Encoding Rules in the Brain

Daniel S. Levine
Department of Psychology
University of Texas at Arlington
Arlington, TX 76019-0528, USA
levine@uta.edu
www.uta.edu/psychology/faculty/levine/levine.htm

Article and book chapter

Angels, devils, and censors in the brain, *ComPlexus*, 2, 35-59, 2005, http://www.uta.edu/psychology/faculty/levine/downloads/complexuspageproofs102405.pdf

How does the brain create, change, and selectively override its rules of conduct? In R. Kozma & L. Perlovsky (Eds.), *Neurodynamics of Higher-level Cognition and Consciousness*. Heidelberg: Springer-Verlag, 2007,

http://www.uta.edu/psychology/faculty/levine/downloads/kozmaperlovskybookchapter121106.doc

Conference proceedings

Levine and Perlovsky, A network model of rational versus irrational choices on a probability maximization task, Proceedings of WCCI2008 (World Congress on Computational Intelligence, June, 2008, Hong Kong)

The need for rules

Brains, or any complex high-order cognitive systems, require rules for what actions to perform and what actions to refrain from performing.

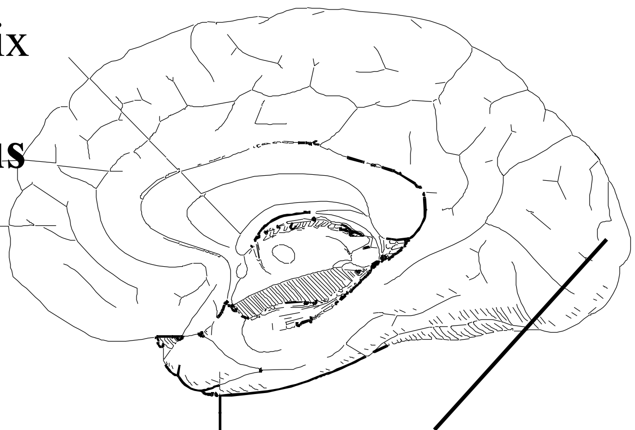
The more complex the system's environment, the more flexible and context-sensitive those rules need to be.

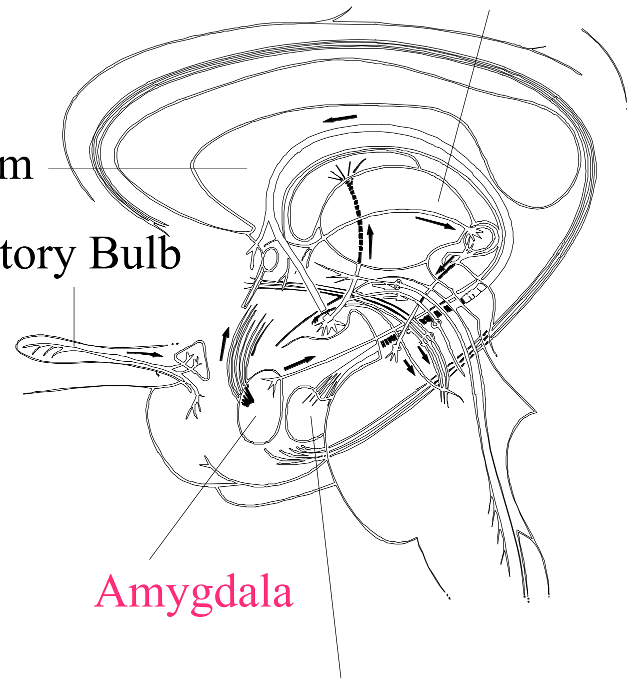
Brain regions involved in rules

Hypothalamus and midbrain: encoding primary needs of the organism (food, water, sex, and maybe social bonding).

Amygdala: closely connected with hypothalamus and midbrain but "one step up" in evolution. Attaches positive or negative emotional valence to specific sensory events or motor actions.

Where is the amygdala?





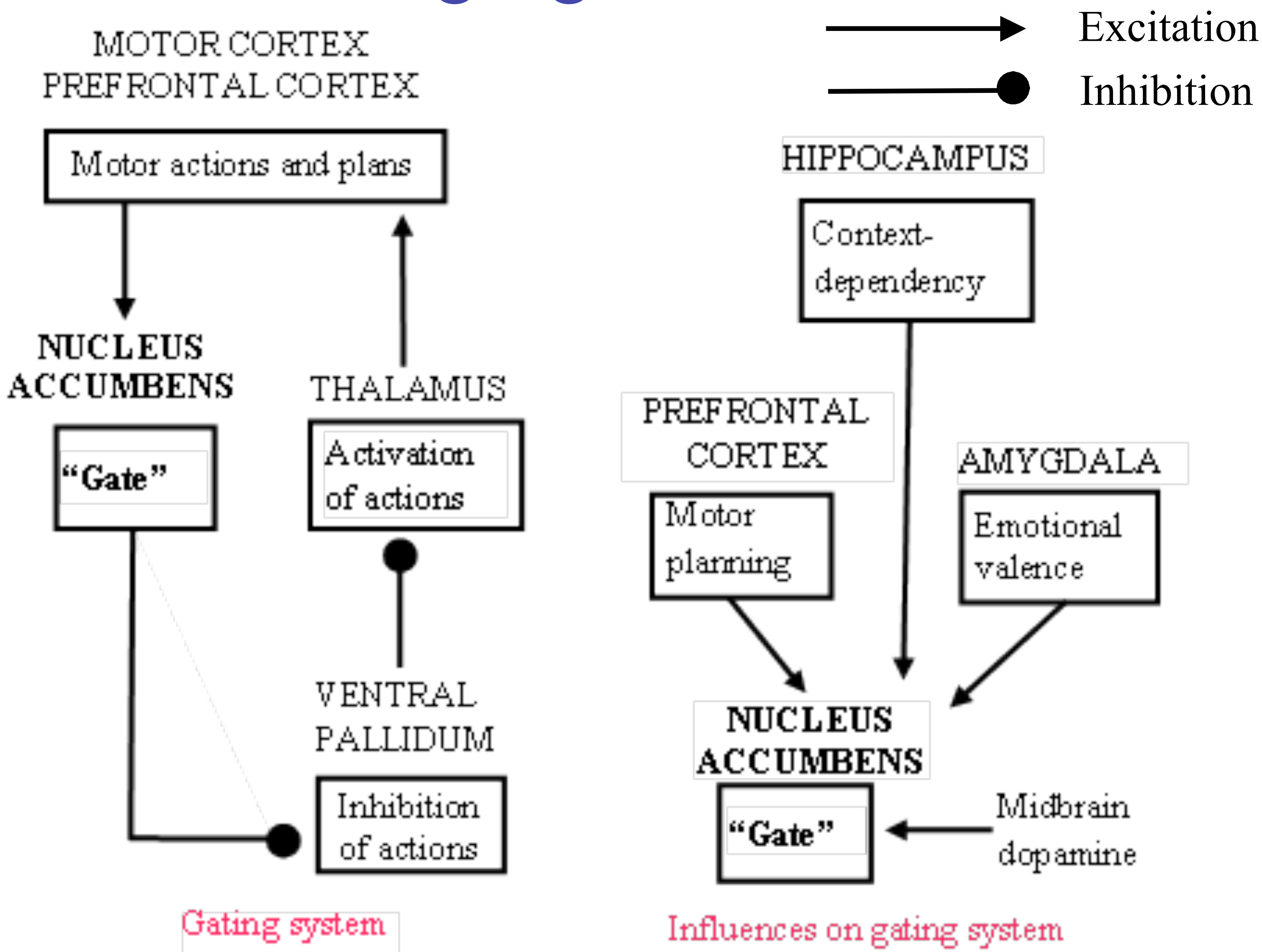
Adapted from Nieuwenhuys, Voodg, & van Huijzen, 1981, with permission of Springer-Verlag.

