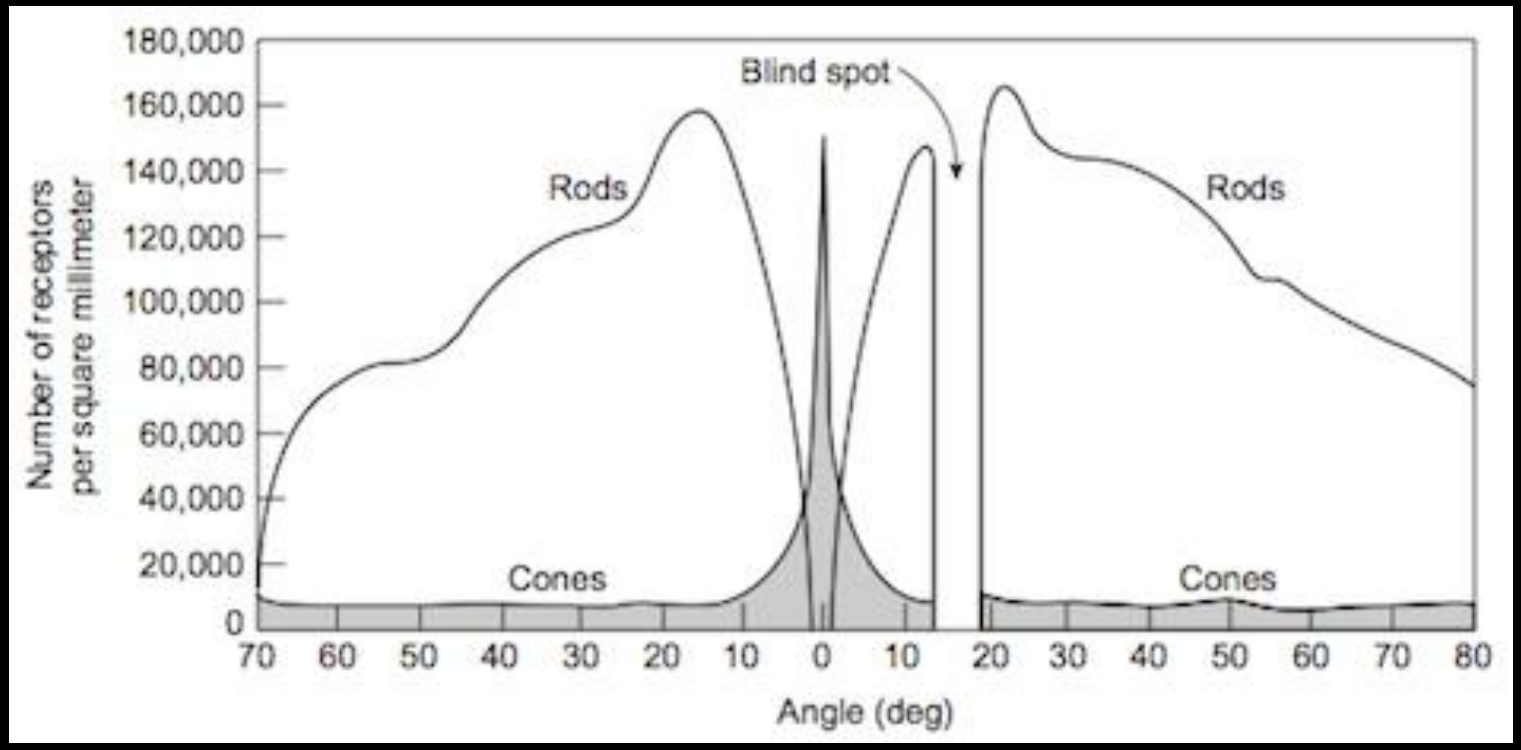
Prospection: The Wish and the Future

Peter Railton
University of Michigan

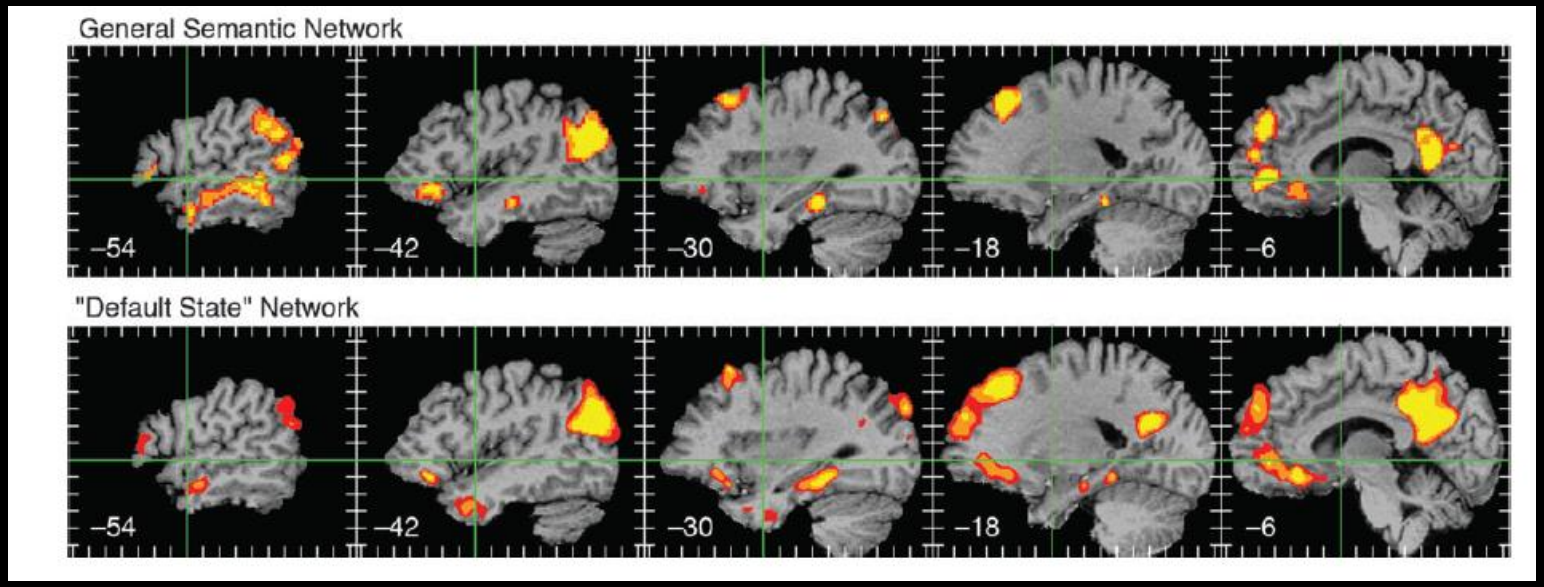
Vienna, 11 July 2016

Distribution of cone cells on the retina

(Johnson, 2010)



Episodic memory, prospective simulation, semantic memory, and evaluative judgment (Binder et al., 2008—150 studies)



The language of desire—why *wish*?

