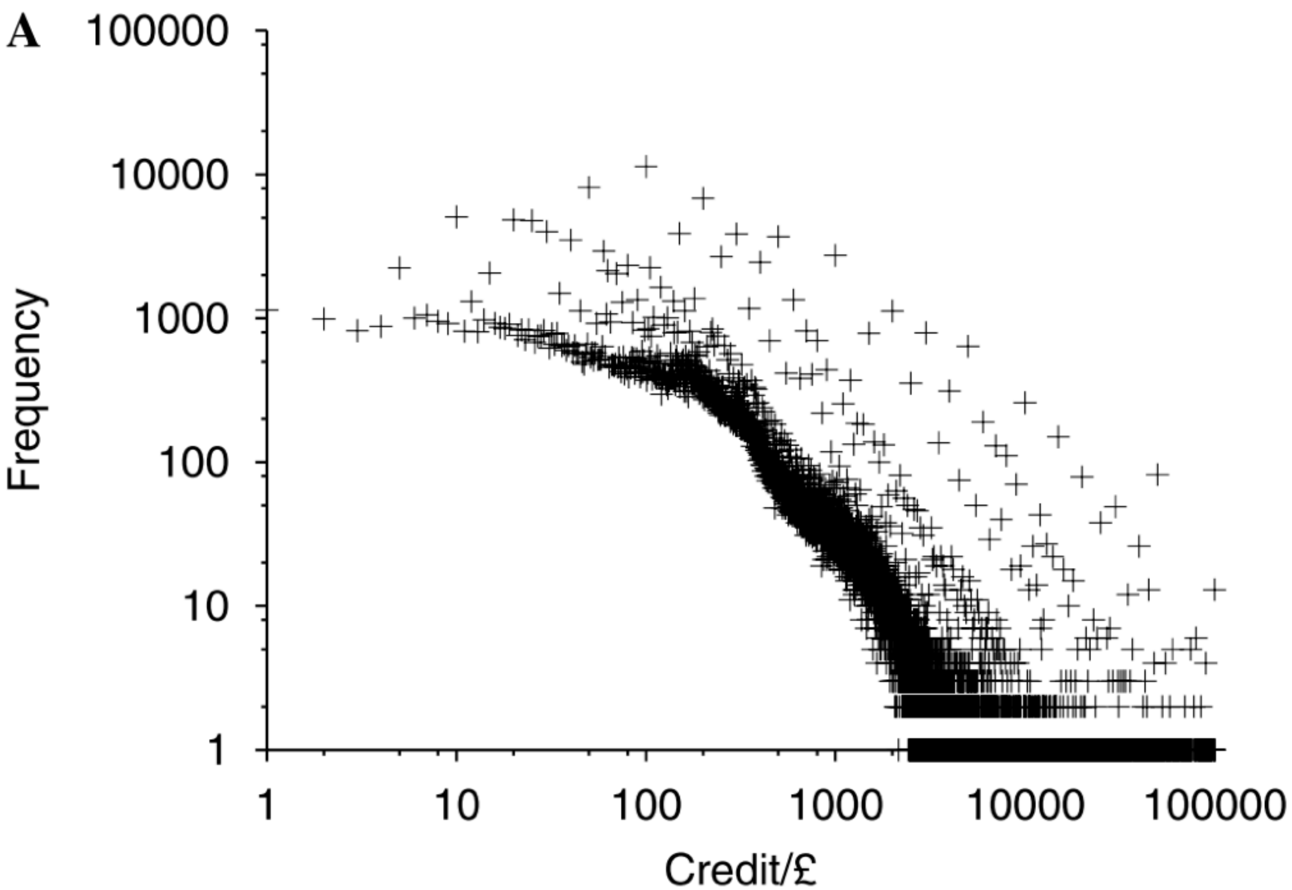
- DbS considers distribution of amounts of money
  - Only rank matters
  - So changes in money value will be valued by change in rank position in samples of amounts

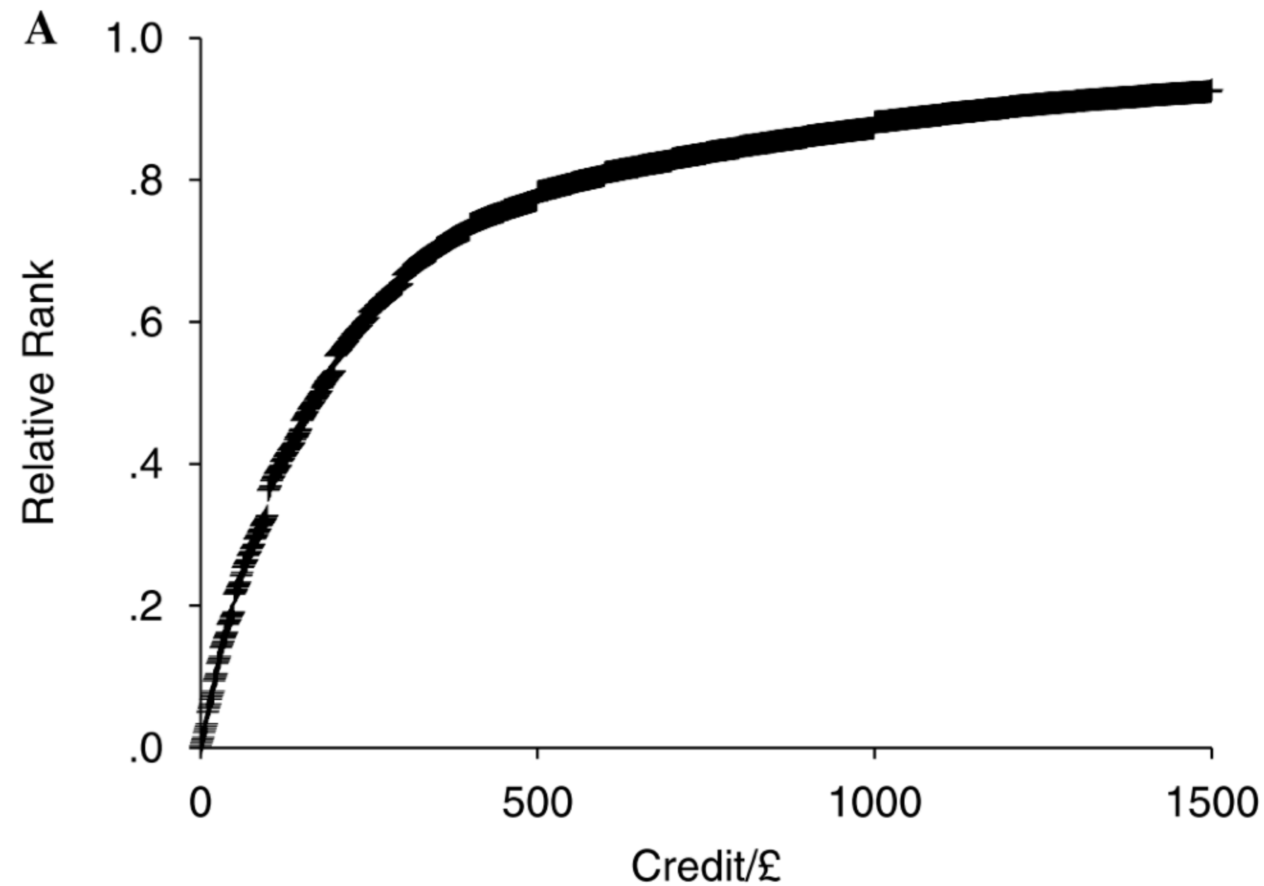
Gains & Losses

Key questions:

1) What is the distribution of gains/losses in people's memories?
2) What effect will this distribution have on the subjective valuation of gains/losses?

Gains



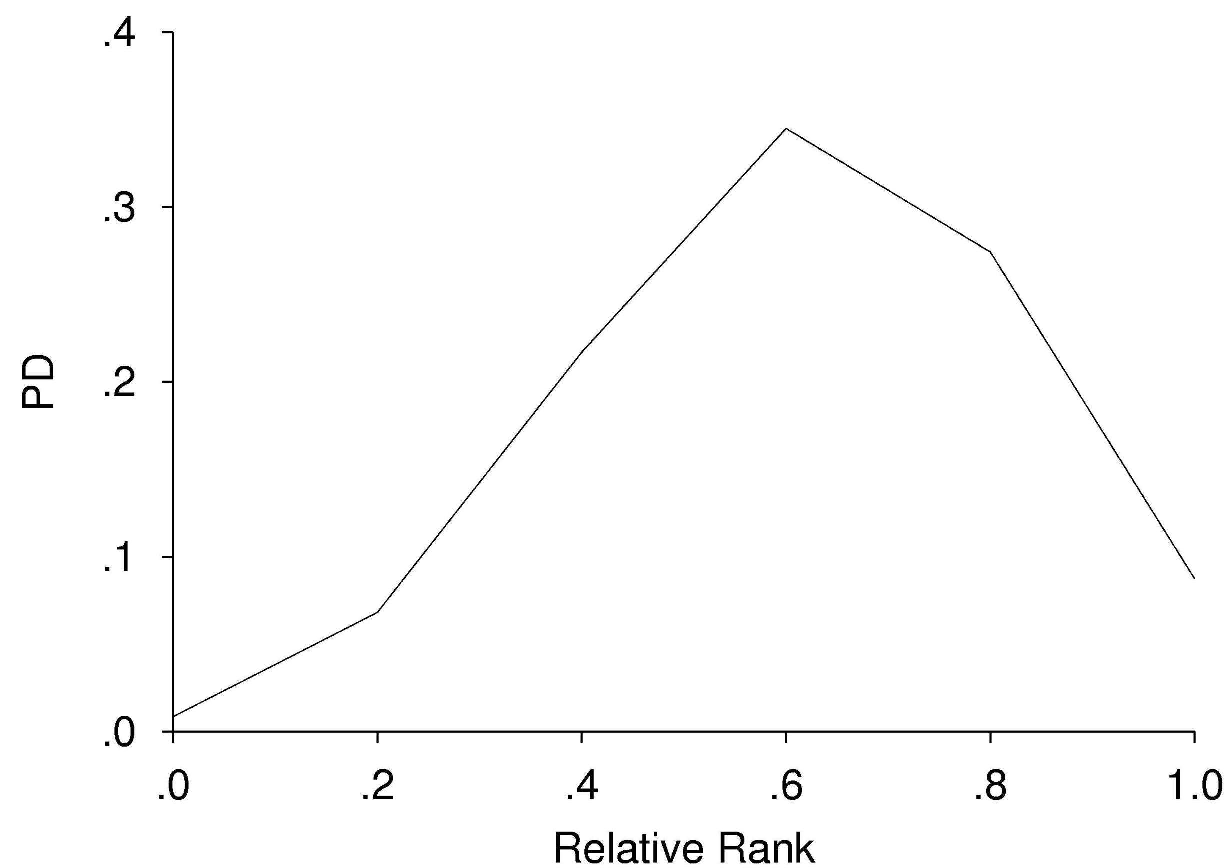


Credits to Bank Accounts in a UK High Street Bank

(random sample of one year of credits to current accounts held by a leading UK bank)

Stewart, Chater, Brown (2006)

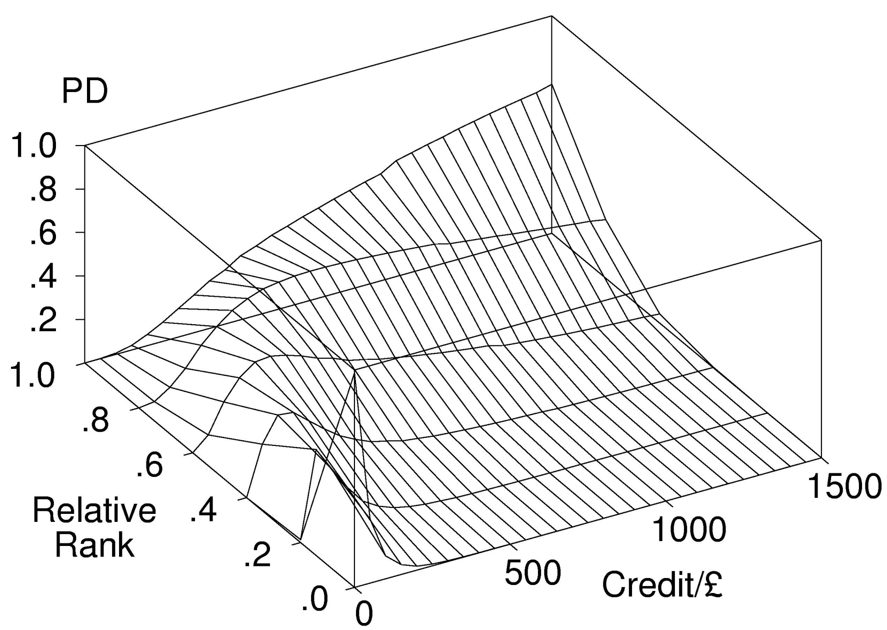
Assume that a credit of, say, £250,
is compared against 5 random credits

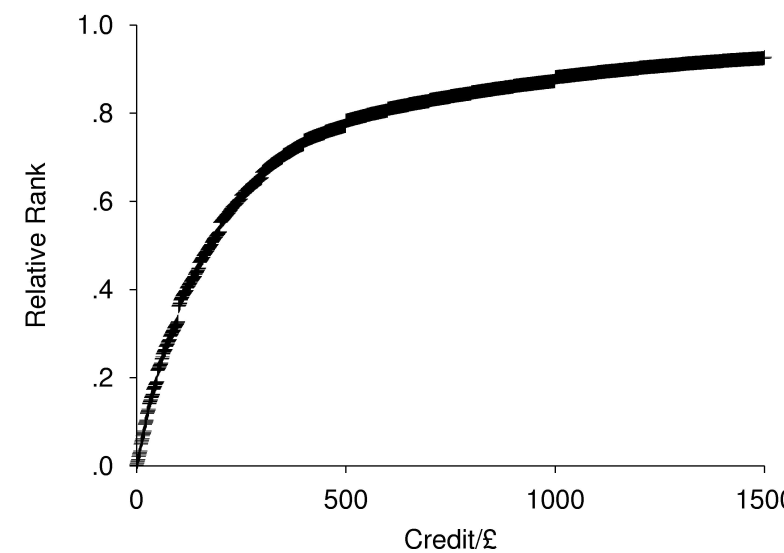


Probability distribution of relative ranks
relative rank of 0 = worst; relative rank of 1 = best.