Basal ganglia and thalamus: "gating"

How do we translate positive and negative emotional linkages into action tendencies or avoidances?

Gating system: brain network that selects sensory stimuli for processing and motor actions for performance. Embodied in pathways between prefrontal cortex, basal ganglia, and thalamus. Link from basal ganglia to thalamus plays role of *disinhibition*: allowing (based on contextual signals) performance of actions whose representations are usually suppressed.

Nucleus accumbens: motivational part of basal ganglia



And on top of all this, the prefrontal cortex executive system

Prefrontal cortex is association area of the frontal lobes.

Last part of the brain to develop, both in evolution and the individual (not fully wired until one is in the mid-20s).

Major area for planning, often called executive of the brain (Pribram and Luria, 1973). Three key executive regions: *orbital* (OFC), *anterior cingulate* (ACC), *dorsolateral* (DLPFC).

Locations of prefrontal areas

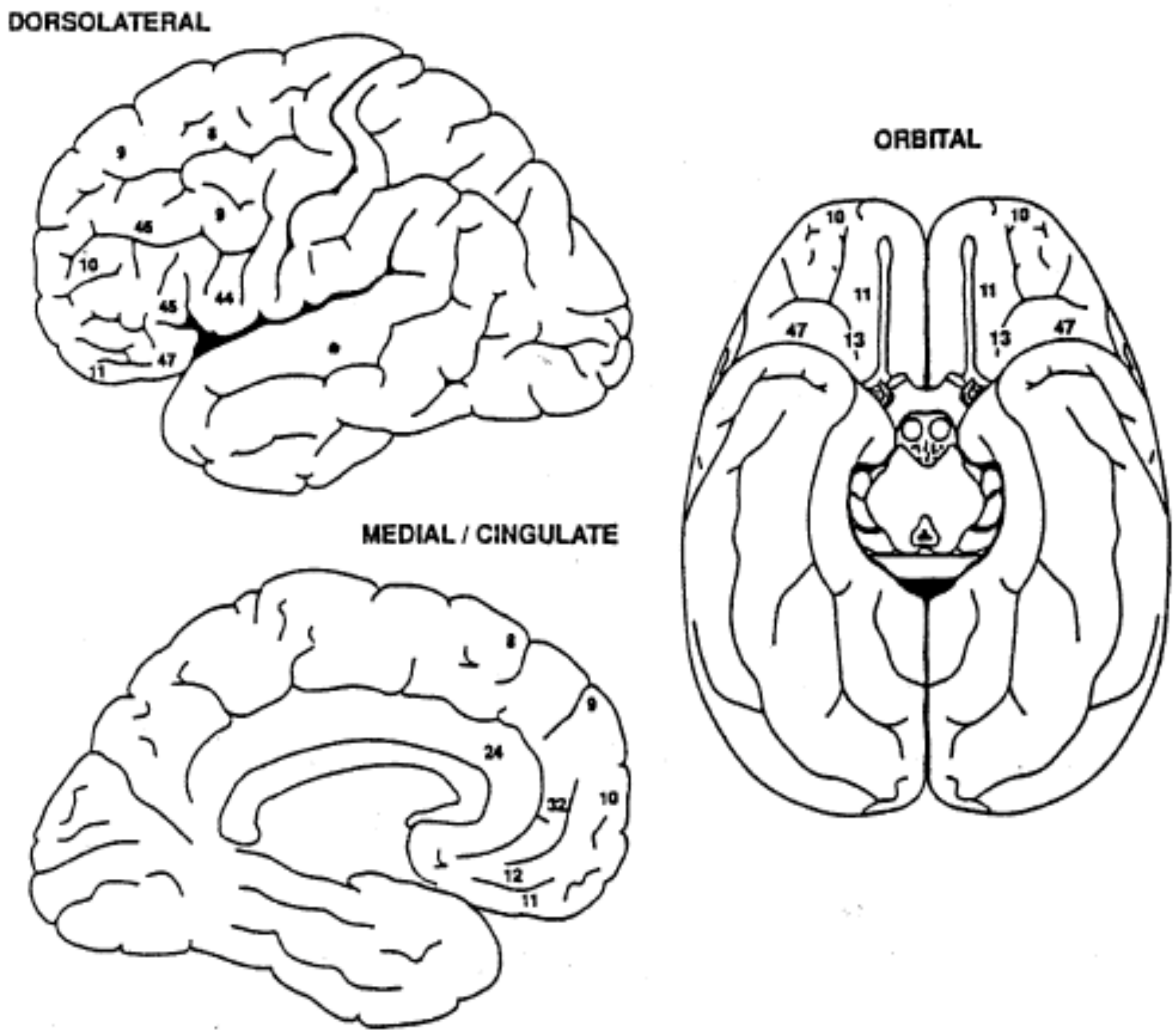


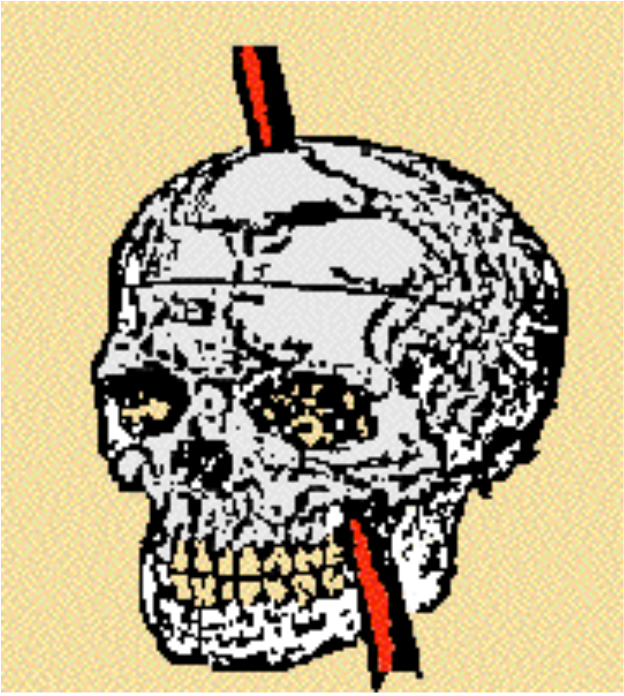
Fig. 1. Fuster diagram. Three schematic views of the human brain with frontal cytoarchitectonic areas indicated according to Brodmann's map. From J. Fuster (1999), Guilford Press.

Functions of orbital prefrontal

Orbital (OFC): forms and sustains mental linkages between sensory events, or motor actions, and positive or negative emotional states.

Long-term storage of emotional valences is at synapses between OFC and amygdala (Schoenbaum et al., 2003).

The 19th century patient Phineas Gage lost planning and social appropriateness after a railroad accident. Modern reconstruction (Damasio, 1994) showed Gage's primary area of damage was OFC.



Functions of anterior cingulate

Anterior cingulate cortex (ACC): involved in selection or switching among different interpretations or aspects of a stimulus (Posner & Petersen, 1990).