- Aristotle (384-322 B.C.E.)
- Organisms are structured *teleologically*
- Depending upon their nature, they have
  - … a distinctive *telos*
  - … and a form of flourishing that is based upon the realization of that *telos*
  - … the original Positive Psychology (indeed, his idea of the human *telos* as a rational, social animal gives rise to a theory of human well-being remarkably similar to PERMA).

At the core …

- … of any animal's distinctive way of relating to the world, or to other members of its kind, or to its own life, is *motivation (orexis*).
- There is a fundamental distinction in motivation:
  - *Epithumia* – "appetite": found in all animals
    - A kind of drive state related to a deficit condition (thirst, hunger, etc.).
    - A state that primes the individual for action, makes certain things salient to it (hunger makes food salient)
    - … but that lacks a *representational content*, so that we cannot *reason* with it.

The language of desire

- For contrast, consider the other chief form of motivation, *boulesis* – translated as *desire, rational desire,* or **wish** (hence my title).
  - Unlike mere *epithumia, boulesis* has a cognitive content.
  - It contains a *representation* of the desired object, action, person, or state.
- Thus, even if there are some desires that nearly everyone has, desire itself is a general capacity, like *belief*, and not specialized to any particular goal or behavior.

Desire thus depends upon imagination …

- … to give reality to absent possibilities, unrealized values, and abstract goals.
- That is its core job—translating what we learn by way of ideas and values into structured, motivated action that can make a difference in our life, and the lives of others.
- Humans are creatures who survive and flourish via complex social, psychological, and physical adaptations
  - This requires a capacity for complex, adaptive motivation as well as belief.
  - Though we often speak of desire as the *enemy* of prospection, but imaginative "mind wandering" and prospection have a key **ally** in desire.

"Don't make it a matter of course, but as a remarkable fact, that pictures and fictions give us pleasure, occupy our minds." **Ludwig Wittgenstein**

“Imagination rules the world.”

Napoleon

The content of desire

- In this respect, *what* one desires, and *how* one responds to one's desires, is one of the chief sources of:
  - Individual difference
  - Success or failure
  - Happiness or misery
  - Rationality or irrationality
- Desire is as different from mere appetite as the erotic is from a "sex drive".
- We can speak of *learning* what one really desires—a difficult process itself—and of learning what *to* want. It is one of our greatest capacities.

Desire?

- "He, on his side, had that faculty of beholding at a hint the face of his desire and the shape of his dream, without which the earth would know no lover and no adventurer."

Marlow on Jim

From Joseph Conrad, *Lord Jim*

Desire and reasoning

- I cannot talk you out of feeling thirsty by telling you that it would be bad for your health to drink.
  - By contrast, I *might* be able to talk you out of wanting your son to be a doctor—by appealing to your deeper beliefs and values, or his.
- Desire is no more *essentially* about the self than it is about others.
- At the same time, desire is what makes our beliefs and values or ideals *practical*—it gives them the power to move us, to change the world, not merely represent it.
- It is an engine of *prospective* guidance in life.

All action, even moral duty, presupposes desire for the object of that action.

The *faculty of desire* is the faculty to be, by means of one's representations, the cause of the objects of these representations. The faculty of a being to act in accordance with its representations is called *life*. [*Metaphysics of Morals* 6:211]

The origins of the Prospection Project

- A slogan, "Driven by the past, or drawn into the future?"
  - It was clear which side we were on,
  - … but the formulation was metaphorical
  - … and concealed some important flaws.

- I'd like to explore a bit the path that led us to this slogan,
  - … and to its clarification
  - … and eventual crystallization in the idea of *prospection*, and of *navigating* into the future.

It's notable that …

- … both halves of the original slogan are about **motivation**.
  - Motivation is pivotal both in the story about the human mind that we are trying to overcome,
  - … *and* in the story we are trying to develop.

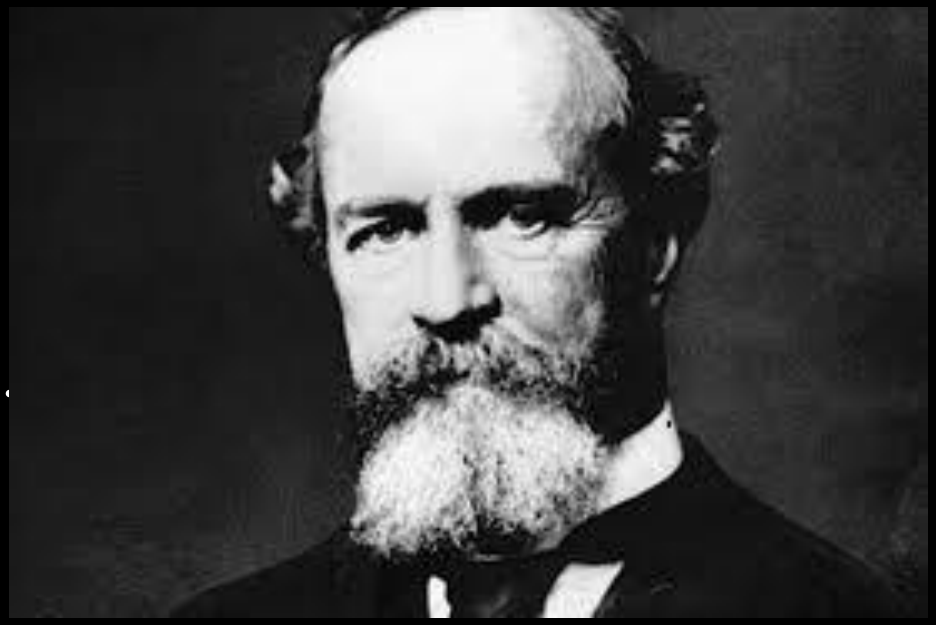
Here's a way of seeing why

- On the classical picture, action is the product of two basic kinds of states:
  - **belief + desire → action**
- As Aristotle noted, belief on its own cannot produce action—it must engage with motivation to produce actual movement.

- And think in **evolutionary** terms:
  - We cannot improve our chances for survival and reproduction simply by forming accurate beliefs—these must make a difference to how we act.

William James drew the relevant conclusion:

- "**My thinking is first and last and always for my doing.**"
- (*Principles of Psychology*, I: 333).



James could have added …

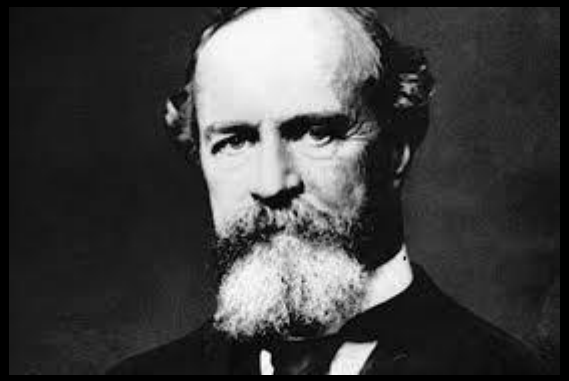
- … that my doing always extends *forward* in time, not backward.
- So thinking that is for doing must be thinking that is *prospective.*

- But thinking that is forward-looking would be of no avail in guiding action if it could not engage motivation.