Stewart, Chater, Brown (2006)

Combine probability distributions, for different credits

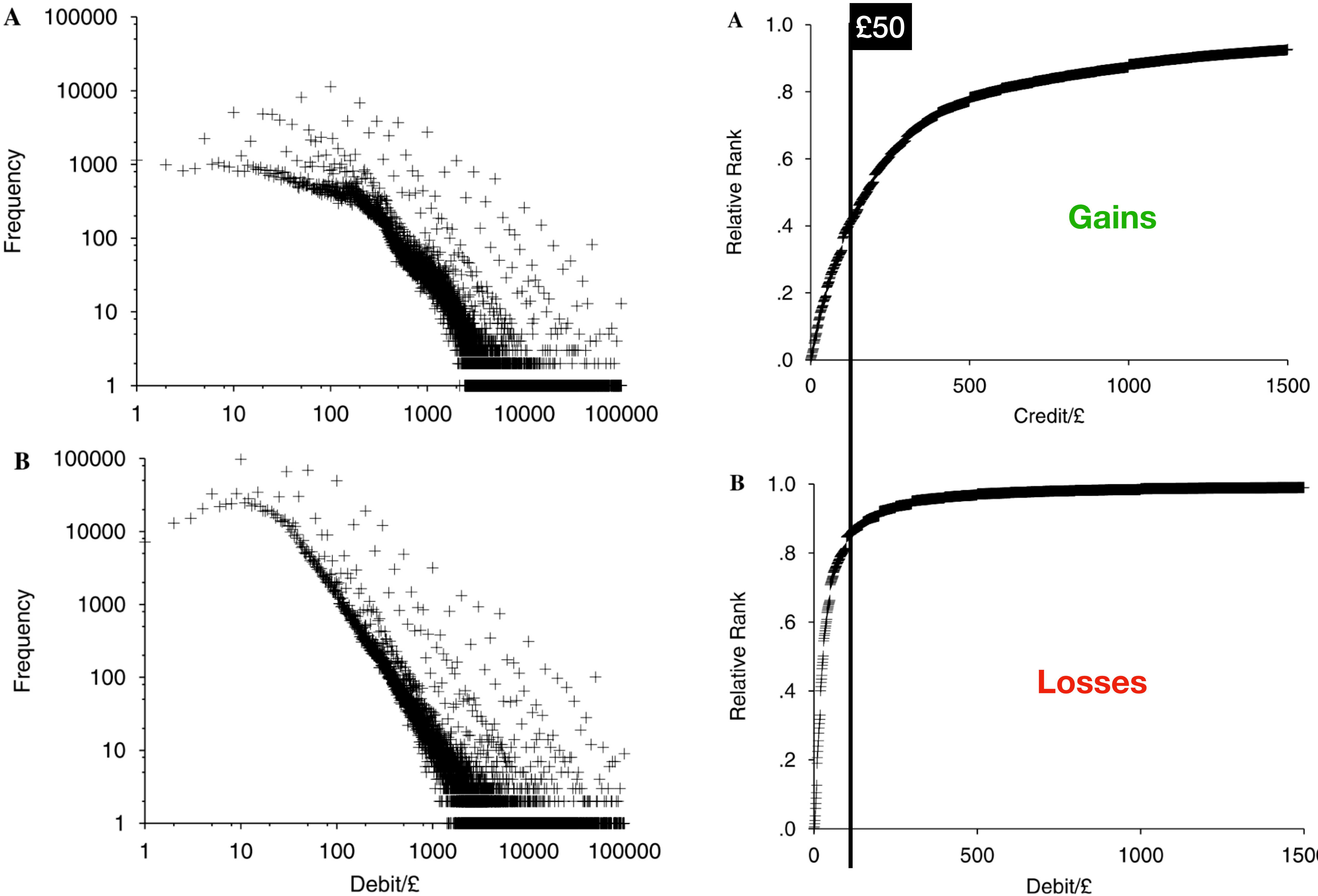




The ‘bend’ in relative rank, for small samples (left) is analogous to the curvature in the full credit-rank sample (right)

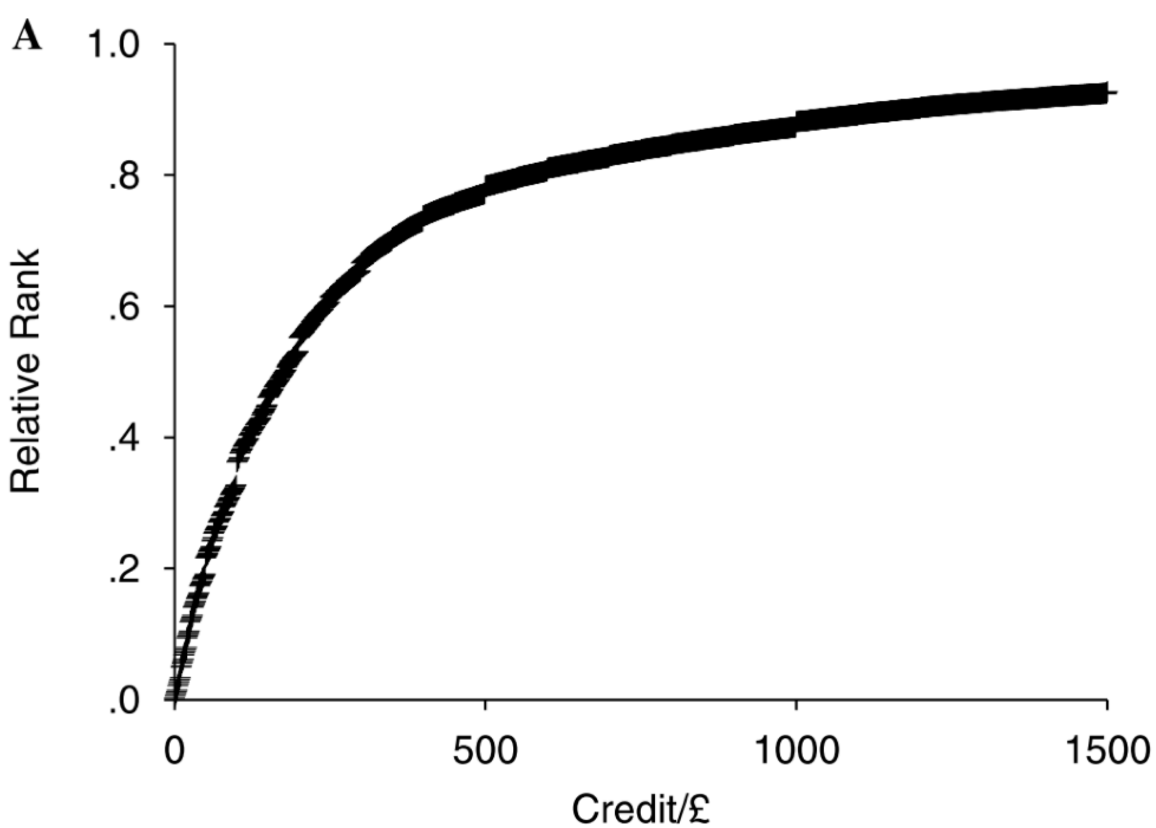
Stewart, Chater, Brown (2006)

Gains vs. Losses

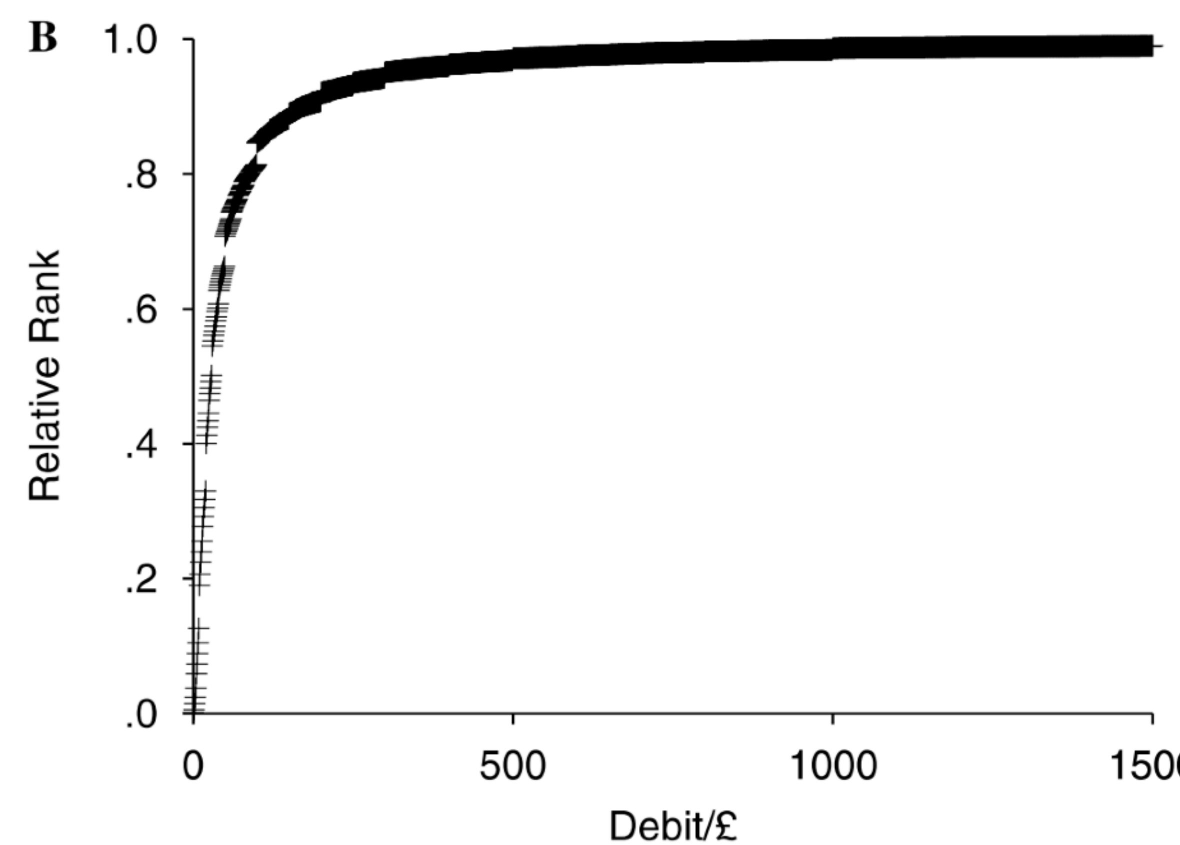


Losses loom larger than gains (credits and debits, bank data)