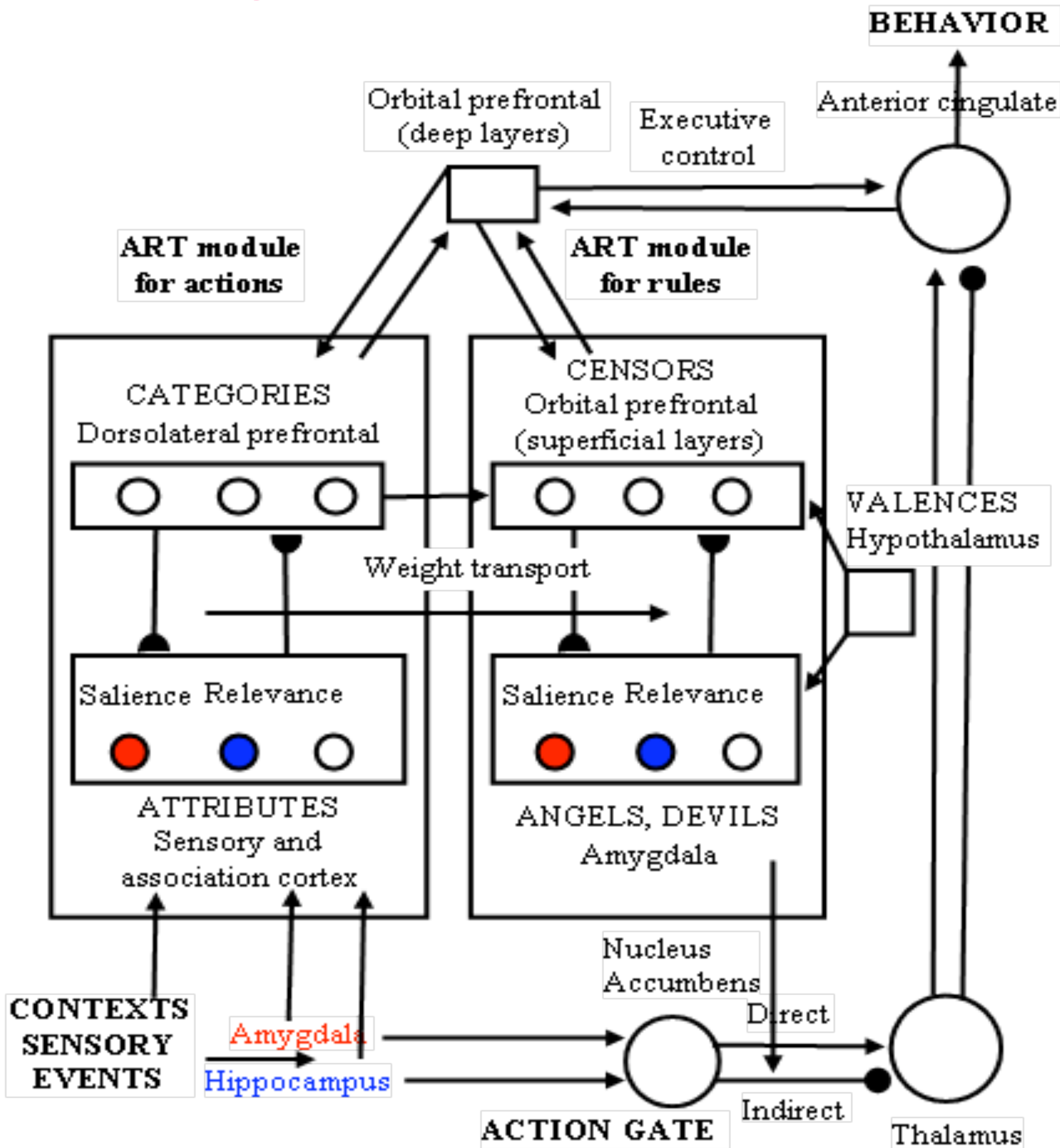
Attentional task with emotional distractors: ACC is only area activated by BOTH targets and distractors (Yamasaki, LaBar, & McCarthy, 2002).

Theories: detection of potential response error or of response conflict (Botvinick et al., 2001; Brown & Braver, 2005)

Functions of dorsolateral prefrontal

- Dorsolateral prefrontal cortex (DLPFC): complex working memory region. Involved in information processing at a higher level of abstraction than the OFC.
- Monkeys: OFC lesions in monkeys impair learning of changes in reward value within a stimulus dimension; DLPFC lesions impair learning of changes in which dimension is relevant (Dias, Robbins, & Roberts, 1996)

Network relating behaviors to action tendencies



But there can be competing rules!

Knowledge instinct (KI): a biologically driven impulse to make coherent sense of the world at the highest level possible (Perlovsky, 2001, 2006).

Yet decision making results suggest a contrary biological drive for effort minimization (EM) by solving problems using simplifying heuristics.

KI vs. EM varies between individuals and contexts.

Brain involvement

Neuroimaging studies:

Using simple heuristics (EM) activates amygdala (primary emotional experience).

More complex decisions (KI) activate:

- Orbitofrontal cortex (OFC) (cognitive-emotional linkage);
- Anterior cingulate cortex (ACC) (detecting risk or conflict);
- Dorsolateral prefrontal cortex (DLPFC) (working memory, complex ideas).

Is there a drive to know?

Perlovsky (2001, 2006): drive to create realistic internal models of the world.

This involves a search for consistent mental representations across different hierarchical levels of the brain

What is the scientific evidence for a knowledge instinct?

1. Exploratory drive in humans and monkeys

Curiosity about their environments regardless of instrumental value

Harlow (1953): monkeys learn to solve mechanical puzzles when no external reinforcement is provided other than the puzzle itself

2. Cognitive dissonance

Festinger (1957): if people find cognitive dissonance between actions and beliefs, they may try to (a) change actions, (b) change beliefs, or (c) synthesize.