Enter "drive theory"

- Yet motivation, as conceived by James and others in the "scientific" tradition of psychology that followed him, is grounded in innate and conditioned drives, which are "bound" to present stimuli.

James on motivation



- "**Not one man in a billion, when taking his dinner, ever thinks of utility**. He eats because the food tastes good and makes him want more. If you ask him *why* he should want to eat more of what tastes like that, instead of revering you as a philosopher he will probably laugh at you for a fool. The connection between the savory sensation and the act it awakens is for him absolute and *selbsverständlich*. *(Principles of Psychology,* II: 386)

Karl Lashley

- "Our current psychological language is a weird composite of teleological and mechanistic terms. …
- "An introspective description of my purpose would not reveal an influence of the future on the present, nor does the behaviorist account" (1923, 349).
- "**The two systems, mechanistic explanation and finalistic valuation, stand out as incompatible points of view**, scientific versus humanistic" (1923, 346).

The shared drive construct

- An internal "deficit" state yields discomfort and dissatisfaction, driving pursuit
  - … and making external stimuli "hot"
  - … triggering actions that in the past were associated with reducing that deficit
  - … and the reduction of discomfort serves as the reward which entrenches the stimulus-response connection as an "automatic" or habit-like response.
  - It is this entrenching process that constitutes *learning*—not the learning of value, but the engraining of habit.

And the “scientific psychologists” were not alone …

- … the drive construct figured essentially in Freudian psychology, Lorenzian psychology, and contemporary work on automaticity (Ferguson & Bargh, 200).

If this really *were* the story of motivation …

- … then future-oriented thinking would be *hobbled* by motivation.
  - Even with elaborate mental representations of the future, if all motivation came from such drive processes, then we would still be "driven by the past".

“Drawn into the future”

- Would the story be significantly changed if we replaced “driven by the past” with “drawn into the future”?
  - In many respects, yes, but the story would still be of an automatic mechanism,
    - … like the iron filings drawn to a magnet or the moth drawn to a flame.

What is *intelligent* belief?

- Make the comparison with belief:
  - Suppose there were certain beliefs we were inevitably "driven" to form …
  - … or that we were inevitably "drawn" to.
- These would hobble our learning, since they would be limits on our learning that do not arise from considerations of justification or evidence …
  - … when what we want in belief is to be free to respond appropriately to the evidence, whatever it might be.

Well, what would be *intelligent* motivation?

- Just as intelligent belief would enable us to respond to evidence to develop an accurate map of the world and its possibities,
  - … and use this to generate *expectations* that could *guide* action accordingly,
  - … intelligent motivation would enable us to respond to evidence to develop an accurate *evaluation* of the world and its possibilities,
    - … and use this to generate expectations that could guide our action accordingly.

Combining intelligence in belief and desire

- With intelligent belief and intelligent motivation, we could *navigate into the future,*
  - … steering by the light of the values we have learned
  - … and setting ourselves up to learn yet more from the outcomes of acting on these values.

Could desire be such an intelligent faculty?

- Somehow we must break away from hydraulic or magnetic imagery, to see how valuation could *regulate* motivation,
  - … creating the possibility of genuinely novel, reason-based action
  - … just as we can form genuinely novel, reason-based beliefs.
- Desire, then, must be able to mobilize and guide activity on behalf of values not yet realized, but represented imaginatively.
- What would such desire be like?

- **Paris!**

Paris!

Paris …

Ecole M
7
LE BON MARCHÉ
RIVE GAUCHE

j'adore
Dior

PRADA

Heineken
SHINE
THE ORIGINAL QUALITY
BEER
AMSTERDAM 1873
MARKS
PARIS 1875
PARIS 1889

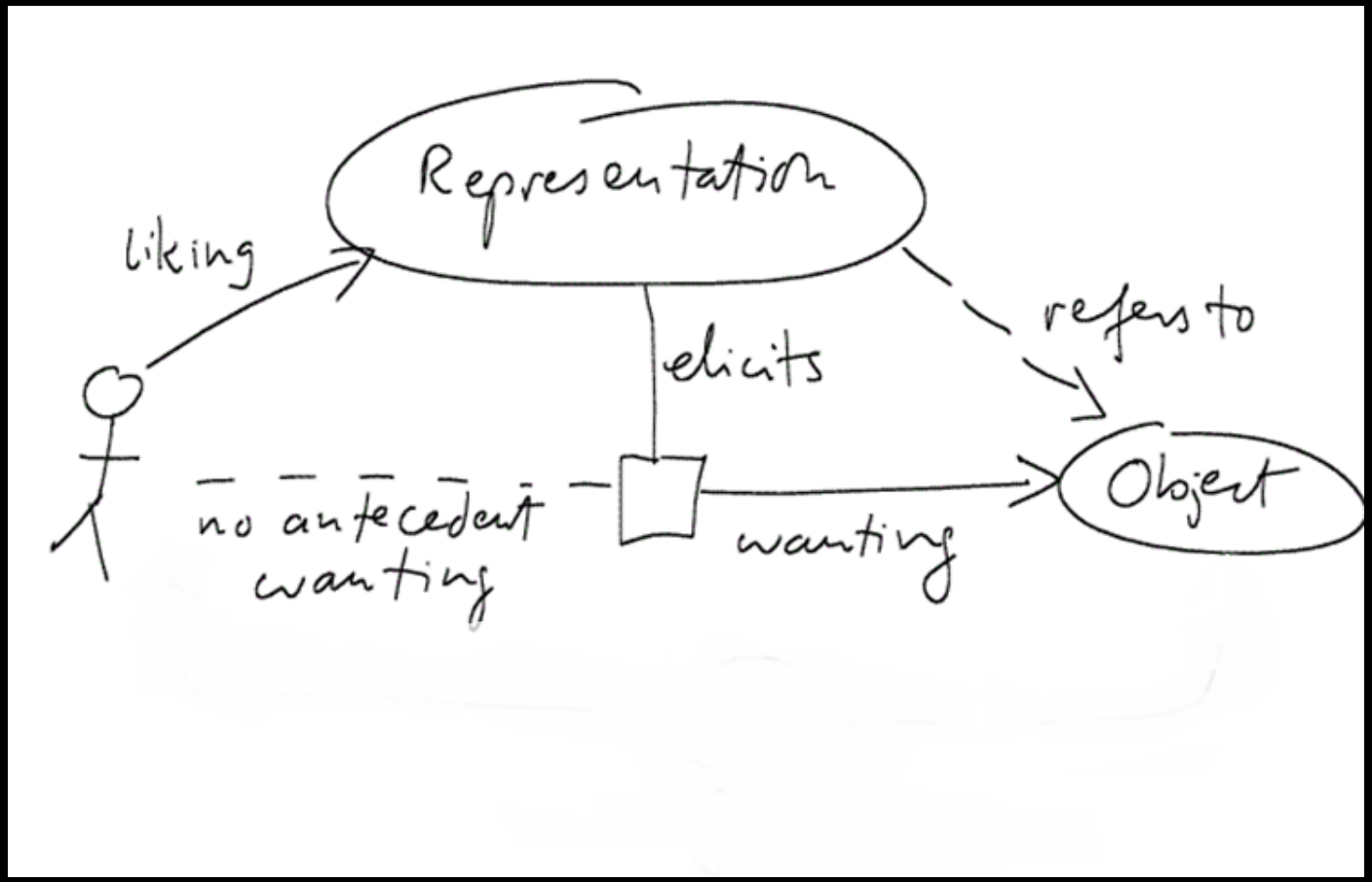
38

BREITLING
1884

BREITLING
BREITLING for BENTLEY

41

Desire – First Pass



Liking – a species of positive affective interest, attraction to a prospect; hedonic evaluation

Wanting – a species of "incentive salience" that galvanizes attention, motivation, and action; not inherently hedonic (Berridge, 2004)

Indo-European root

- desire

dē- + *sīder*

apart + star

Star of wonder,

Star of light,

Star of royal beauty bright.