Gains



Losses

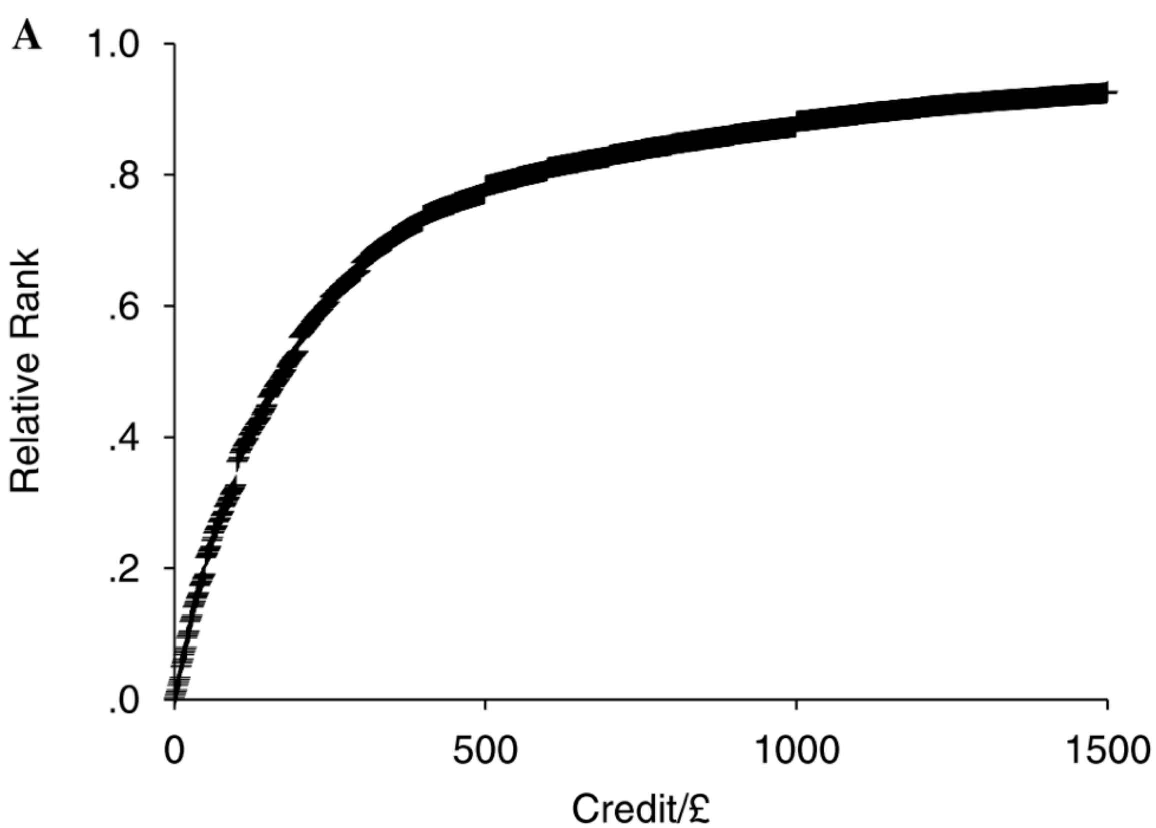


More small losses than small gains:
£50 is a “bigger” subjective loss than gain

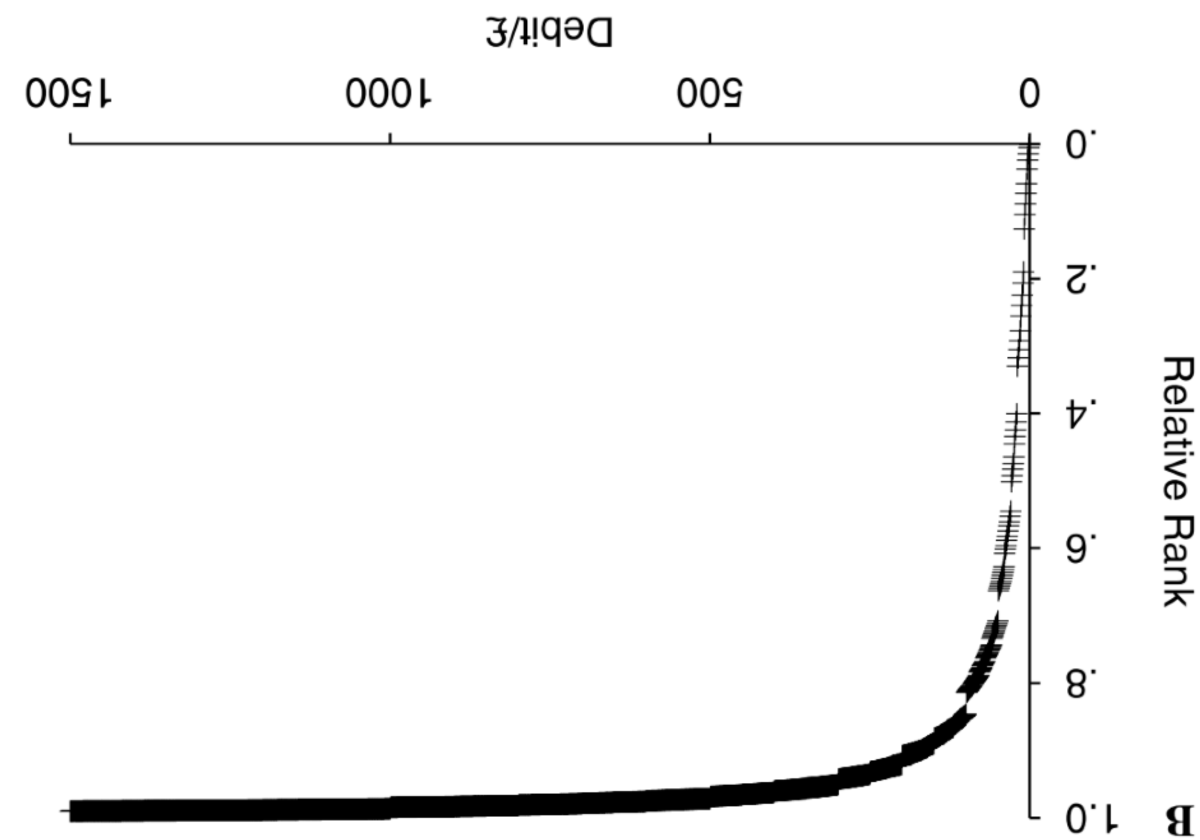
Stewart, Chater, Brown (2006)

Losses loom larger than gains (credits and debits, bank data)

Gains



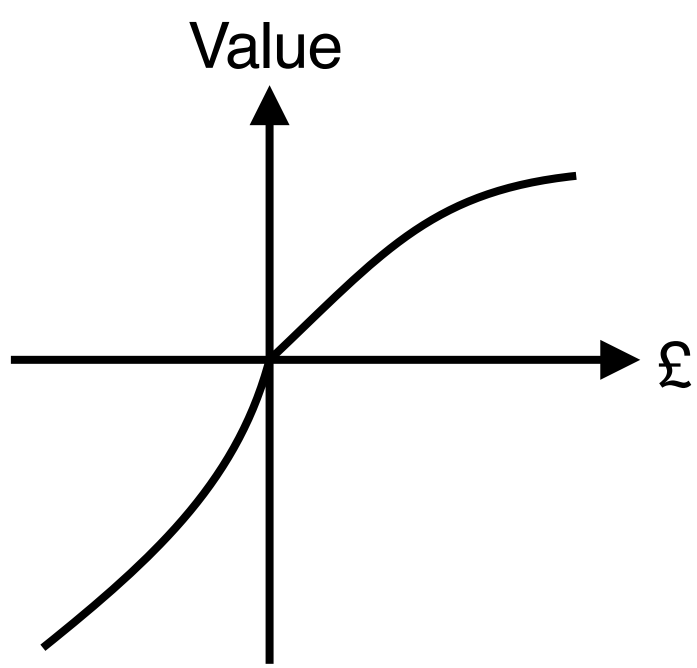
Losses



More small losses than small gains:
£50 is a “bigger” subjective loss than gain

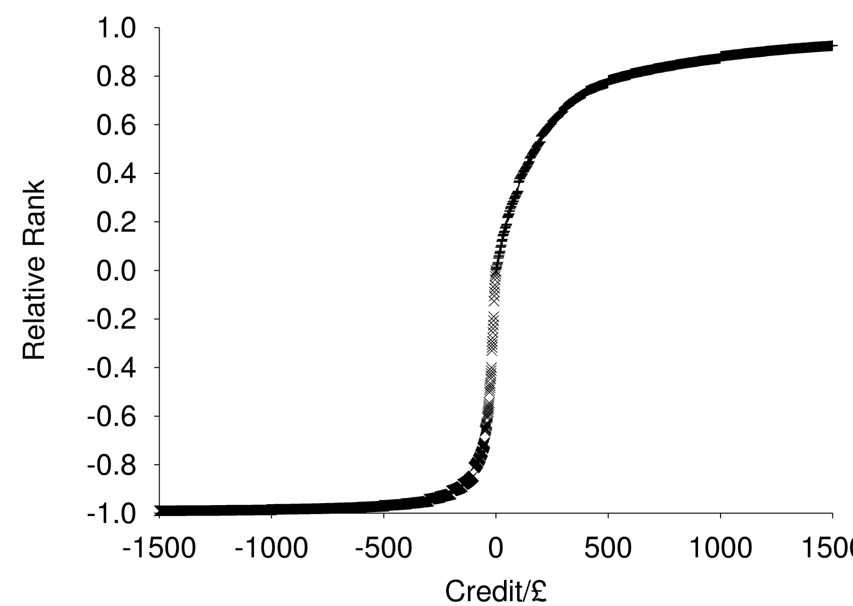
Combining gains and losses creates an analog "value function" as in prospect theory (Kahneman & Tversky, 1979)

Value function



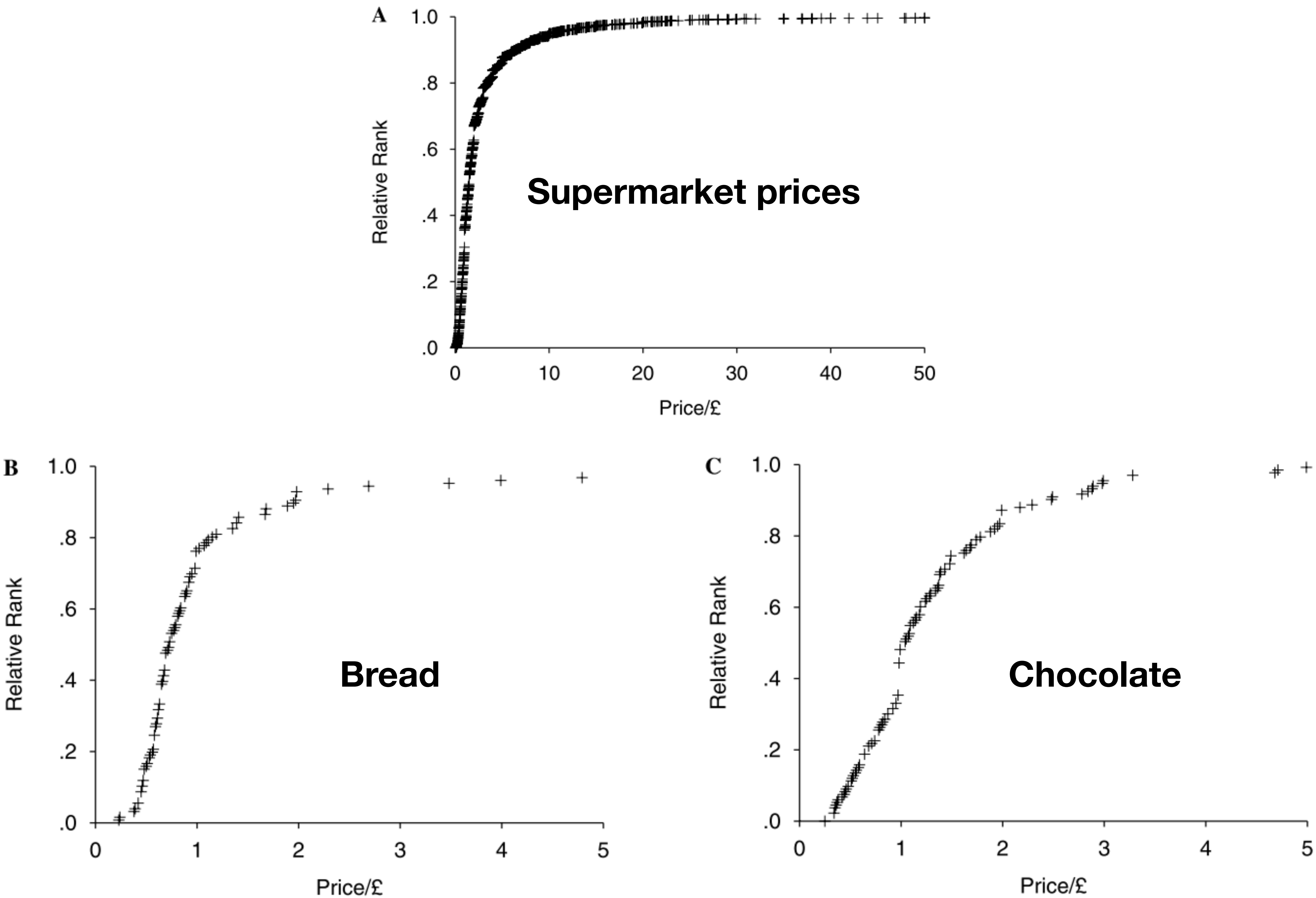
-ively accelerating for gains
+ively accelerating for losses