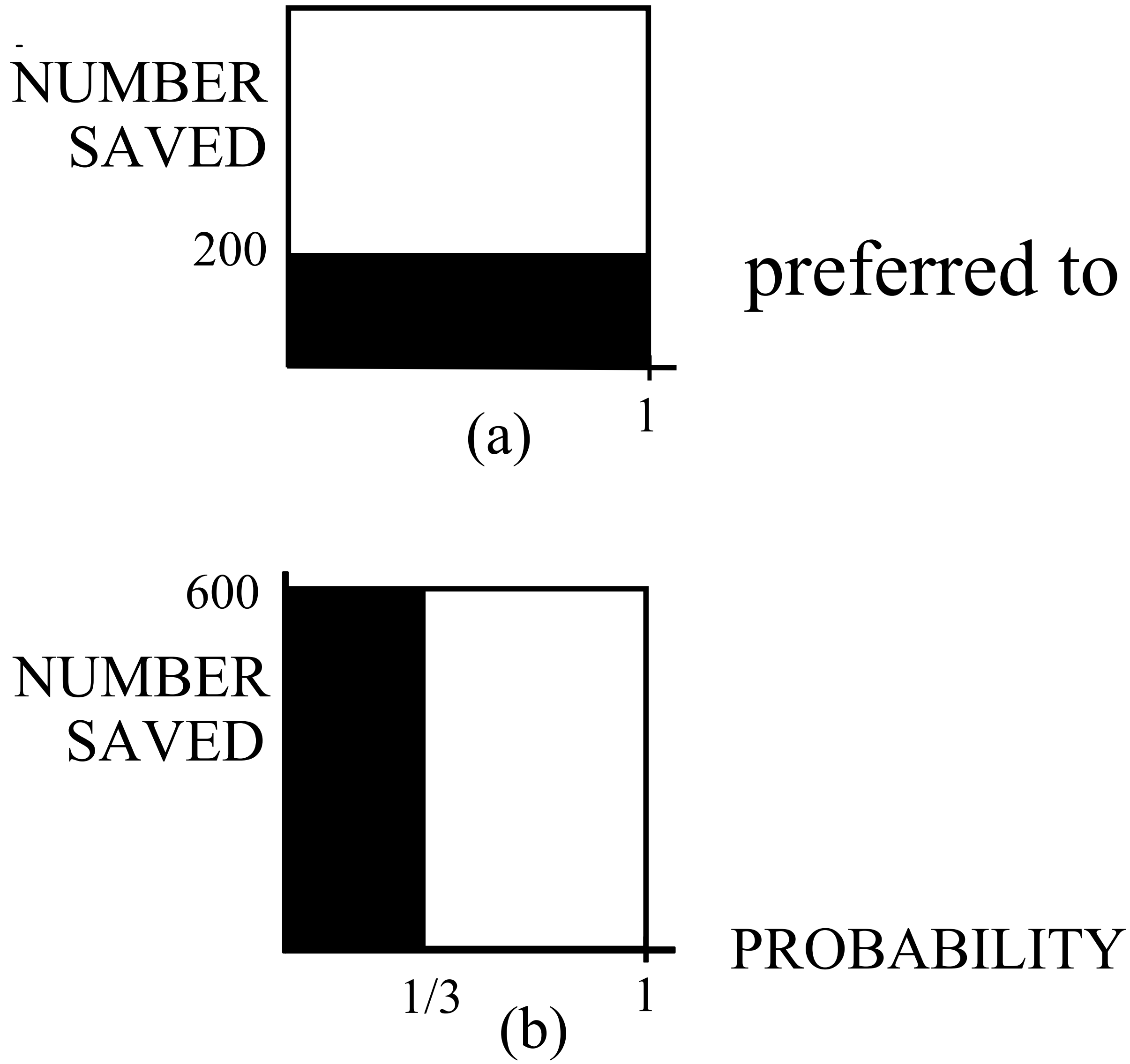
Elkins and Leippe (1986): signs of physiological discomfort (skin conductance) may remain with change of beliefs. **Is there less discomfort with synthesis? My laboratory may test this!**

Results from cognitive neuroscience

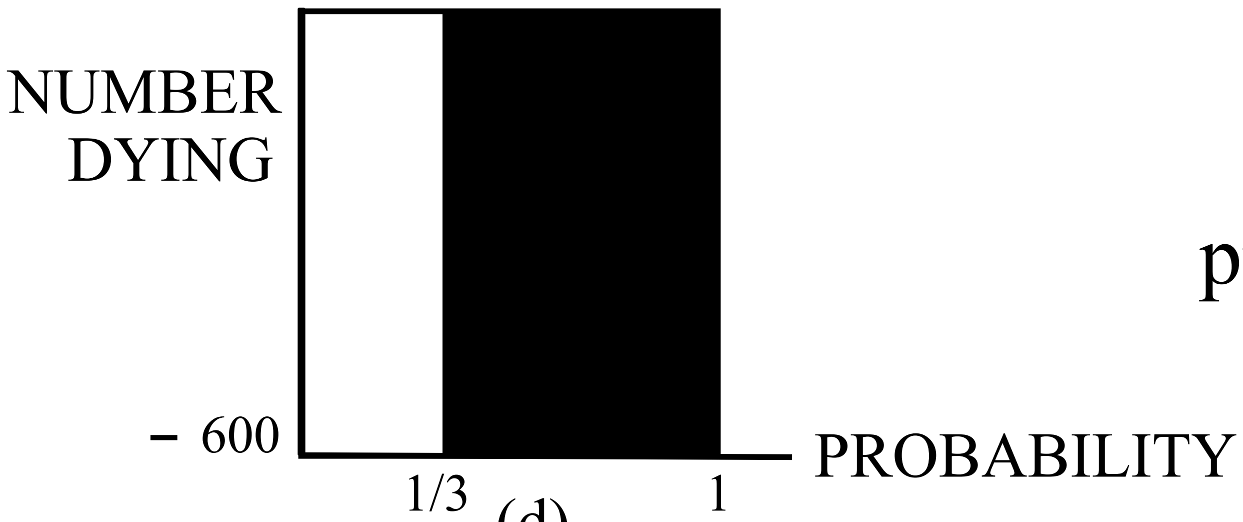
Different brain processes can be involved in individuals who follow heuristics versus those who violate heuristics.

Example: "Asian disease problem" due to Tversky and Kahneman (1974, 1981). Subjects are asked to consider two programs to combat an Asian disease expected to kill 600 people.

Save 200 for sure or 600 with probability 1/3?

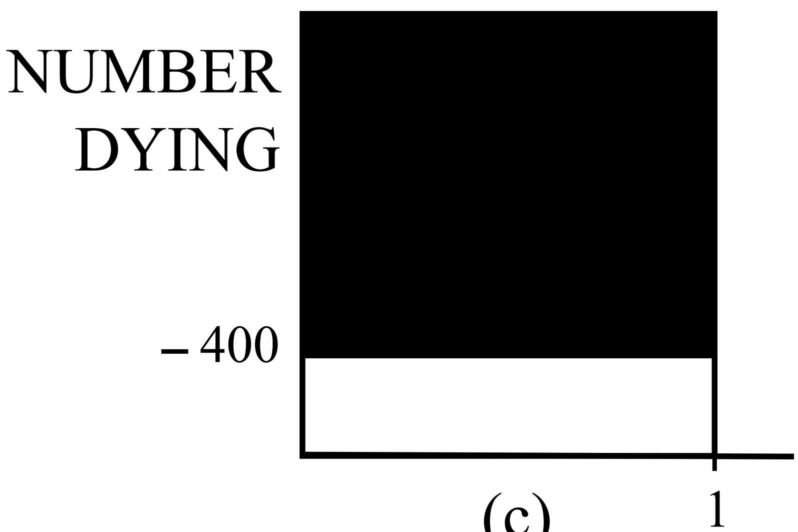


400 die for sure or 600 with probability 2/3?



(d)

preferred to



(c)

These choices are inconsistent!

(a) is preferred to (b), (d) is preferred to (c). But

200 out of 600 saved for sure (a)

is identical to

400 out of 600 die for sure (c).

1/3 chance all saved, 2/3 chance none saved (b)

is identical to

2/3 chance all die, 1/3 chance none die (d)