Westward-leading, still proceeding

Following thy perfect light.

La planète désir

www.galerieslafayette.fr

Why don't the advertisers always win?

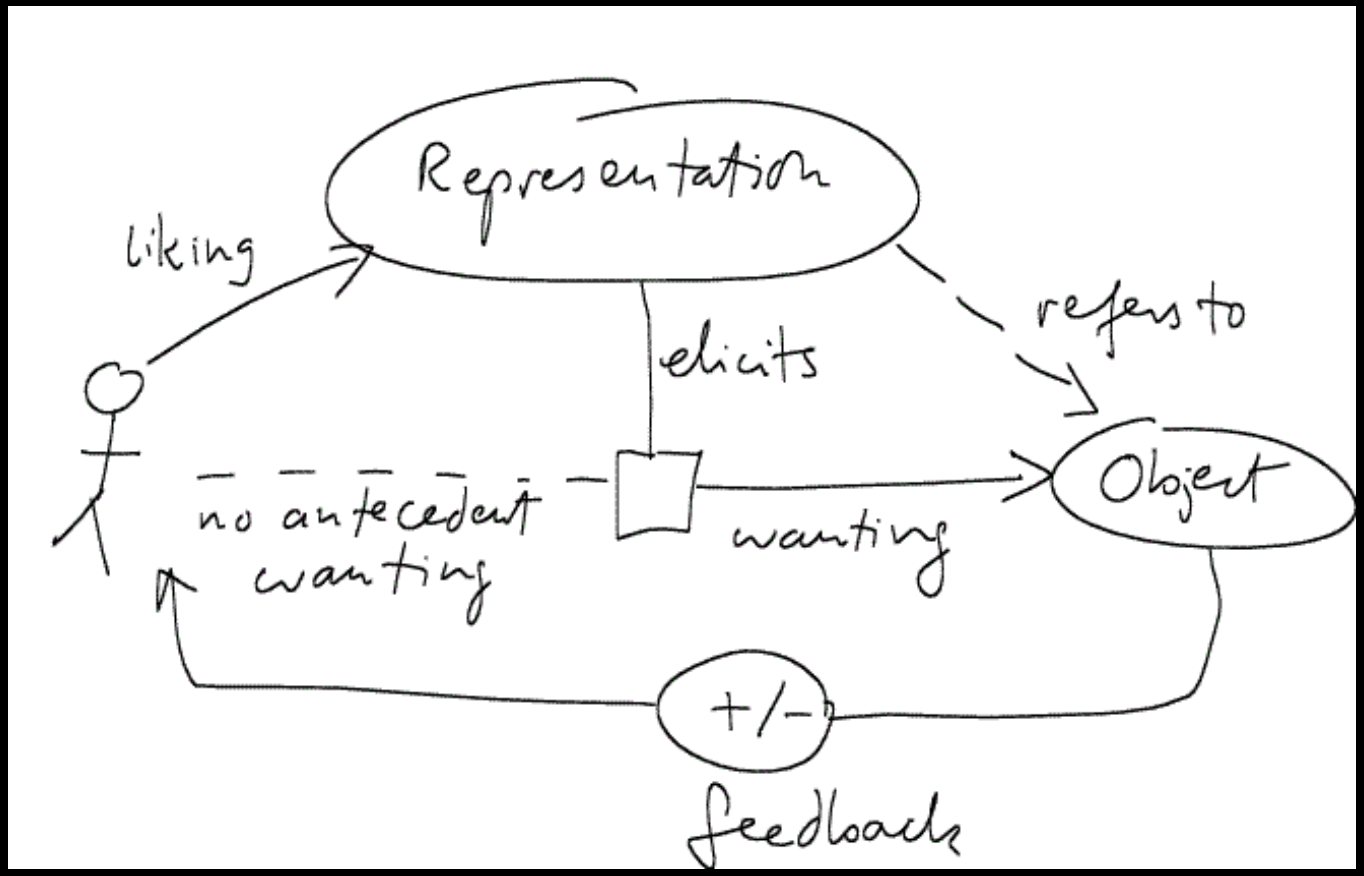
- Oscar Wilde: "There are only two tragedies in life:
  - one is not getting what you want …

Why don't the advertisers always win?

- Oscar Wilde: "There are only two tragedies in life:
  - one is not getting what you want …
  - the other is getting it."

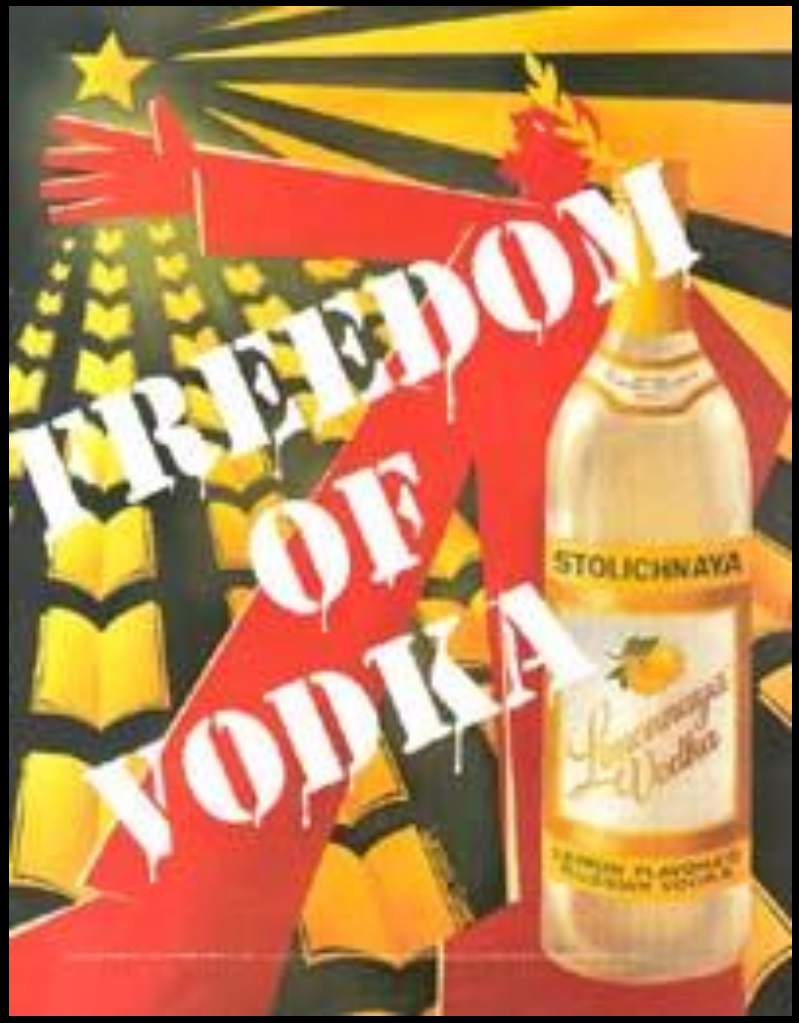
RENCONTRE
#03
RICARD/
GLACE
RICARD
UN RICARD
DES RENCONTRES
ricard.com
L'ABUS D'ALCOOL EST DANGEREUX POUR LA SANTÉ. À CONSOMMER AVEC MODÉRATION.

