Information & the Cognitive Sciences Workshop

Marius Usher
Tel-Aviv University

From Mutual Information and Bayesian decision models to Mental Representations (and Misrepresentation)

Natural Information and Mental Representations

Mental states are **intentional**: *they are about* or *represent* items in the world. **How do we determine the content of mental representation?**

- **Causation**: R represents objects that caused R; the mis-representation problem: "dog" caused by cat should represent cats; see Fodor (1990)
- **Mutual-Information** (Dretske, 1981), but with P(R|S)=1, it leaves no room for mis-representation; no distinction between information and veridicality (Floridi)

A Statistical Referential Theory of Content: Using Information Theory to account for Misrepresentation; **Usher M (2001).** ***Mind & Language*****.**

A **number of objections to probabilistic theories of content**:

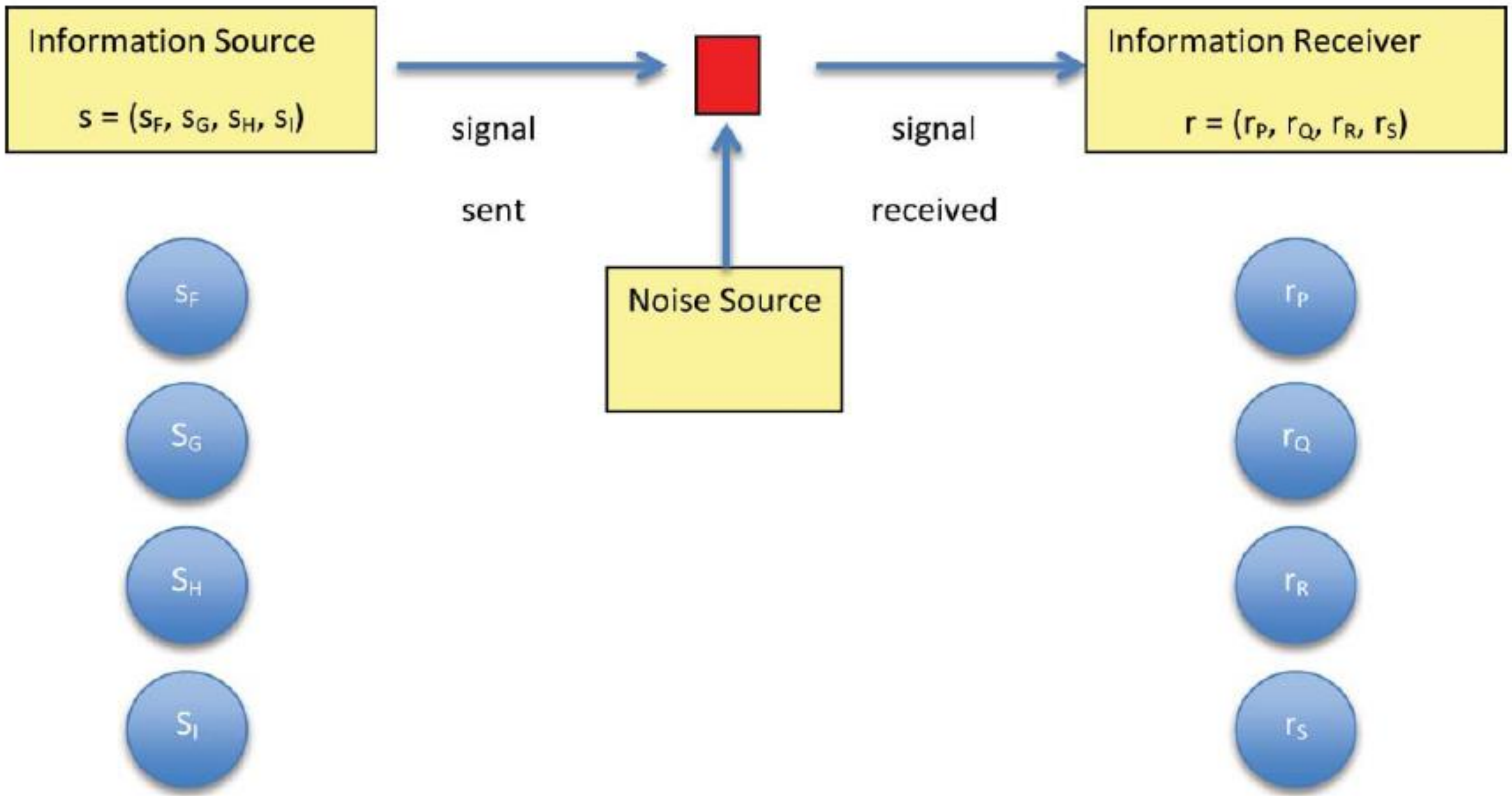
- The problem of arbitrary criteria
- The problem of bias in categorization (Milikan, 1989)