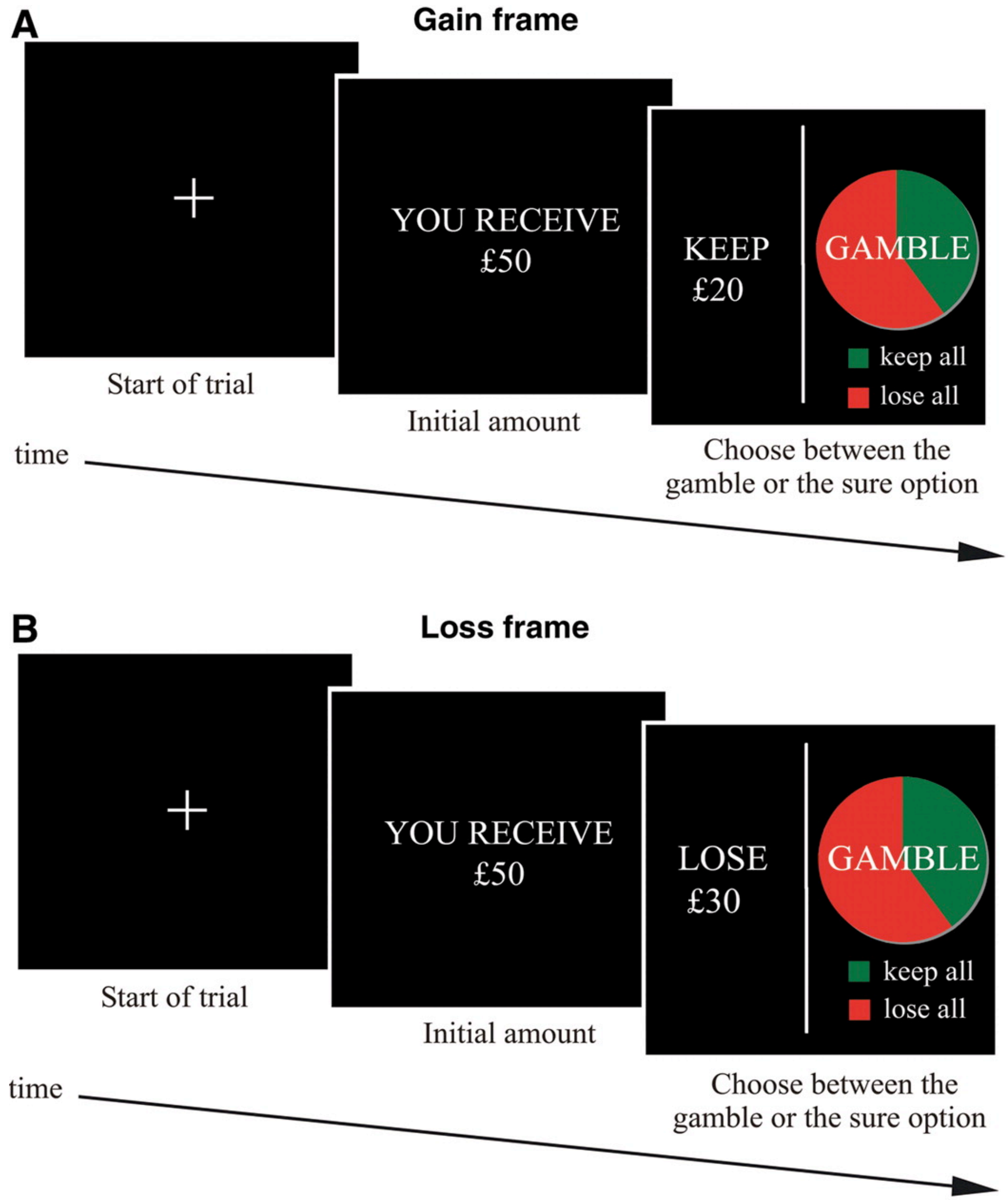
Gain vs. loss frame changes preference!

fMRI study (DeMartino, Kumaran, Seymour, & Dolan, 2006)

Monetary variant of “Asian disease”:

All participants are given the same amount of money at first.

Then they are given a choice of a sure thing or a gamble, described either in terms of what they keep (gain frame) or in terms of what they lose (loss frame): see next slide.



From DeMartino et al., Science, 313:684-87, 2006.

Brain activation patterns

Gamble chosen about 60% of time in loss frame, 40% in gain frame.

Those who DIDN'T follow heuristics (risk seekers on gains or risk avoiders on losses) had more activation of orbitofrontal cortex (OFC) and anterior cingulate cortex (ACC).

Those who DID follow heuristics (risk-seeking for losses, risk-averse for gains) showed more activation of the amygdala.

(Oversimplified) Functional summary

Knowledge maximization tends to activate prefrontal executive regions

vs.

Simplifying heuristics tend to activate primary emotional regions

AUTOMATIC vs. CONTROLLED processing (Shiffrin & Schneider, 1977; Shallice, 1988)

Both modes are needed

What is the network parameter that mediates switches between modes?

Some form of

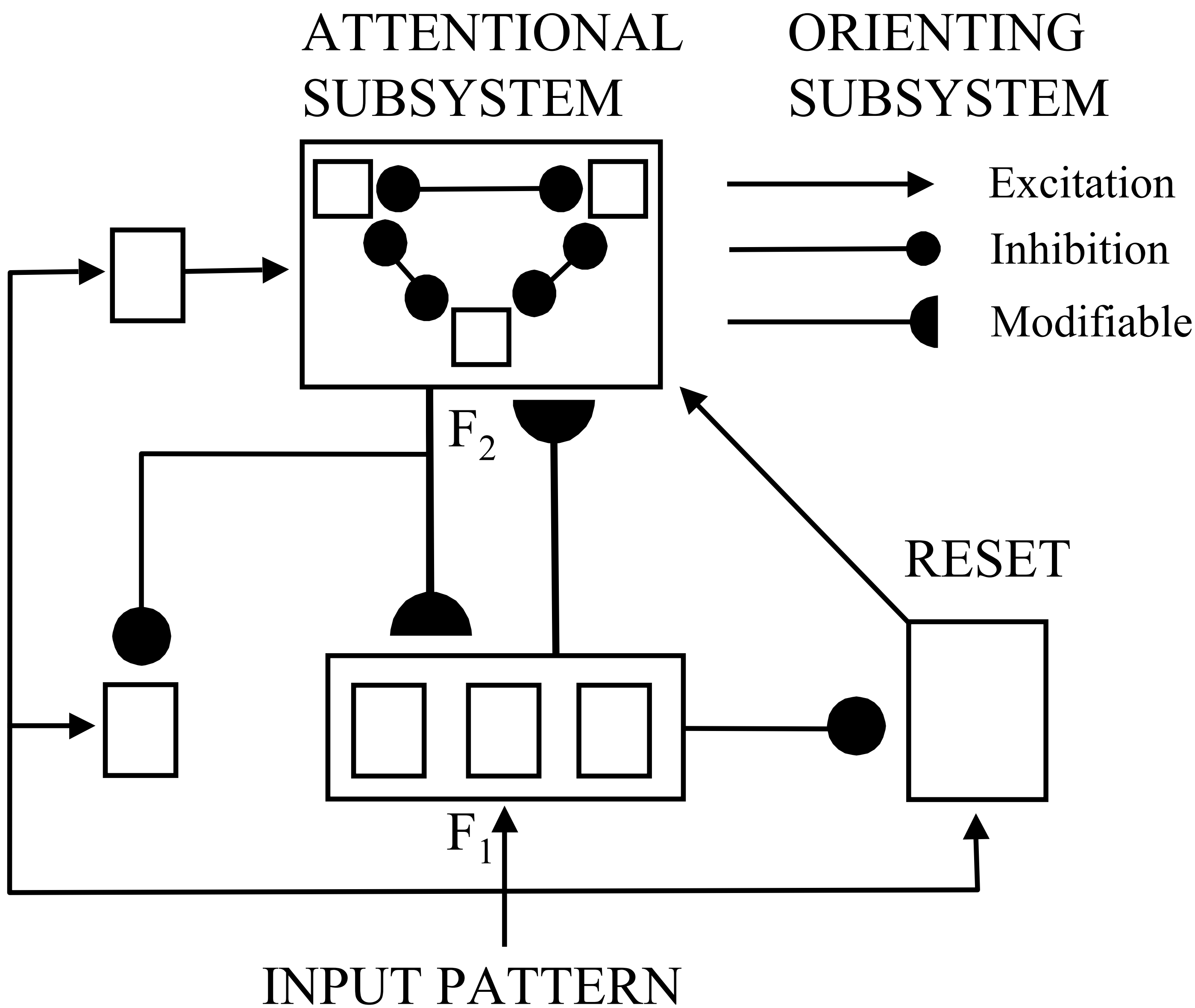
VIGILANCE (ART: Carpenter & Grossberg, 1987, etc.)