Desire: Second Pass



Once as tragedy …

Once as farce …



What about the brain?

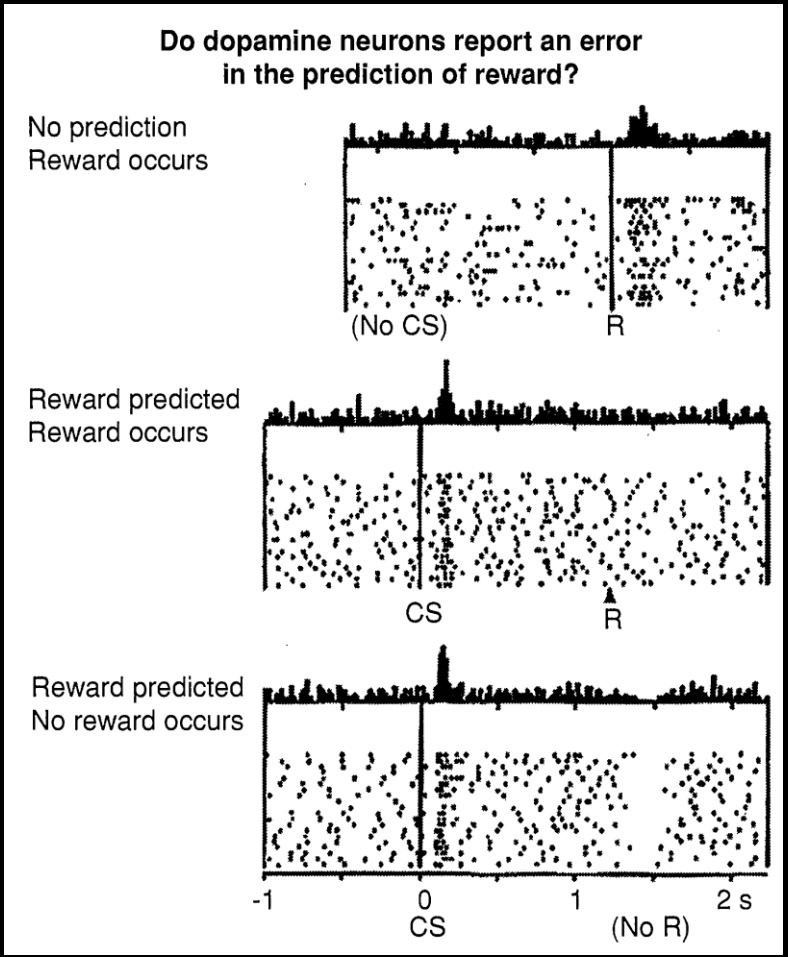
The denouement of drive theory

- Eating behavior persisted when the deficit state was artificially removed (Miller & Kessen, 1952; Turner *et al.*, 1975).
- Eating electrodes were hedonically positive, not negative.
- **The drive construct "is so intuitive that it was thought to be self-evident for decades** … [and] some behavioral neuroscientists today still talk and write as though they believe it. **All the more pity, perhaps, that the idea turns out not to be true**." (Berridge, 2004, p. 191).

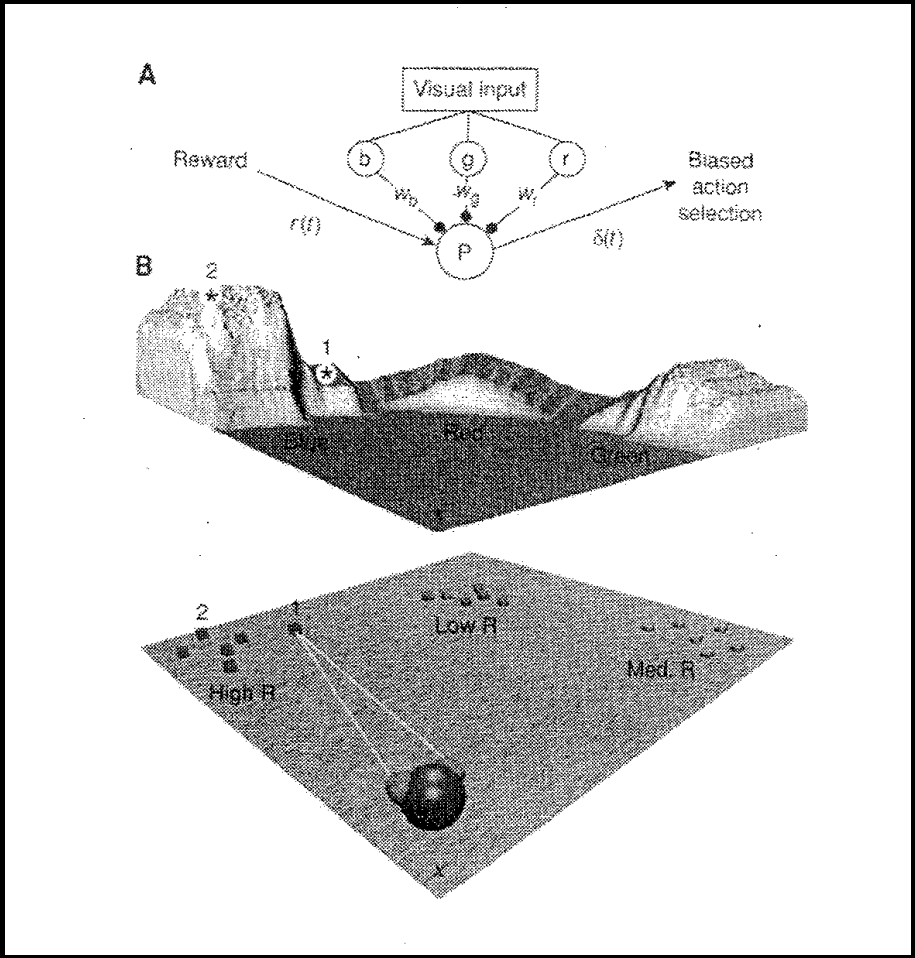
- What is the alternative? The dog eats because it *likes* eating—its affective system assigns a positive value to eating. That is, it *does* consult utility when deciding about whethere to teatit

The neutral substrate of reward-based learning

(Schultz *et al.*, 1997)