Can be solved on the basis of Local MI

$$MI_{ij} = \log \frac{P(R_i|S_j)}{P(R_i)} = \log \frac{P(R_i,\ S_j)}{P(R_i)\ P(S_j)} = \log \frac{P(S_j|R_i)}{P(S_j)}$$

The *Probabilistic Difference Maker Theory* (Scarantino 2015)

Fixes the natural meaning of a ***signal***



PDMT's Informational Content: That state 1 is incrementally supported/countersupported to degree $log_2 \frac{p(\textit{state 1} \mid \textit{signal \& background data})}{p(\textit{state 1} \mid \textit{background data})}$ and overall supported to degree p(state 1 | signal & background data) & … & that state n is incrementally supported/countersupported to degree $log_2 \frac{p(\textit{state n} \mid \textit{signal \& background data})}{p(\textit{state n} \mid \textit{background data})}$ and overall supported to degree p(state n | signal & background data)[16]

Plan of this talk

- Start from the PDM theory and apply it to content of mental representations: *use the mental representation as the signal;* how do we pick the content? (selection procedure)
- **Decision Neuroscience**: Bayesian algorithm that allows neural organisms to make use of stochastic samples of signals generated by objects in order to select representations that satisfy ***mutual-information*** conditions and to compute ***degrees of belief***
- Show this can solve problems related to decision biases and unequal priors e.g., **P1=.4**; P2=.1; P(R1|s) =**.48**; P(R2|s) = .42

Ratio1 = .48/.4 < **Ratio2 = .42/.1**

The likelihood of R1="tiger" may be high (danger) while tiger-frequency low (Millikan, 1989)

- This will require to distinguish between *degree of belief* and accuracies

Content of a Mental Representation

Content of mental representation → object mostly likely to have tokened the mental representation in a probabilistic process of perceptual categorisation; we assume the world is made of ***"object"*** entities (durable), O_j, not merely stimuli; assume also neural representation states, R_i, – *winner take all*

Content of a Mental Representation

Content of mental representation → object mostly likely to have tokened the mental representation in a probabilistic process of perceptual categorisation; we assume the world is made of ***"object"*** entities (durable),O_j, not merely stimuli; assume also neural representation states, R_i, – *winner take all*

- Examine matrix of ***conditional probabilities.*** *Example:*

$P(R_i|O_j)$

$$\frac{P(R_i,O_j)}{P(R_i)P(O_j)} = M_{ij} = \frac{P(R_i|O_j)}{P(R_i)} = \frac{P(O_j|R_i)}{P(O_j)}$$

O \ R	R1	R2	R3
O1	40	30	30
O2	30	40	30
O3	.39	0	61
P(R) =	109/300 .36	70/300 .23	121/300 .40

1.08	1.30	.75
.81	**1.74**	.75
1.07	0	**1.5**

Compare among Objects for given R-state

Maximal MI-value on the diagonal

Introducing response bias

Assume there is response bias to favor R1 (danger)

Simple guess model: $P'(R_i|O_j) = g\,\delta_{i,1} + (1-g)P(R_i|O_j)$

g=.5 –> $P'(R_i|O_j)$

O \ R	R1	R2	R3
O1	**70**	15	15
O2	65	**20**	15
O3	69	0	**31**
P(R) =	204/300 **.68**	35/300 **.12**	61/300 **.20**

M_{ij}

1.03	1.30	.75
.96	**1.74**	.75
1.01	0	**1.5**

- Even with response bias, MI picks content on the basis of best match among objects for given R-state. Given R_2, the ratio likelihood of O_2 is larger that the ratio likelihood for O_1.
- This follows from the fact that $P(R_2|O_2) > P(R_2|O_1)$
- All we need to worry is forward (causal) probabilities for each R

The decision mechanism

How does the decision system compute its best guess of O_i? How is the $P(R_i|O_j)$ obtained?