Basic ART heterarchical structure (can repeat at many levels):

Two interconnected layers of nodes, called F_1 ("bottom") and F_2 ("top") (diagram on next slide).

F_2 nodes respond to categories of F_1 node activity patterns.

Learning: synapses between the two layers are modifiable in both directions.


ATTENTIONAL SUBSYSTEM
ORIENTING SUBSYSTEM
Excitation
Inhibition
Modifiable
F2
RESET
F1
INPUT PATTERN

Match vs. mismatch

Inhibition from F_2 layer to F_1 shuts off most neural activity at F_1 if there is mismatch between the input pattern and the active category's prototype.

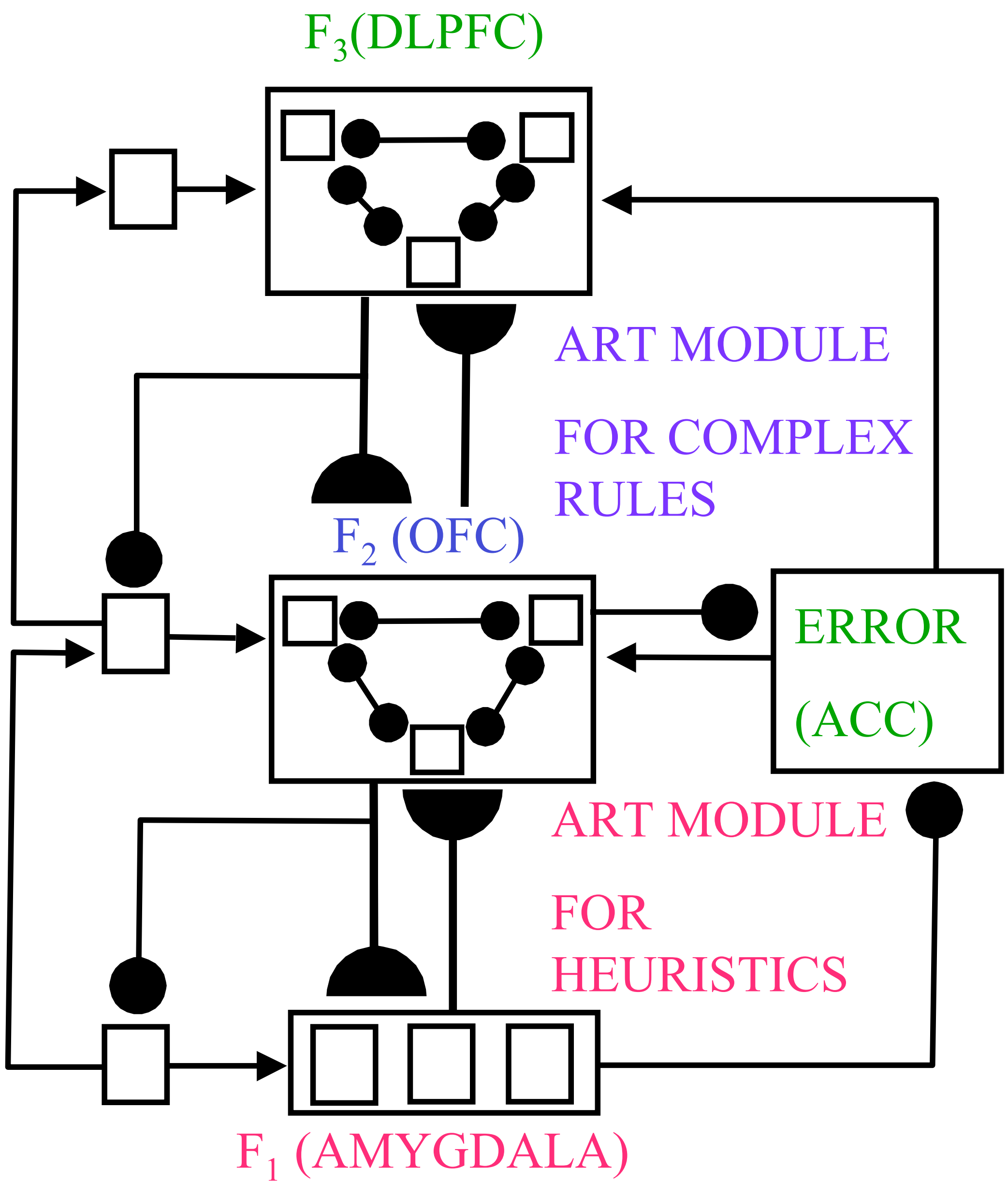
Only with a sufficiently large match are enough of the same F_1 nodes excited by both the input and the active F_2 category node, which is needed to overcome nonspecific inhibition from F_2.

Vigilance

Criterion for matching: some function representing degree of match between top-down and bottom-up patterns must be greater some positive constant *r*, called the *vigilance* of the network.

Low vigilance → broad categories
High vigilance → specific categories.

Network combining both modes



Consistency of this network with brain imaging data

Recall: De Martino et al. (2006) found that heuristics violators on their framing task had increased activation of OFC and ACC but not DLPFC.

Is this inconsistent with the network? My belief is that the task is not cognitively complex enough to engage the DLPFC.

Simpler model of knowledge/heuristics tradeoff

Work sponsored by grant from the Air Force via General Dynamics Information Technology, administered by Dr. Leonid Perlovsky.

Many decision making tasks evoke two or more competing rules, one of which is normatively superior to the others.

Example: choosing a larger probability versus a larger (absolute) frequency of either a gain or a loss (Denes-Raj & Epstein, 1994; Pacini & Epstein, 1999).