Relative rank



losses worse than gains
Discontinuity at zero

Not specific to bank accounts…



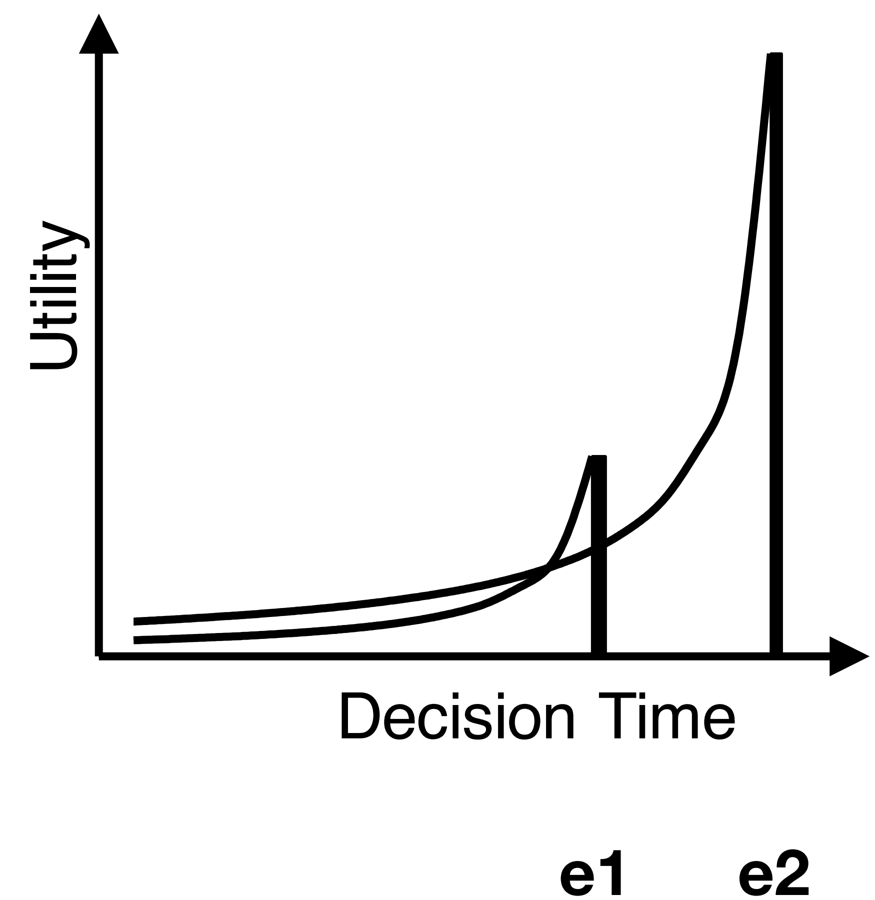
Stewart, Chater, Brown (2006)

Temporal discounting

Animals (Chung & Herrnstein, 1967; Herrnstein & Prelec, 1991)

People (Ainslie & Herrnstein, 1983; Herrnstein & Prelec, 1991)

- As events e1, e2 approach, they get more attractive. Crucially, preferences between events can reverse
- Reversals are commonly observed, but inconsistent with rational choice theory



Time

Empirical distribution of times:
Number of Google hits for 1 day, 2 days…

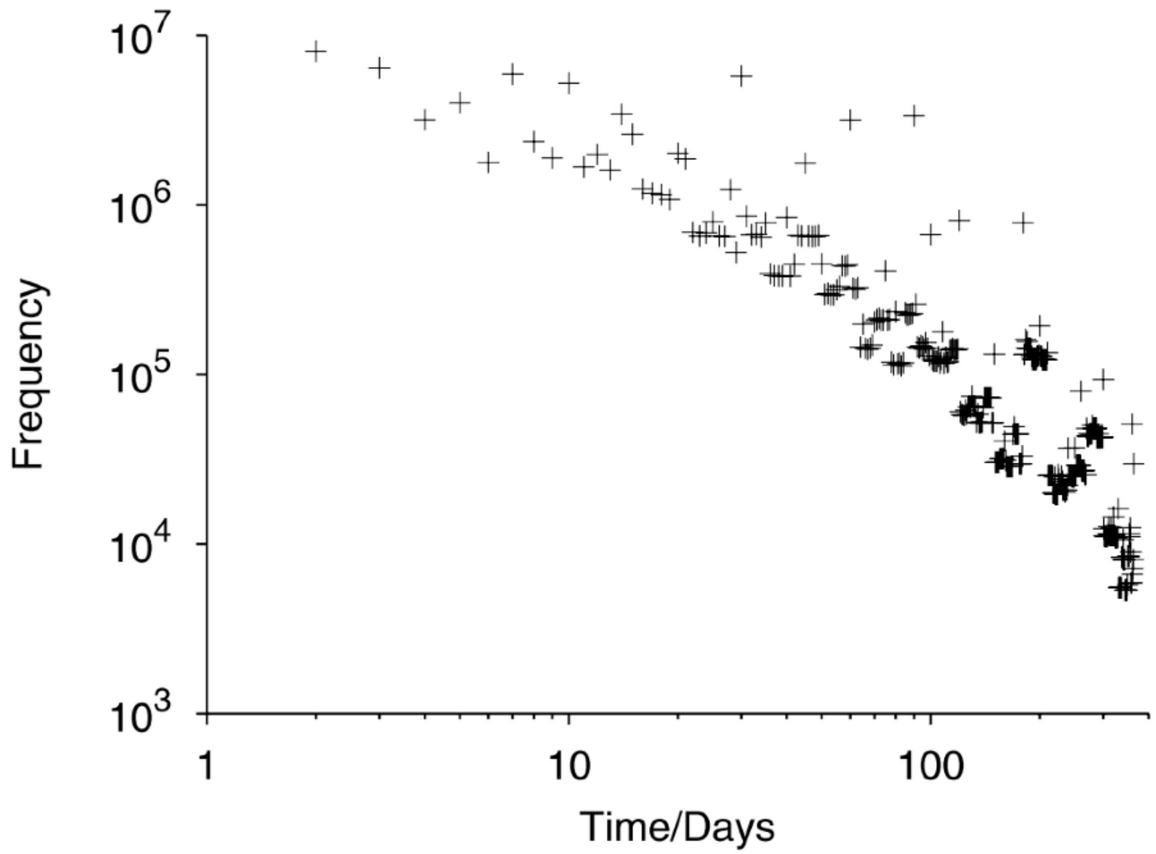
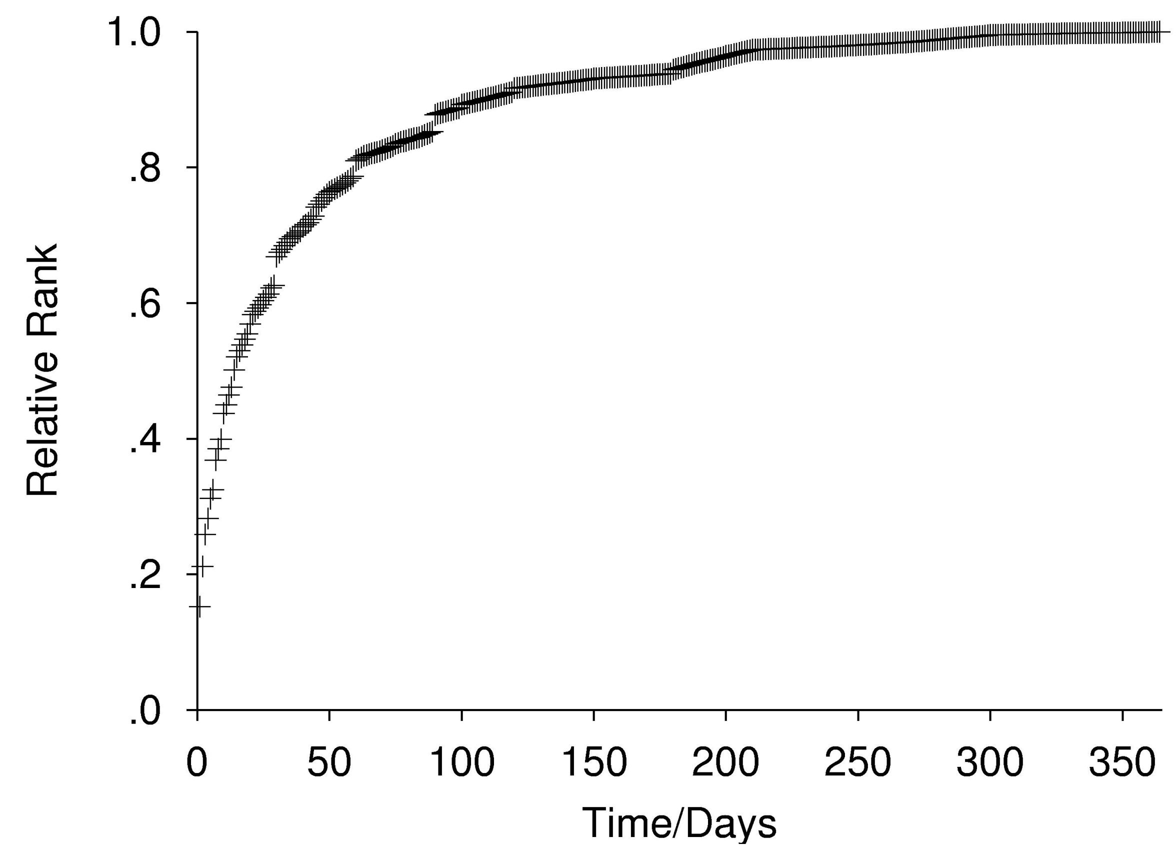


Fig. 5. The distribution of time delays on the internet.



Frequency vs #days
power = -1.5

Cumulative distribution

Stewart, Chater, Brown (2006)

Consistent with other estimates

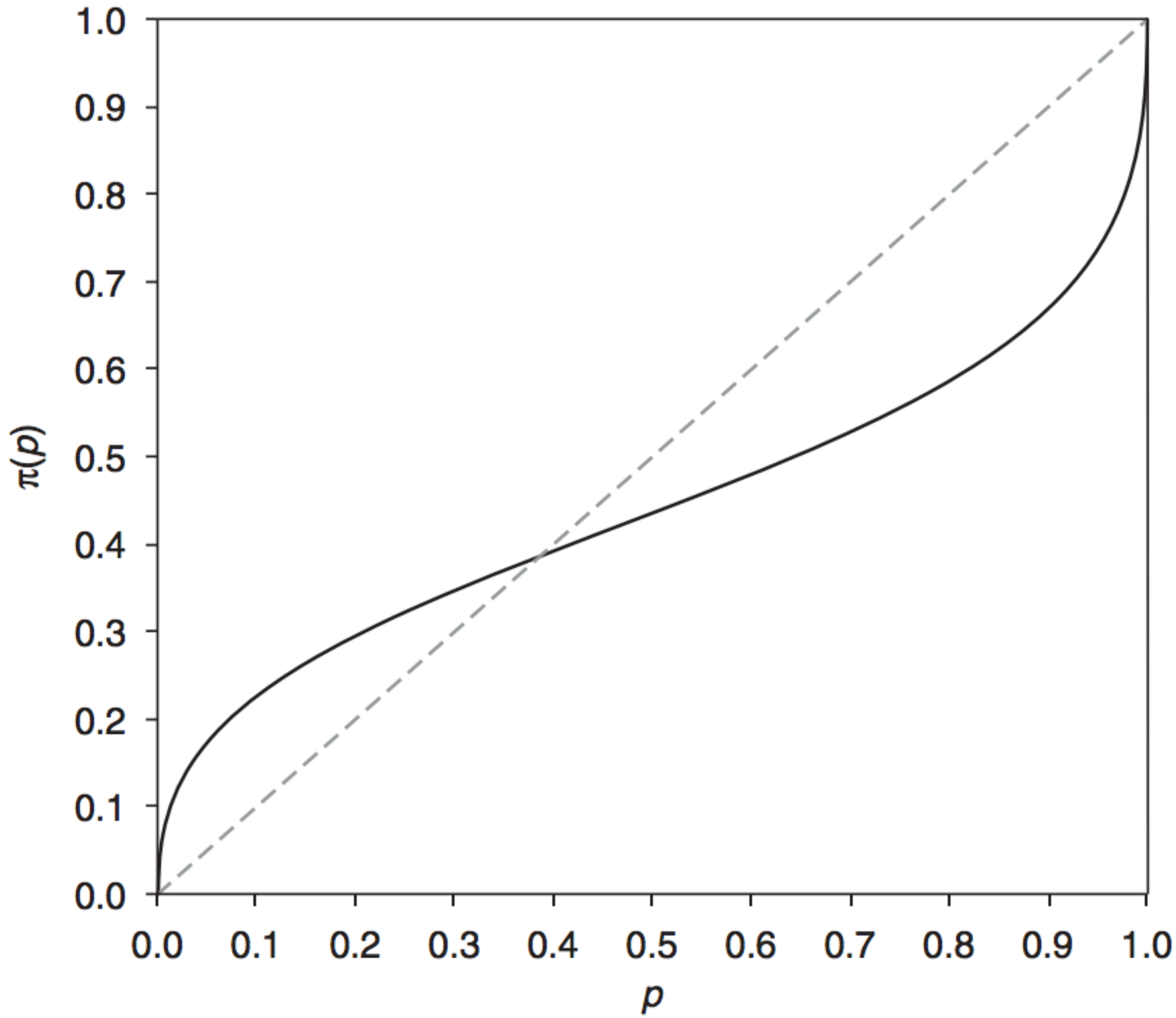
Source	Range	Power
Google™ hits	1 year	-1.5
Telegraph	30 days	-1.7
Frankfurter Allgmeine Zeitung NRC/Handelsblad International Herald Tribune (Pollmann & Baayen, 2001)	500 years	-1.4