- The standard assumption: Each O^i, generates a temporal sequence of stimuli, $\{x^i_1, x^i_2, \ldots x^i_t\}$ based on some generative distribution (e.g., Normal)
- The **ideal observer decision problem**: select object that is the most likely, given evidence $e=\{x_1, x_2, x_t\}$ and priors, $P(O_i)$.

Treat each O^i as an Hypothesis and compute Ratio-likelihood for posteriors. Assume the case of n=2:

r = **P(O1|e)/P(O2|e)** > 1

Bayes rule:

P(e|O1)/P(e|O2) * P(O1)/P(O2) = **P(O1|e)/P(O2|e)**

Signal detection with multiple samples of evidence

Likelihood ratio with multiple evidence, e1, e2, ...

$$LR_{1,2|e_1,e_2,\ldots,e_n} = LR_{1,2|e_1} \cdot LR_{1,2|e_2} \cdots LR_{1,2|e_n}$$

Decision rule

$$LR_{1,2|e_1} \cdot LR_{1,2|e_2\ldots} \cdot LR_{1,2|e_n} \cdot \frac{\Pr(h_1)}{\Pr(h_2)} \cdot \quad >1$$

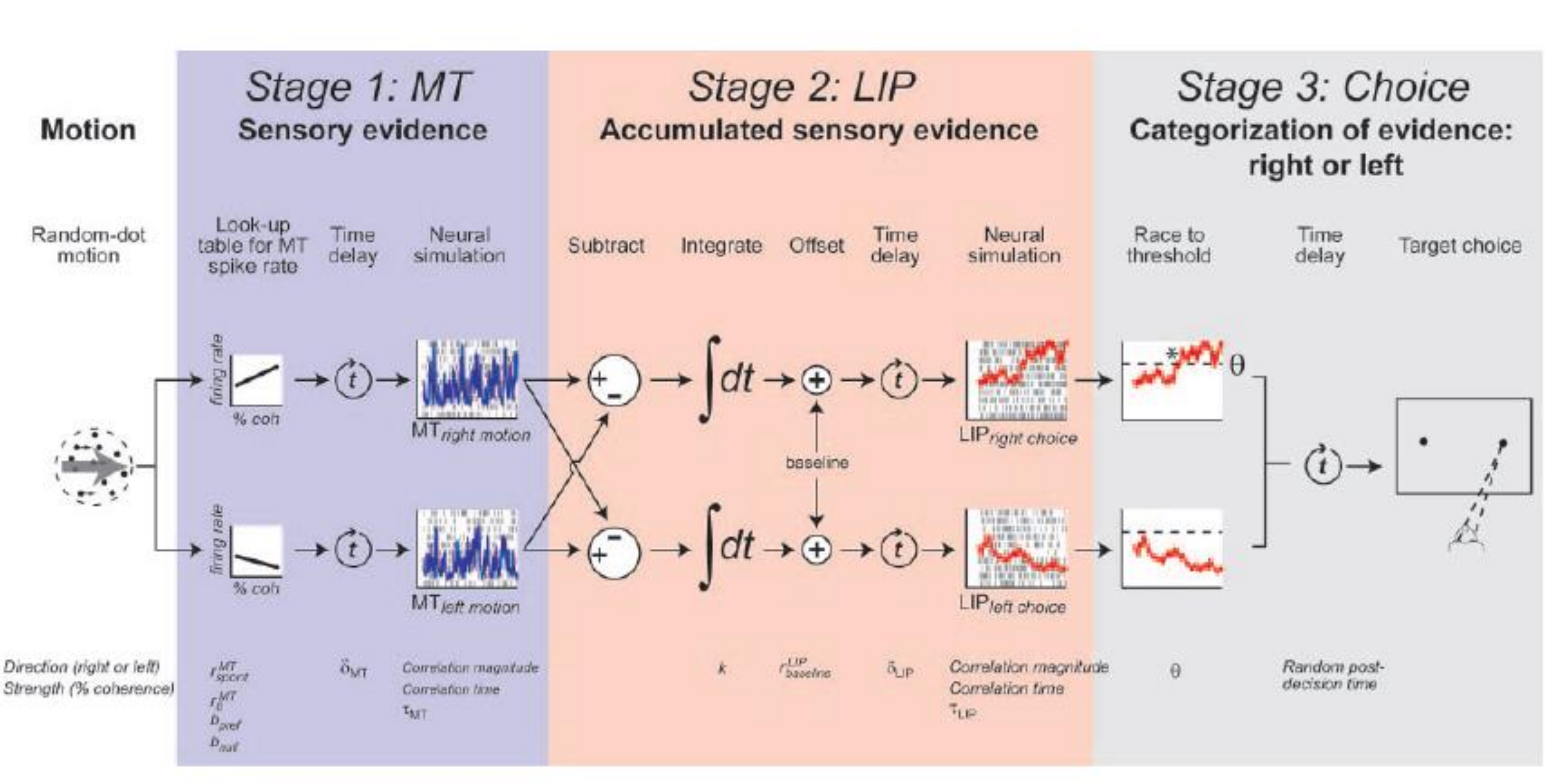
Take Logs

$$\log LR_{1,2|e_1} + \log LR_{1,2|e_2\ldots} + \log LR_{1,} + \log\left[\frac{\Pr(h_1)}{\Pr(h_2)}\right] \cdot \quad >0$$

Integrate evidence until the ratio-likelihood reaches a desired value, say 4/1, corresponding to P(Hi) = 80%.

Neuroscience model of perceptual decisions

(Mazurek, Roitman, Ditterich & Shadlen, 2003; *Cerebral Cortex*)



Decision system integrates evidents (ratio likelihoods) and selects the hypothesis (Ri) that has the highest posterior probability