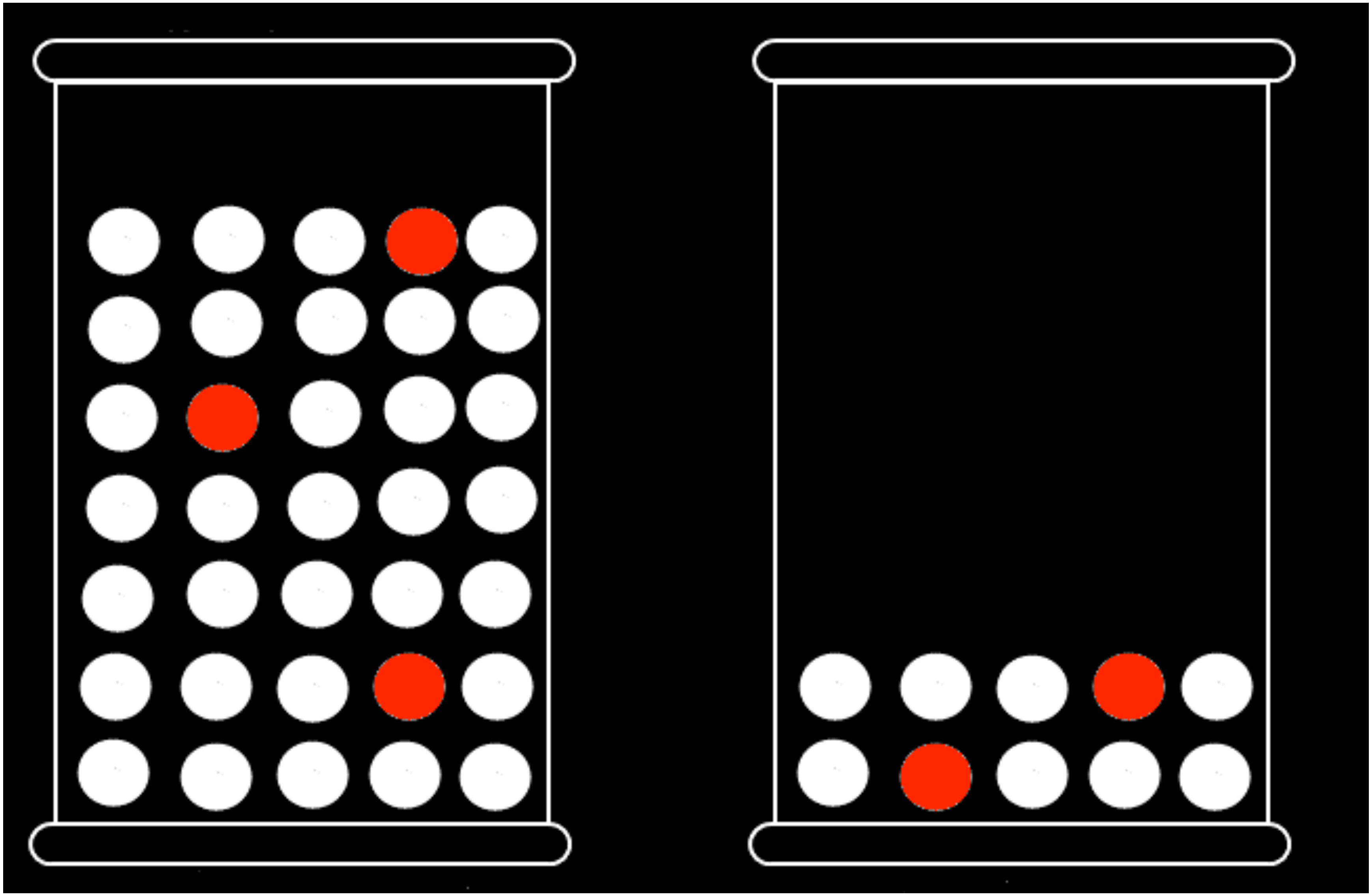
Denes-Raj and Epstein data

Win condition: shown two bowls containing red and white marbles (they used jellybeans), told they would win money if they randomly selected a red marble, told to choose which bowl gave them the best chance of winning money.

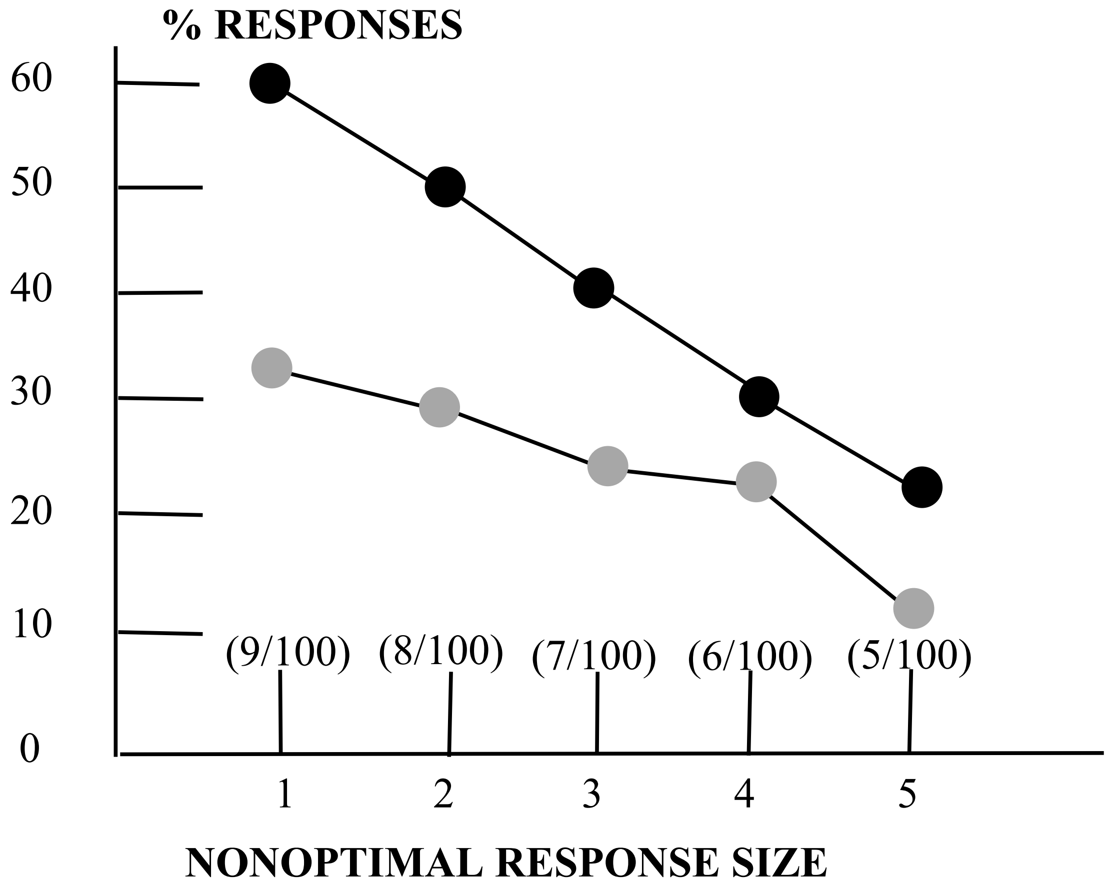
Bowl 1: 10 total marbles, 1 red. (higher prob.)

Bowl 2: 100 total marbles, k red, $1 < k < 10$. (higher freq.)

Loss condition: same two bowls, but told they would lose money if they selected a red marble.



Results (black = win, gray = lose)



Theoretical hypotheses about the "probability/frequency" task

Choosing the high probability correctly involves knowledge maximization.

Choosing the high frequency incorrectly involves a simplifying heuristic.

Three types of decision makers (DMs):

- DMs who choose, say, 8-in-100 over 1-in-10 and are not aware of any reason to do otherwise (heuristic-bound);
- DMs who choose 8-in-100 over 1-in-10 but verbalize a numerical reason for making the opposite choice (conflicted);
- DMs who correctly choose 1-in-10 over 8-in-100 (rational).

Current fMRI experiment

With Dr. Daniel Krawczyk (University of Texas Southwest in Dallas), we are testing this hypothesis:

Types (b) and (c) will show more ACC (conflict detector) activation than type (a).

Type (c) will show more DLPFC (working memory manipulator) activation than either type (a) or (b).

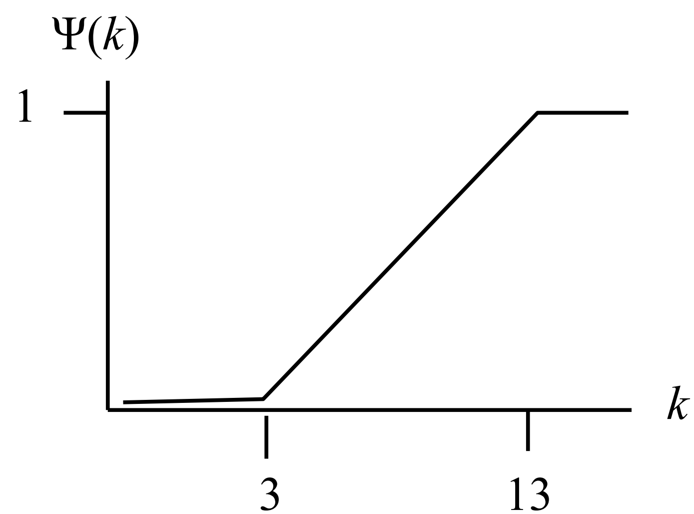
Simplified NN model (Levine & Perlovsky, IJCNN 2008, Hong Kong)

Network includes two parameters representing ACC and DLPFC function.