DbS can explain the shape of the temporal discounting function

Stewart, Chater, Brown (2006)

Risk evaluation

Subjective probability S curve



Distribution of probability phrases

Probability phrases:
- numerical probability equivalent (experimental)
- frequency in natural language (British National Corpus)

Table 2
Judged numerical equivalents and BNC frequencies of probability phrases

Phrase	Judged numerical equivalents				BNC frequency		
	M	*SD*	Mdn	IQR	Raw frequency	Proportion of probability uses	Adjusted frequency
Impossible	0.00	0.00	0.0	0.0	6170	1.00	6170
Not possible	0.00	0.00	0.0	0.0	1217	1.00	1217
No chance	0.00	0.00	0.0	0.0	534	.60	320
Never	0.00	0.00	0.0	0.0	48,217	.80	38574
Extremely doubtful	3.76	2.81	3.0	3.0	20	.95	19
Almost impossible	3.79	3.19	2.5	4.0	486	.90	437
Pretty impossible	5.36	5.86	3.0	7.5	2	1.00	2
Almost unfeasible	6.33	6.14	5.0	8.0	0	.00	0
Highly unlikely	7.11	5.08	5.0	5.0	172	1.00	172
Highly improbable	7.31	5.17	5.0	5.0	27	1.00	27
Very doubtful	8.08	5.73	5.0	5.0	66	.95	63
Very unlikely	8.25	4.58	9.5	5.0	157	1.00	157
Little chance	11.75	7.38	10.0	10.0	273	.80	218
Faint possibility	11.89	8.71	10.0	15.0	7	1.00	7
Pretty doubtful	13.20	8.57	10.0	12.25	1	1.00	1
Improbable	13.28	11.22	10.0	15.0	340	1.00	340

Distribution of probability phrases

Table 2 (*continued*)

Phrase	Judged numerical equivalents				BNC frequency		
	M	*SD*	Mdn	IQR	Raw frequency	Proportion of probability uses	Adjusted frequency
Usually	74.15	10.96	75.0	15.0	18619	.85	15826
Rather likely	74.25	9.88	75.0	11.3	1	1.00	1
Very feasible	74.26	10.15	75.0	10.0	3	.00	0
Most of the time	78.74	10.78	80.0	15.0	580	.95	551
High likelihood	79.73	8.50	80.0	16.3	5	1.00	5
Fairly certain	79.83	12.16	85.0	20.0	56	1.00	56
Great likelihood	80.82	9.64	80.0	12.5	1	1.00	1
High possibility	80.93	7.30	80.0	11.0	1	1.00	1
Most likely	81.05	11.86	80.0	15.0	1341	.00	0
Very likely	81.53	8.05	80.0	13.5	296	.85	252
Great possibility	82.49	8.04	80.0	10.0	1	1.00	1
Quite certain	82.85	10.27	85.0	15.0	97	.90	87
Pretty certain	85.30	9.19	89.5	10.0	45	1.00	45
Very certain	89.78	7.35	90.0	11.3	15	.87	13
Almost certain	92.32	5.76	95.0	5.0	1694	1.00	1694
Most definitely	95.13	5.32	95.0	7.8	109	.20	22
Sure thing	97.53	4.34	100.0	5.0	27	.35	9
Always	100.00	0.00	100.0	0.0	41,869	.90	37682
Absolute certainty	100.00	0.00	100.0	0.0	37	.40	15
Certain	100.00	0.00	100.0	0.0	36,121	.25	9030
Definitely	100.00	0.00	100.0	0.0	3233	.80	2586

Stewart, Chater, Brown (2006)

Risk evaluation

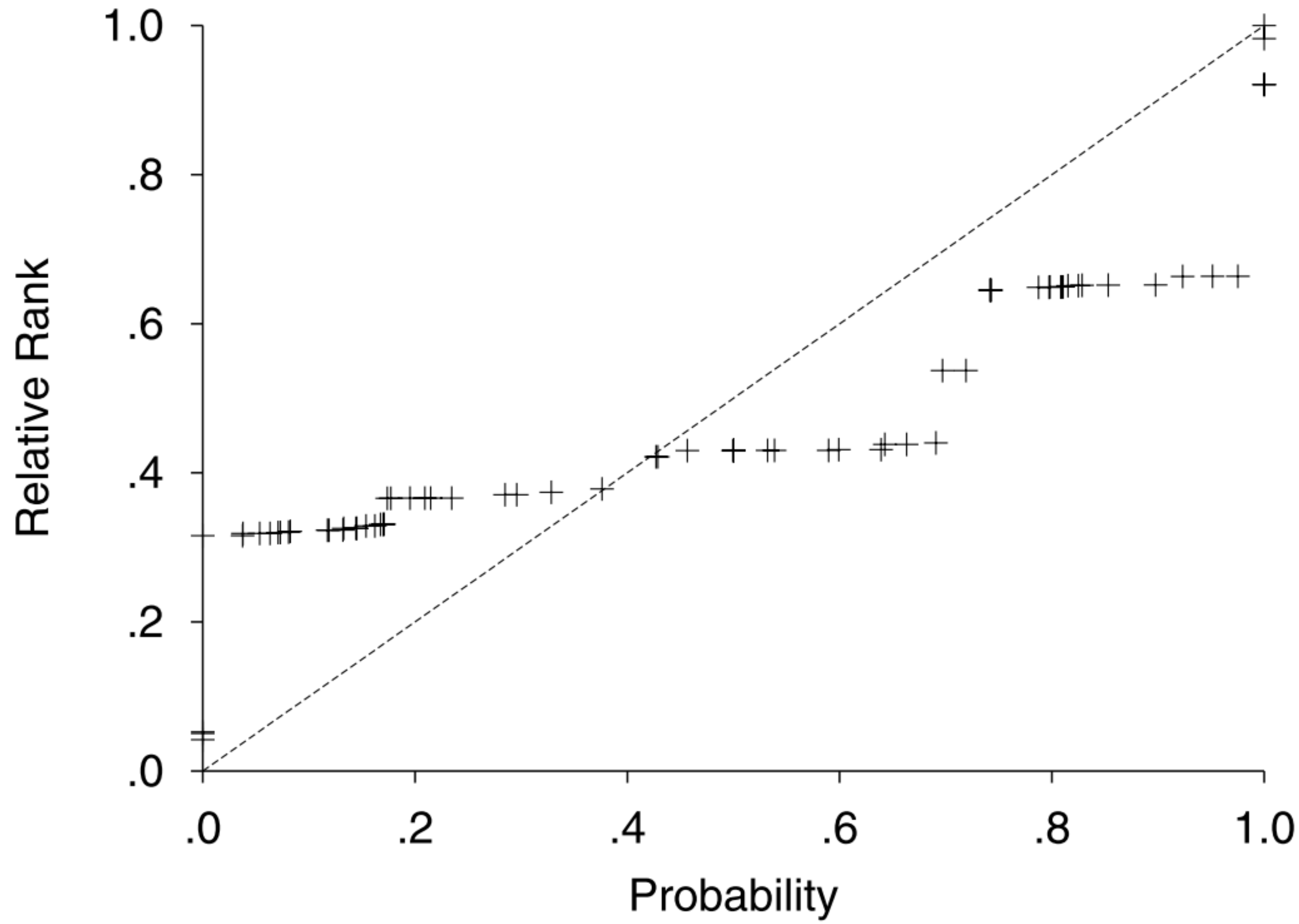


Fig. 7. The relative rank of probability phrases.

Stewart, Chater, Brown (2006)

Conclusion

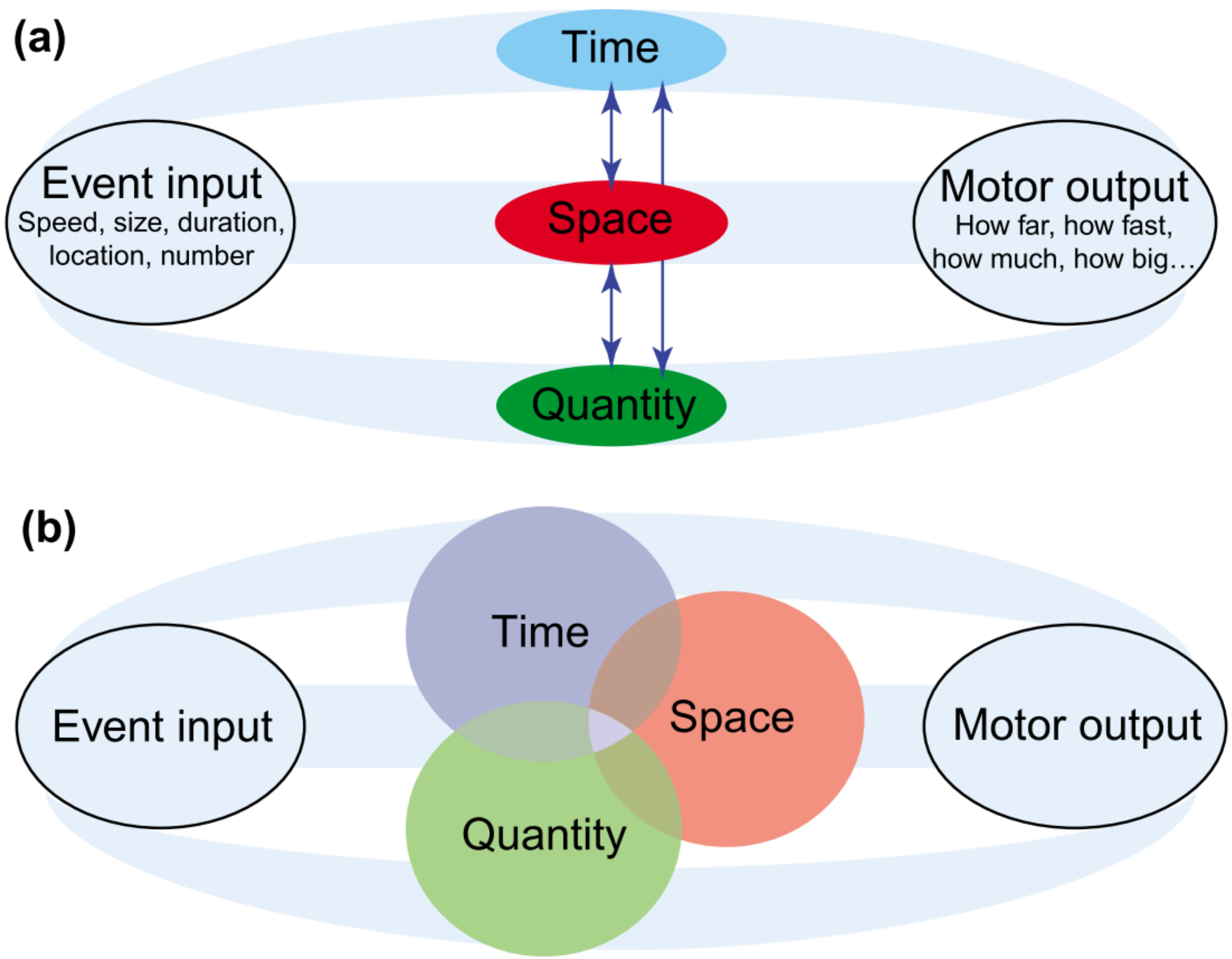
- DbS provides explanations for the descriptive functions of:
  - utility of gains and losses
  - temporal discounting
  - subjective value of probabilities
- Two key claims:
  - people make only binary, original comparisons
  - attribute values are compared with values from a decision sample drawn from memory
- Assumption: distribution of values in memory reflect distribution of attribute values in the world

3. ATOM: A Theory of Magnitude

Walsh (2003)

- brings together disparate literatures on time, space and number
- shows similarities between these three domains
- indicative of common processing mechanisms?
- common neural substrate
- parietal cortex -> reaching, grasping, space, quantity and time

3. ATOM: A Theory of Magnitude



Walsh (2003)

What is time for? What is space for? Why do we have numbers?