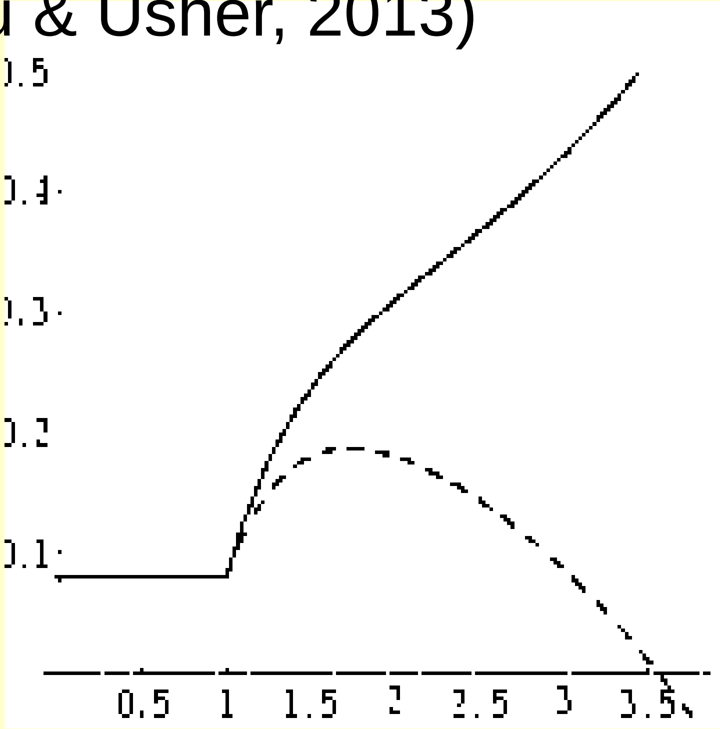
Leaky competing accumulators

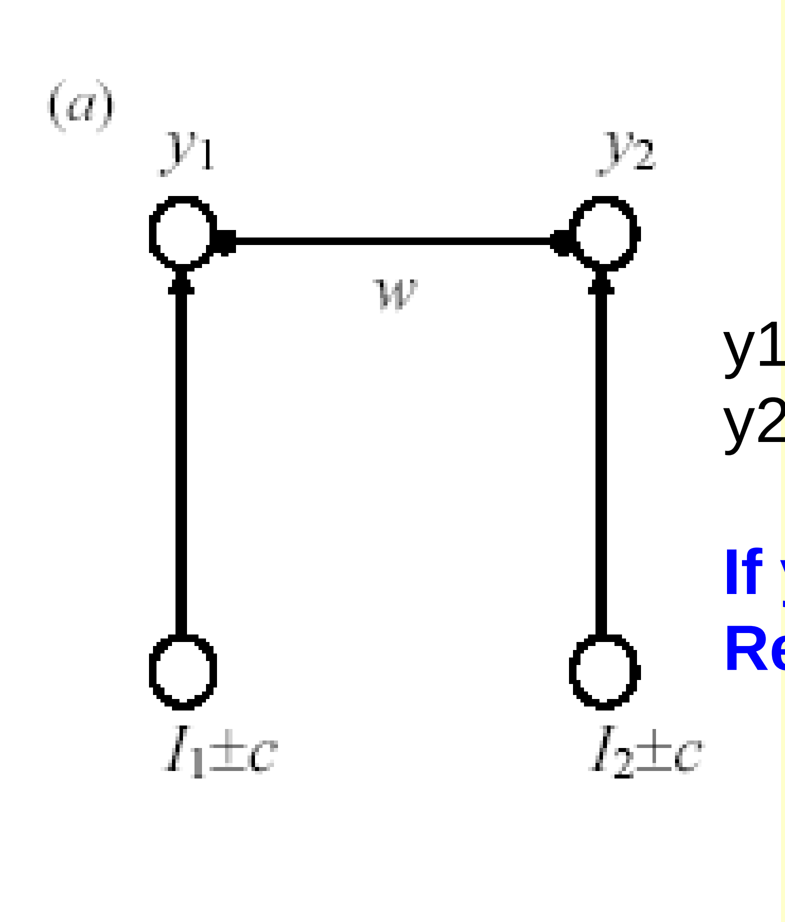
(Usher & McClelland, 2001; Teodorescu & Usher, 2013)

Mutual inhibition; decay of activation;

nonlinear activation function

Common response criterion





y1(t+dt) = y1(t) + dt[I1 – k*y1 -b*y2] + noise_1
y2(t+dt) = y2(t) + dt[I2 – k*y2 -b*y1] + noise_2

If y_i <0, then y_i =0.
Respond when y_i > threshold

Decision bias and priors

- Decision system integrates evidence (ratio likelihoods) and selects the hypothesis (R_i) that has the highest posterior probability
- But assume unequal priors, $P_i(0)$, and evidence e = {s1, .. sn}

P_1=.4; P_2=.1, P3=...=P7=.1; P(R_1|e) =**.55**; P(R_2|e) = .45;
Ratio1 = .55/.4 < **Ratio2 = .45/.1**

This can happen because the decision algorithm includes the ratio of prior terms; without it we have prior neglect (suboptimal)

- **A riddle:** What is the natural information of e? Favours O1 or O2?

PDMT may appear to favour O2? (P(O2|e)/ P(O2) is largest)

However, the organism has just committed to O1, it cares about priors too... (it should if it wants to survive).

- Solution: we need to distinguish between "objective" conditional probabilities, $P(R_i|O_j)$, and "subjective" degrees of belief, $P(O_i| e)$

Subjective vs Objective Information

- The subject has access to a sequence of stimuli, e = {s1, .. sn}, and based on this it makes an informed guess, about the O_i
- This guess takes priors into account; the organism cares of maximal posterior probability more than it does about ratio likelihoods, $P(O_i|e)/P(O_i)$

- However, our question is not to select the ***content of e***, but ***rather of R***$_i$. Thus we need to condition on R_i and discard e. When we do this, all that matters is the forward (causal) conditional probabilities: $P(R_i|0_1)$, $P(R_i|0_2)$. And those satisfy the MI representation condition, even when there are priors or decision biases at play

This scheme allows for R1 to represent O1 even if in the actual case, R1 was triggered by O2 (mis-representation)

Conclusions

- Local Mutual Information (Shannon) allows us to construct a procedure that picks the content to representation-states, as long as the conditional probability of Ri is maximal for Oi (compared with other O's); consistent with PDMT, while taking the R-state as the signal (not the stimulus/evidence).

- This scheme allows to account for representation even under response bias (as per Millikan, 1989; danger bias)

- The scheme does not associate content with causation (or “veridical information”, as per Dretske, 1981) and thus it has room to account for mis-representations

- Since the scheme is competitive is does not rely on arbitrary criteria.

Thanks to Rani Moran for Bayesian consultancy

Funding

German-Israel Foundation; Israel Science Foundation

Israel Science Foundation; ICORE in Cognitive Science