Affectively-coded cognitive mapping—foraging for value (Schultz *et al.*, 1997)

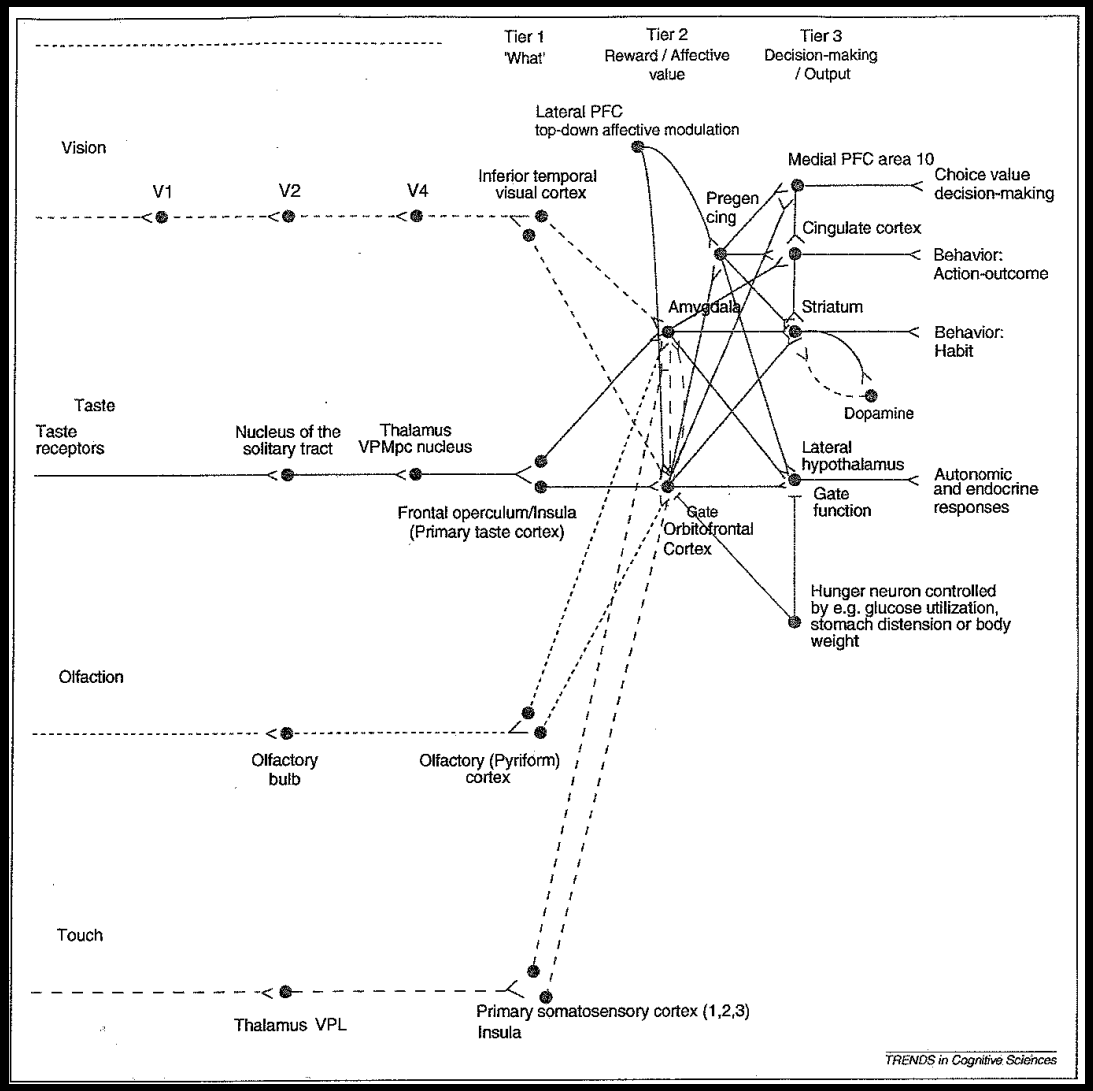


It appears that …

- … just as we are "built for" perception and cognition,
- … we are "built for" evaluation—indeed, evaluation has a *primary* role in both perception and cognition.
  - In perception, we must separate out what needs attending to, what is urgent, what is uncertain, etc.
  - In cognition, we must *interpret* information, weigh its degree of confidence, use it to guide inference and potentiate action, etc.
  - All these are part of the "job description" of the affective/evaluative system: to coordinate our responses across the board in light of relevant values.

Neural processing for valuation and decision

(Grabenhorst & Rolls, 2011)

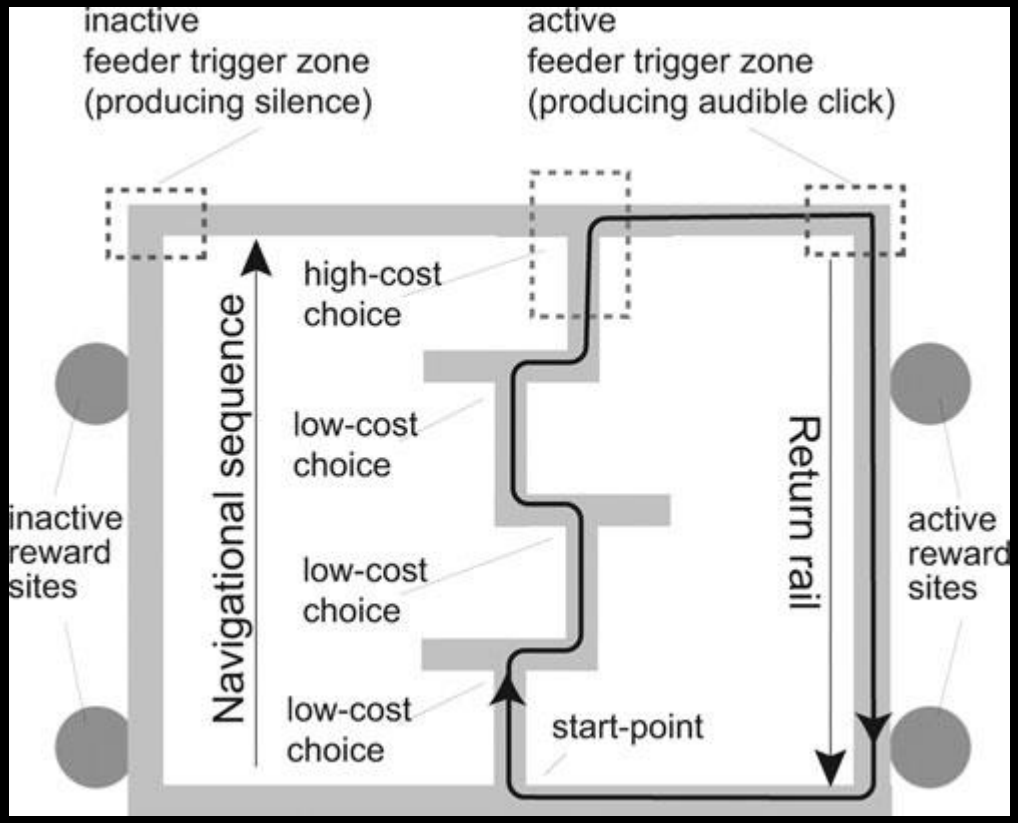


But does utility- and probability-based reasoning really guide action?