- Temporal and spatial information are of course necessary for action: for reaching, throwing, pointing and grasping, and the objects at which we direct these actions are frequently moving
- Time and space are rarely segregated in everyday activities (but they usually are in experiments)
- It would be surprising not to at least be in close proximity in the brain (they are - in the parietal lobe)
- Parietal cortex has an analogue system for action that computes ‘more than–less than’, ‘faster–slower’, ‘nearer–farther’, ‘bigger–smaller’ -> Numbers free load on this system (most efficient)

Distance effect in number processing

2 3

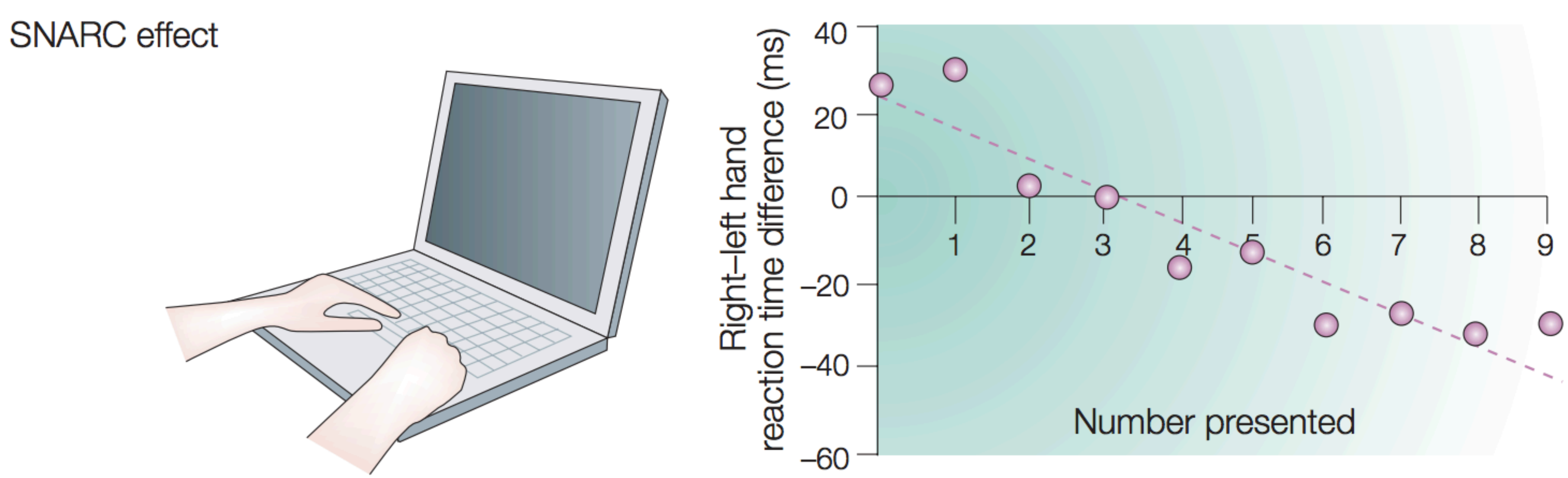
Distance effect in number processing

2 9

Distance effect in number processing

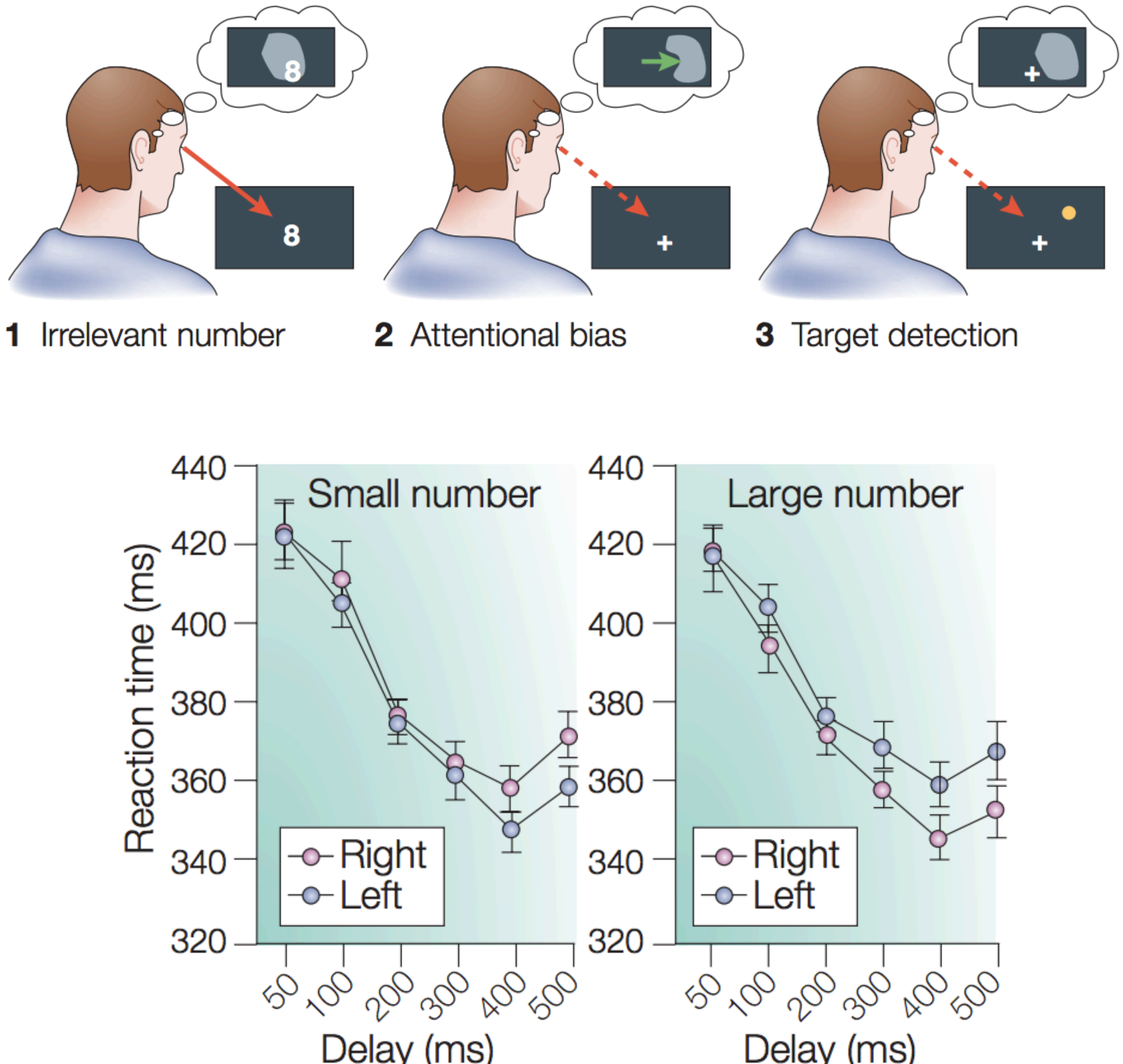
SNARC effect

- SNARC = spatial–numerical association of response codes
- Participants are asked to judge whether a number is even or odd
- Responses to larger numbers are faster on the right side of space whereas those for smaller numbers are faster on the left side