Decisions between two alternative gambles are based on either one of two rules, *heuristic rule* based on frequencies and *ratio rule* based on probabilities. "ACC" parameter, called α, determines likelihood of choosing ratio rule for a given pair of gambles. If ratio rule is chosen, "DLPFC" parameter, called δ, determines probability of optimal response.

Heuristic rule used if ACC parameter α small

If choice is between k out of 100 and 1 out of 10, criterion is $\alpha < \psi(k)$, where ψ corresponds to the fuzzy "much larger than 1":



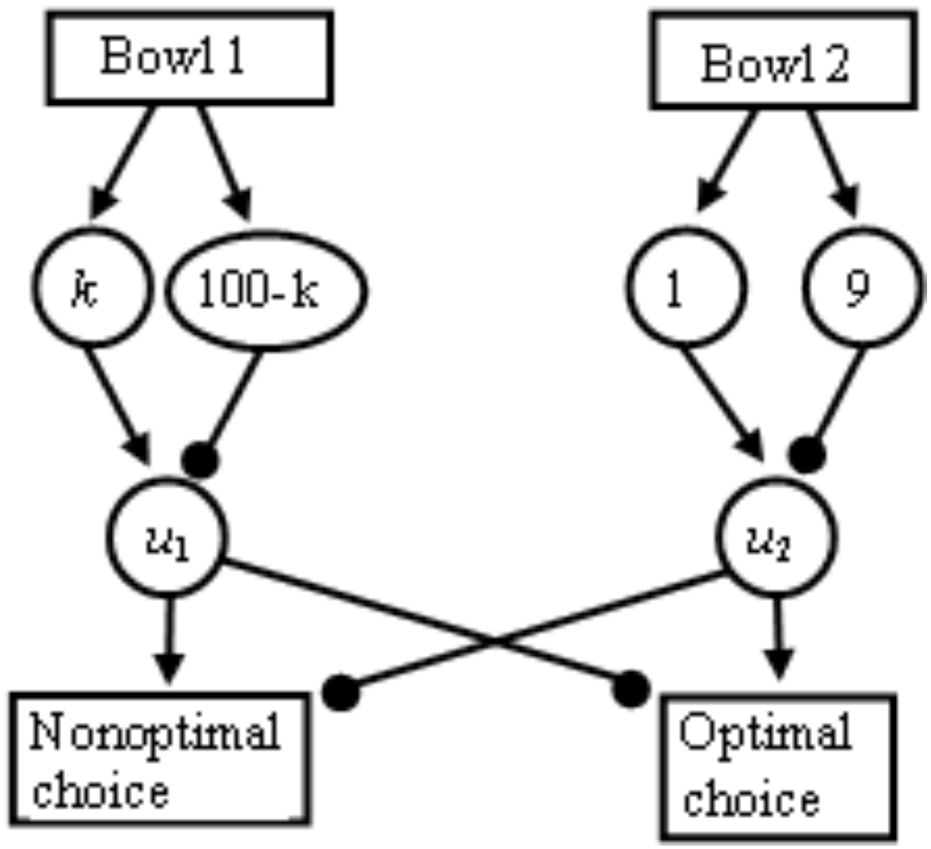
But rational rule does not guarantee higher probability is chosen

Numerosity detectors in parietal cortex are imprecise (Piazza et al., 2004).

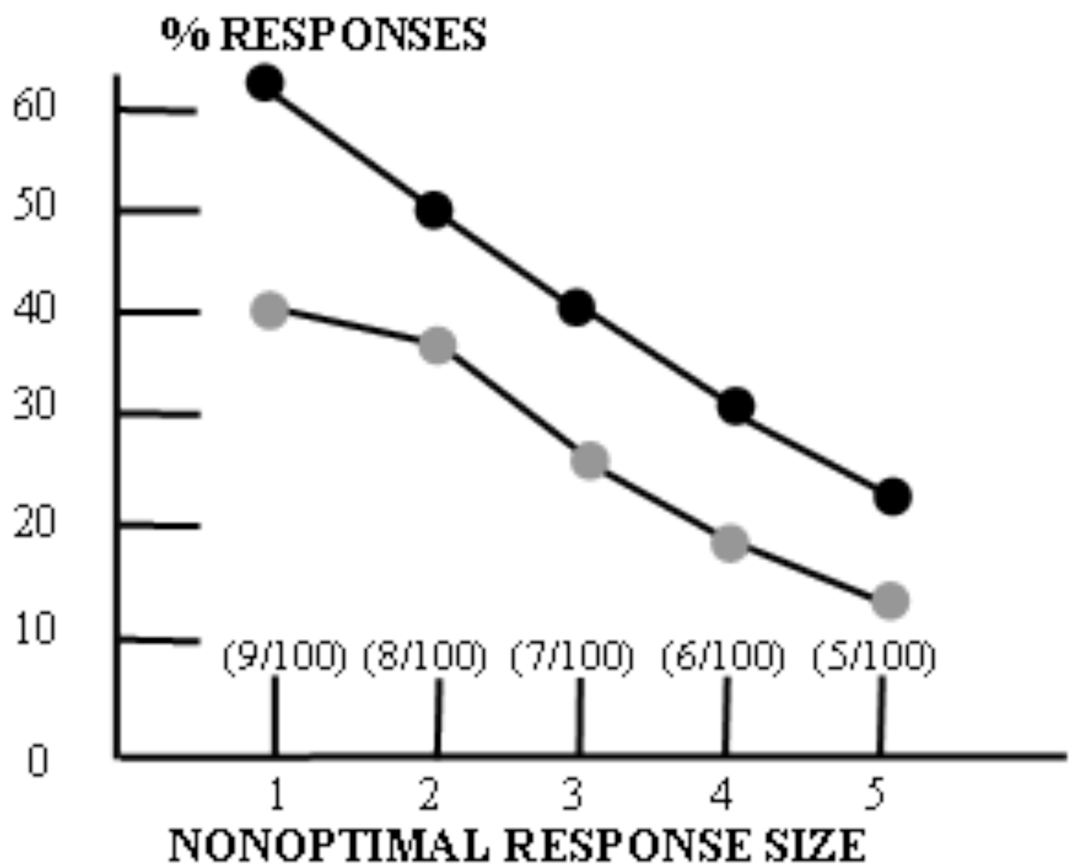
Assume numerators and denominators of both alternatives (*k*, 100, 1, and 10) each activate Gaussian distribution of numerosity detectors. Hence network multiplies each of those numbers by a normally distributed quantity with mean 1.

Assume DLPFC inputs to parietal cortex sharpen tuning of numerosity detectors, so standard deviation of each normal is $.1(1-\delta)$.

Ratios (probabilities) arise from shunting on-center off-surround network:



This simple network closely reproduced Denes-Raj & Epstein data:



Work in progress or planned

1. Application to other decision data (e.g., base rate neglect)
2. Include ART and detailed brain circuitry in neural network model
3. Integrate decision model with “angels, devils, and censors” model