But some unanticipated results—Karl Lashley (1929) and novel behavior without enforcement

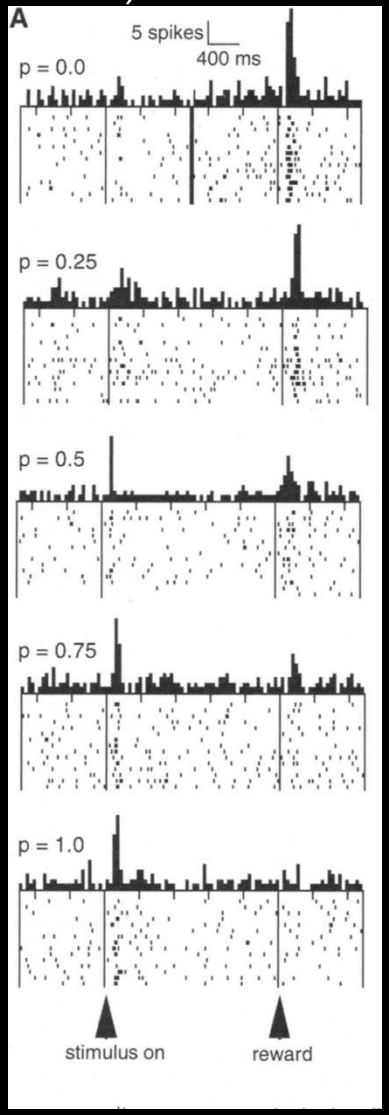
Further unanticipated results: Edward Tolman (1948)—cognitive maps

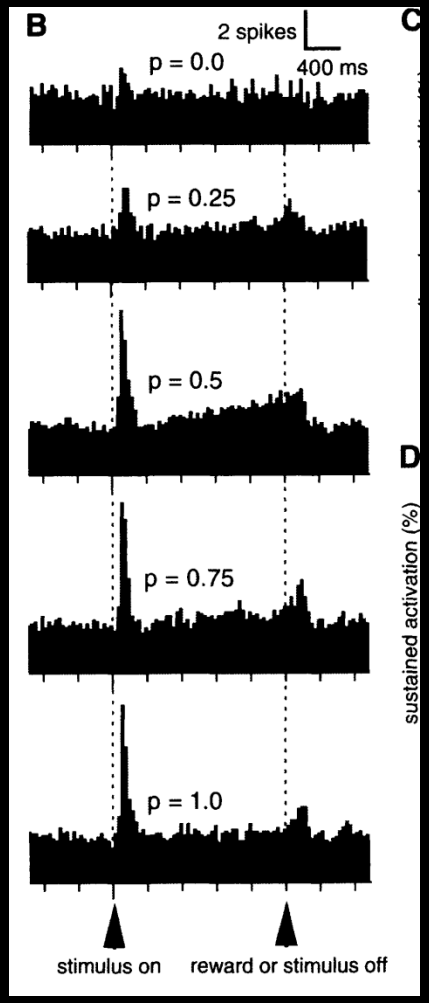
Decision


inactive
feeder trigger zone
(producing silence)
active
feeder trigger zone
(producing audible click)
high-cost
choice
low-cost
choice
low-cost
choice
low-cost
choice
Navigational sequence
Return rail
start-point
inactive
reward
sites
active
reward
sites

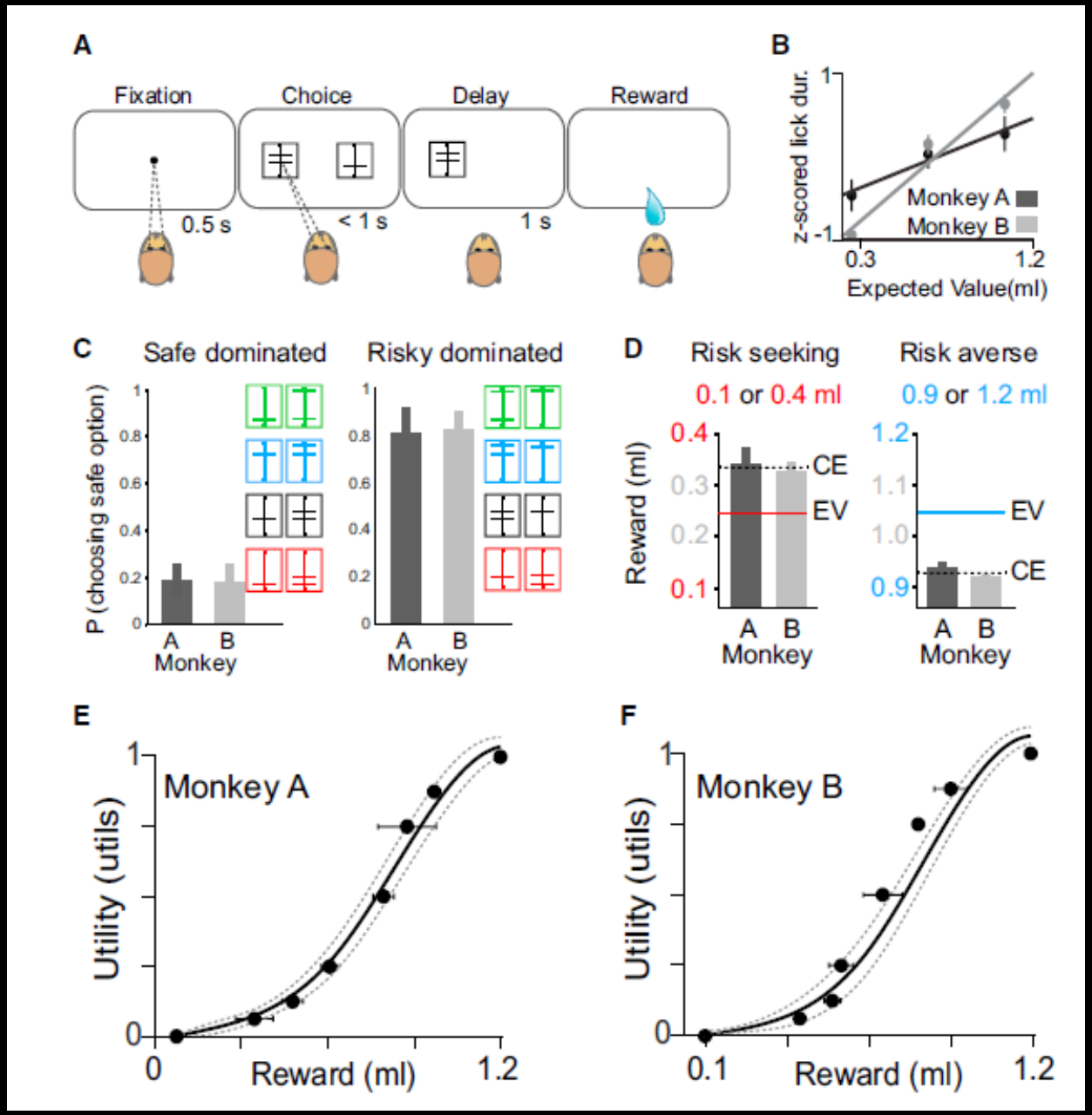
Risk and anticipation

(Fiorillo *et al.*, 2003)