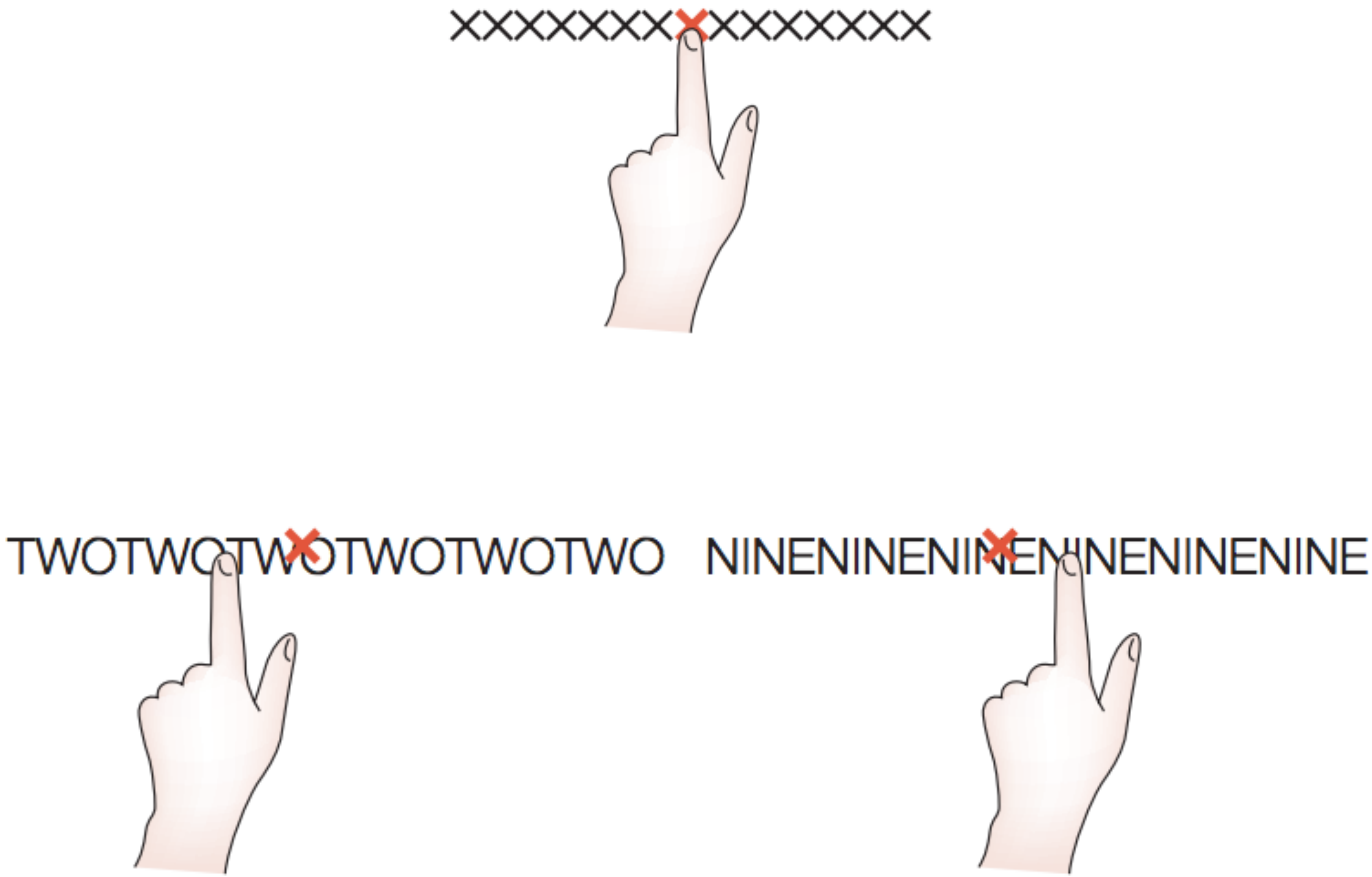
from *Hubbard et al. (2005)*

Attention bias effect



from *Hubbard et al. (2005)*

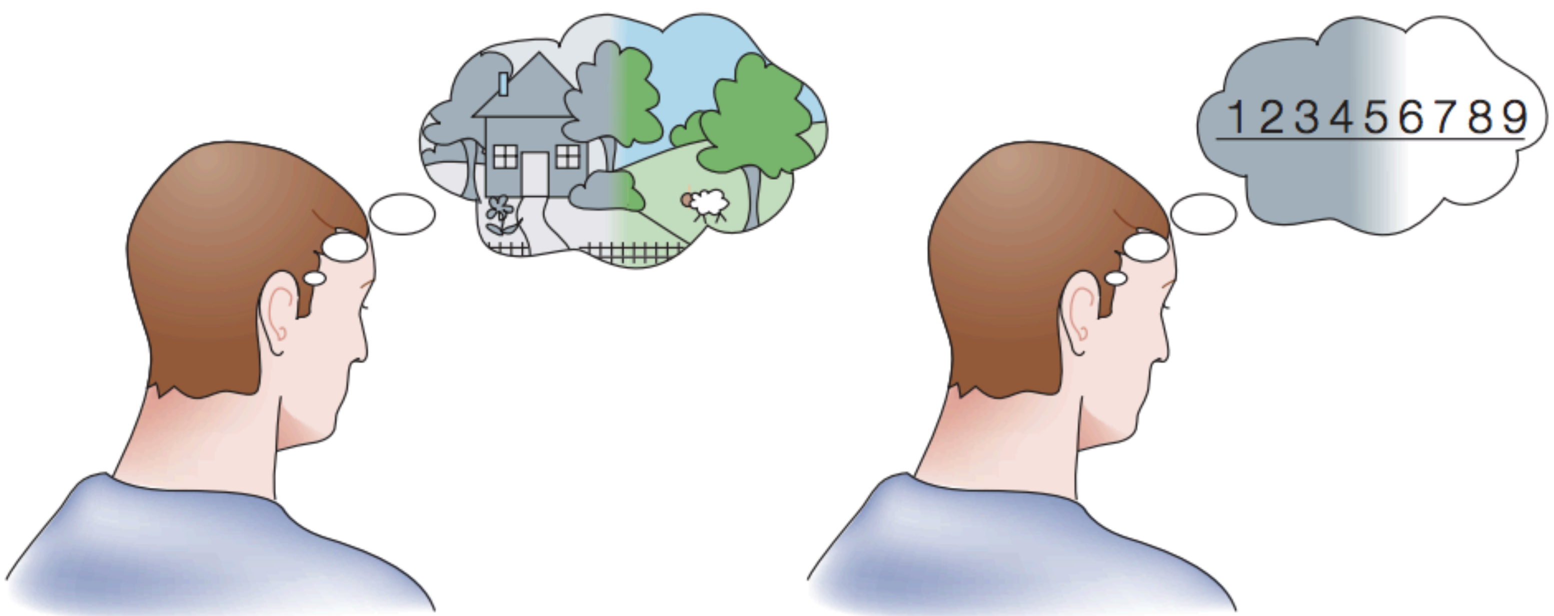
Line bisection effect

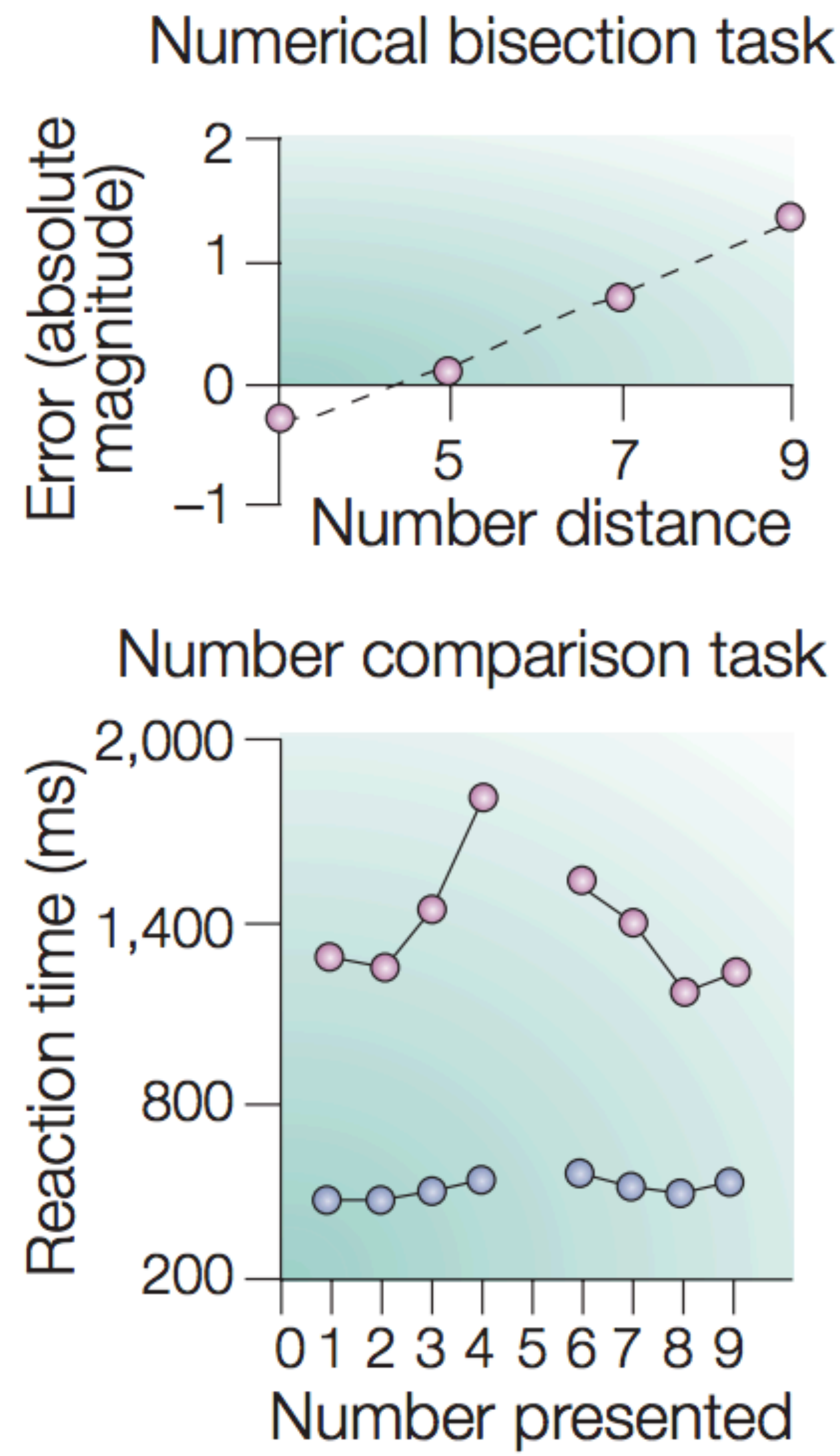


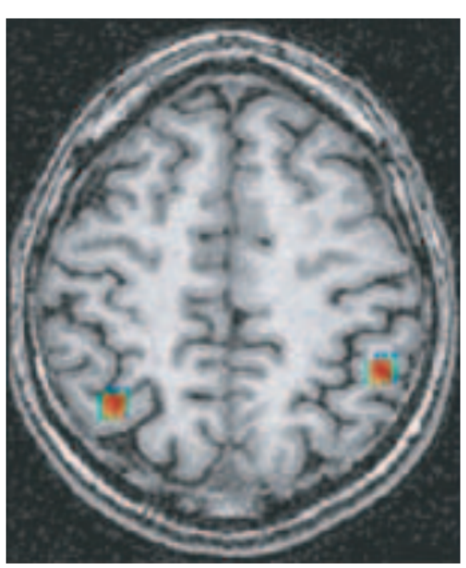
from *Hubbard et al. (2005)*

Number deficits in neglect patients

from *Hubbard et al. (2005)*

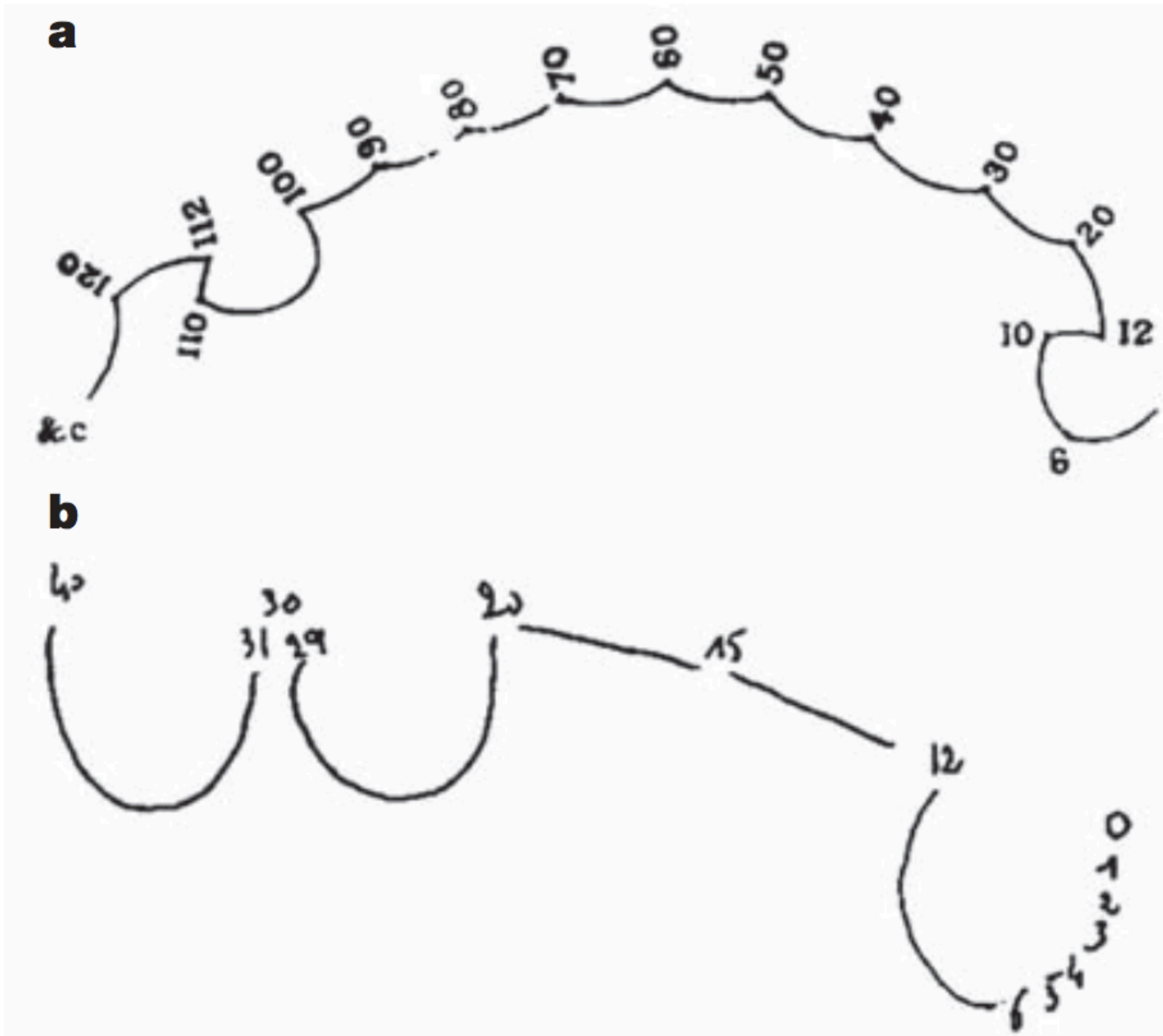






TMS studies too

Number forms

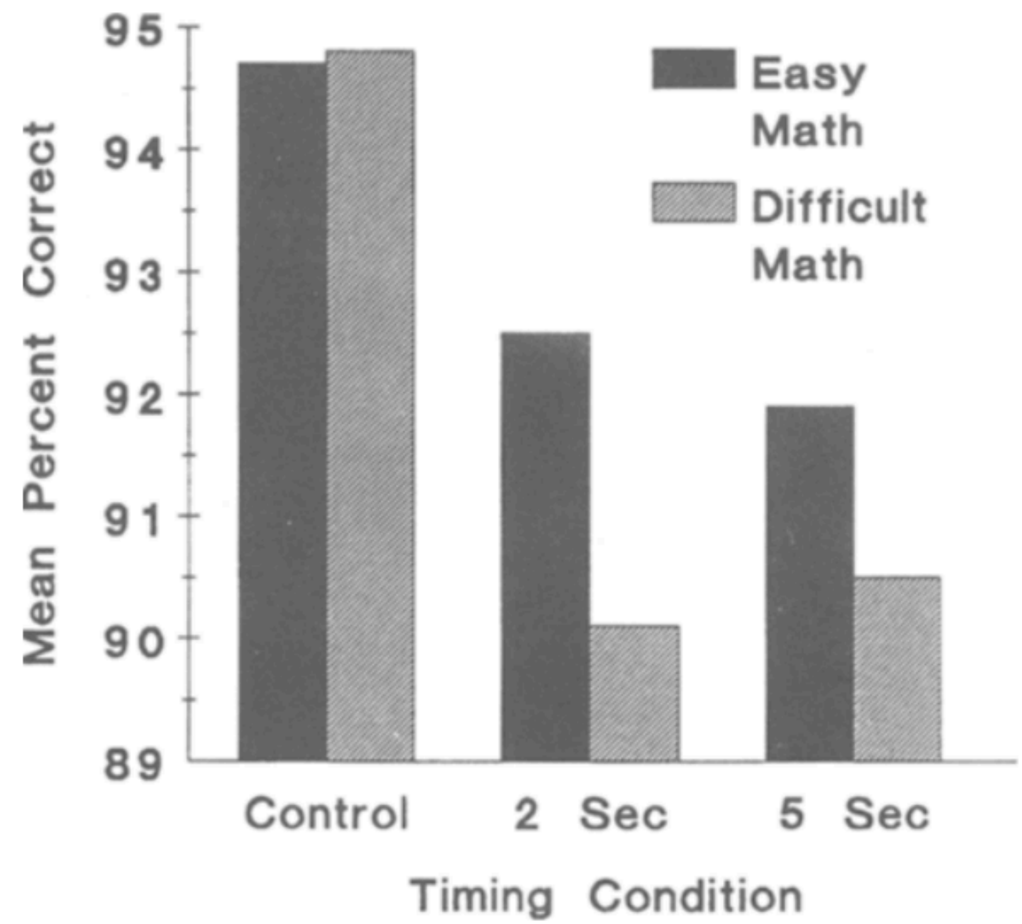


from *Hubbard et al. (2005)*

Time and number (quantity/size)

- Dual-task experiments
- Secondary number task impairs time estimation (Casini & Macar, 1997)
- Time impairs numerical task

(Brown, 1997)



-> time and number draw upon common magnitude mechanisms

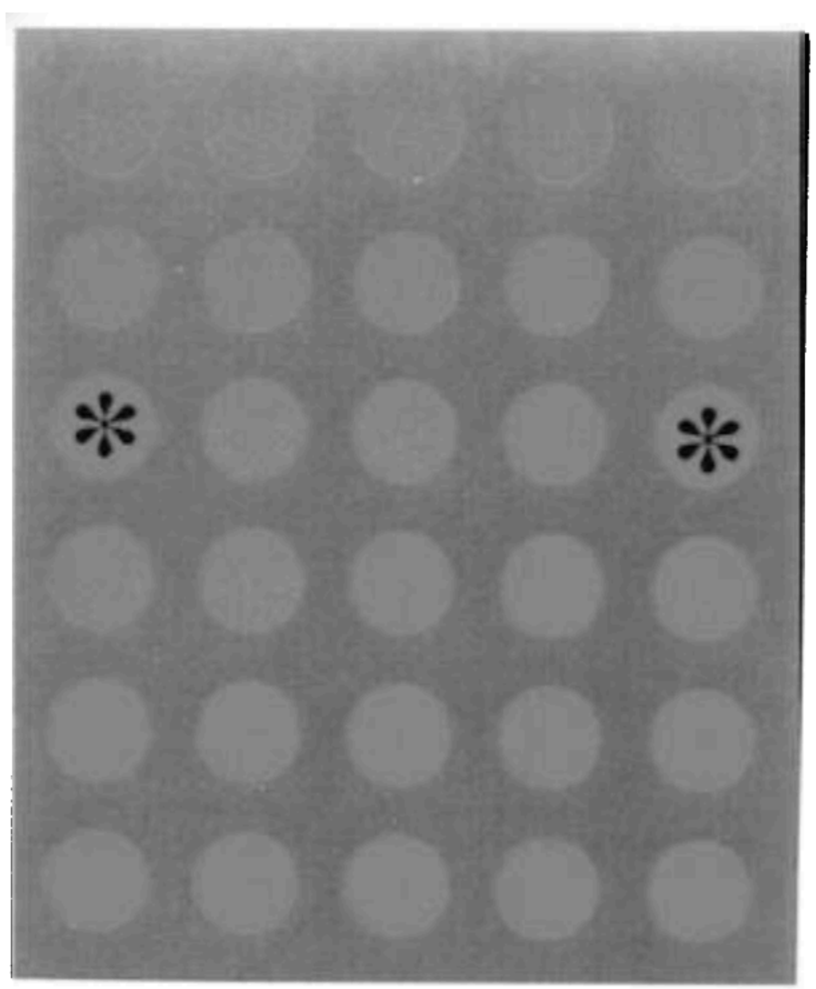
Time and space

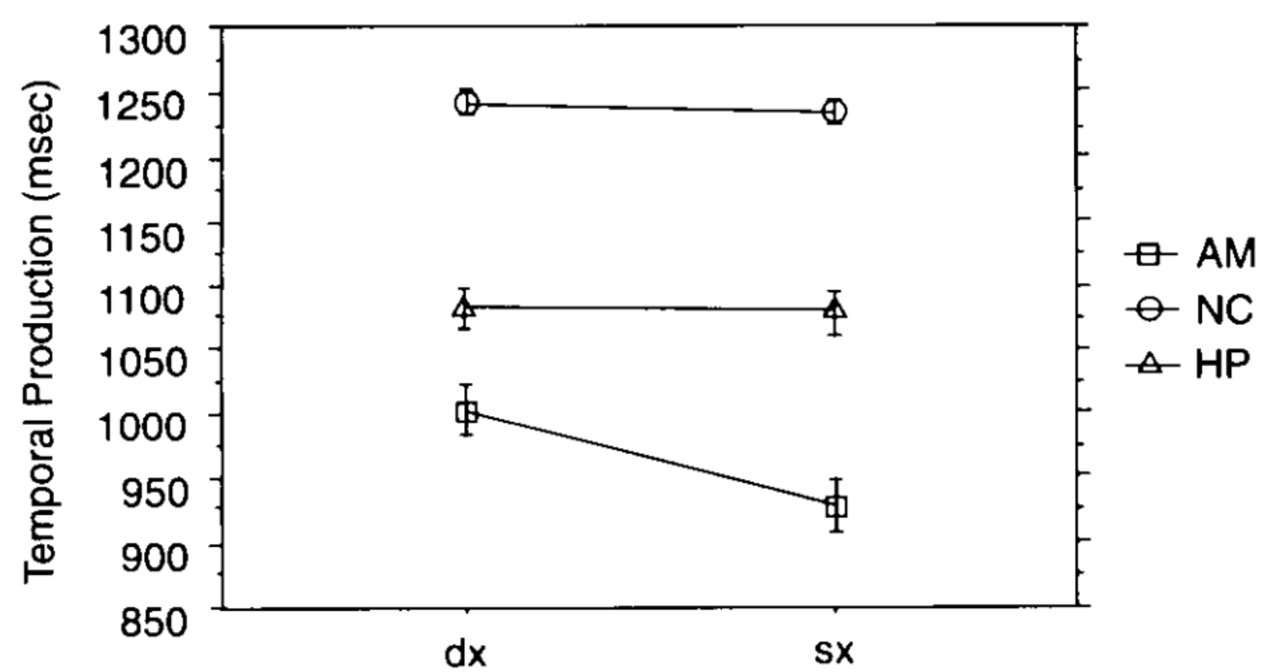
- De Long (1981): spatial scale mediates the experience of time and temporal duration
- participants observed different scale-model environments 1/6, 1/12, and 1/24 of full size; and asked to imagine as a scale figure engaging in 30 minutes activity inside the scale-model environments

Condition	Model scale	N	Elapsed time (min) ($\bar{X}$ ± S.E.M.)	Elapsed time (min) Range
			Experiment 1 (unmasked)	
Single exposure	1/6	20	4.15 ± 0.630	1.73 to 13.83
	1/12	166	2.52 ± 0.170	0.62 to 11.33
Exposure to two scales (same sample)	1/12	124	2.64 ± 0.133	0.35 to 9.75
	1/24		1.57 ± 0.085	0.17 to 4.92

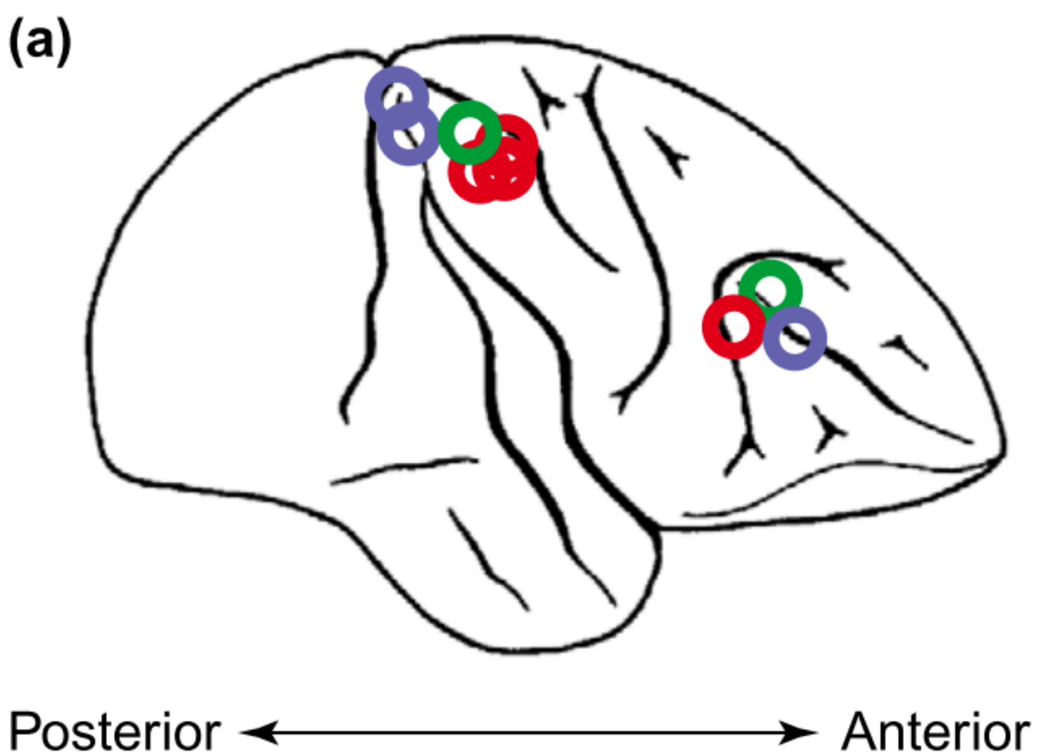
Time & space from neuropsychology

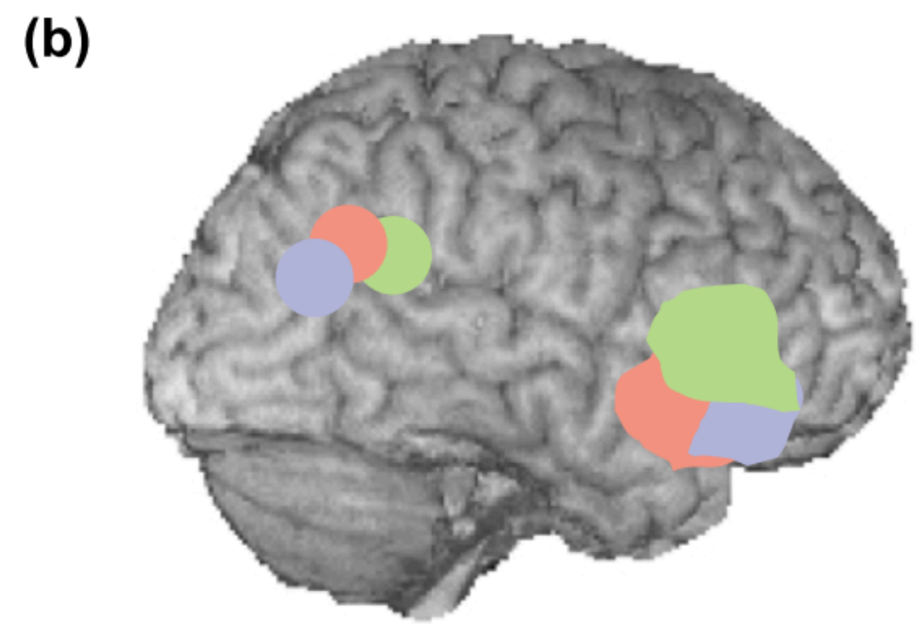
- right inferior parietal cortex (rIPC), is important for time perception
- Basso et al. (1996): spatial neglect patient with temporal deficits
- durations of stimuli presented in the neglected space consistently overestimated





Neuropsychology and brain imaging





Spatial
Numerical
Temporal

Walsh (2003)

Conclusion

- ATOM provides an explanatory account of why temporal processing may share mapping metrics with space, number, size, speed and action
- Discussion:
  - Convinced? Makes sense?
- Most likely, false

References

Scale-invariance

- **Chater, N., & Brown, G. D. . (1999). Scale-invariance as a unifying psychological principle. *Cognition*, 69(3), B17–B24.**
- Chater, N., & Brown, G. (2008). From universal laws of cognition to specific cognitive models. *Cognitive Science*, 32(1), 36–67.
- Kello, C. T., Brown, G. D. A., Ferrer-i-Cancho, R., Holden, J. G., Linkenkaer-Hansen, K., Rhodes, T., & Van Orden, G. C. (2010). Scaling laws in cognitive sciences. *Trends in Cognitive Sciences*, 14(5), 223–232.
- Maylor, E. A., Chater, N., & Brown, G. D. A. (2001). Scale invariance in the retrieval of retrospective and prospective memories. *Psychonomic Bulletin & Review*, 8(1), 162–167.
- Spence, A. J. (2009). Scaling in biology. *Current Biology*, *19*(2), R57-R61.
- West, G. B.,&Brown, J. H. (2005). The origin of allometric scaling laws in biology from genomes to ecosystems: Towards a quantitative unifying theory of biological structure and organization. *Journal of Experimental Biology*, 208, 1575–1592.

DbS

- **Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology*, 53(1), 1–26.**
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–291.
- Slovic, P. (1995). The construction of preference. *American Psychologist*, 50, 364–371.

ATOM

- **Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. Trends in *Cognitive Sciences*, 7(11), 483–488.**
- Basso, G., Nichelli, P., Frassinetti, F., & Di Pellegrino, G. (1996). Time perception in a neglected space. *Neuroreport*, 7(13), 2111-2114.
- Brown, S.W. (1997) Attentional resources in timing: interference effects in concurrent temporal and non-temporal working memory tasks. *Perception & Psychophysics*, 59, 1118–1140.
- Casini, L. & Macar, F. (1997) Effects of attention manipulation on perceived duration and intensity in the visual modality. *Memory & Cognition*, 25, 812–818.
- De Long, A.J. (1981) Phenomenological space-time: towards an experiential relativity. *Science* 213, 681–683
- Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience*, 6(6), 435–448.