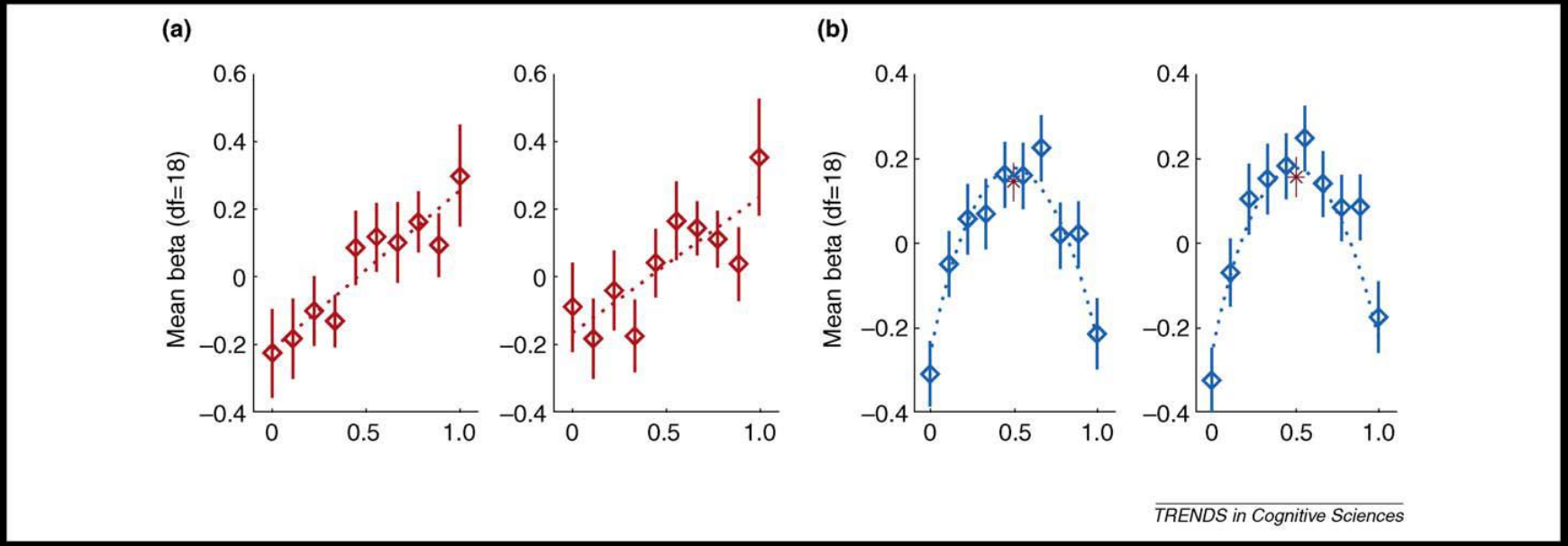




65

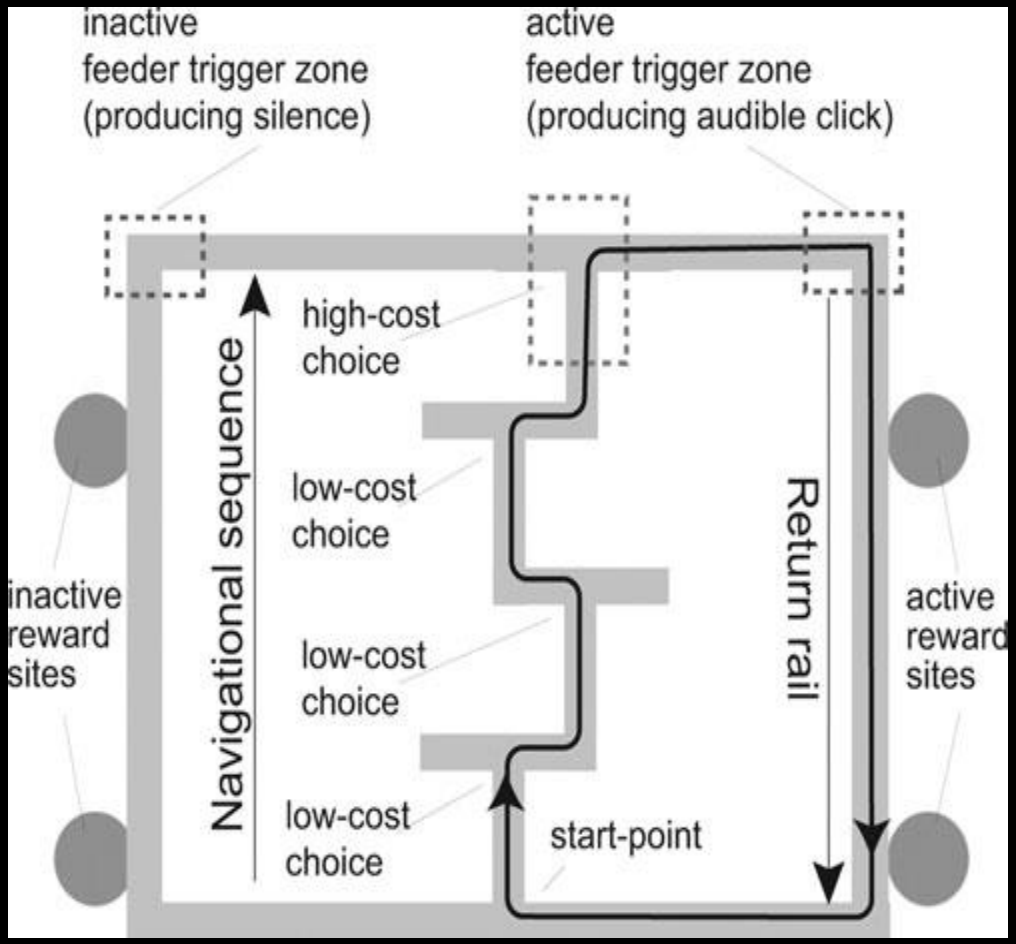

A
Fixation
Choice
Delay
Reward
0.5 s
< 1 s
1 s
B
z-scored lick dur.
1
-1
Monkey A
Monkey B
0.3
1.2
Expected Value(ml)
C
Safe dominated
Risky dominated
P (choosing safe option)
1
0.8
0.6
0.4
0.2
0
A
B
Monkey
D
Risk seeking
Risk averse
0.1 or 0.4 ml
0.9 or 1.2 ml
Reward (ml)
0.4
0.3
0.2
0.1
CE
EV
1.2
1.1
1.0
0.9
EV
CE
A
B
Monkey
E
Monkey A
Utility (utils)
1
0
0
Reward (ml)
1.2
F
Monkey B
Utility (utils)
1
0
0.1
Reward (ml)
1.2

Representation of value vs. risk—activation in the human ventral striatum (Quartz, 2009)



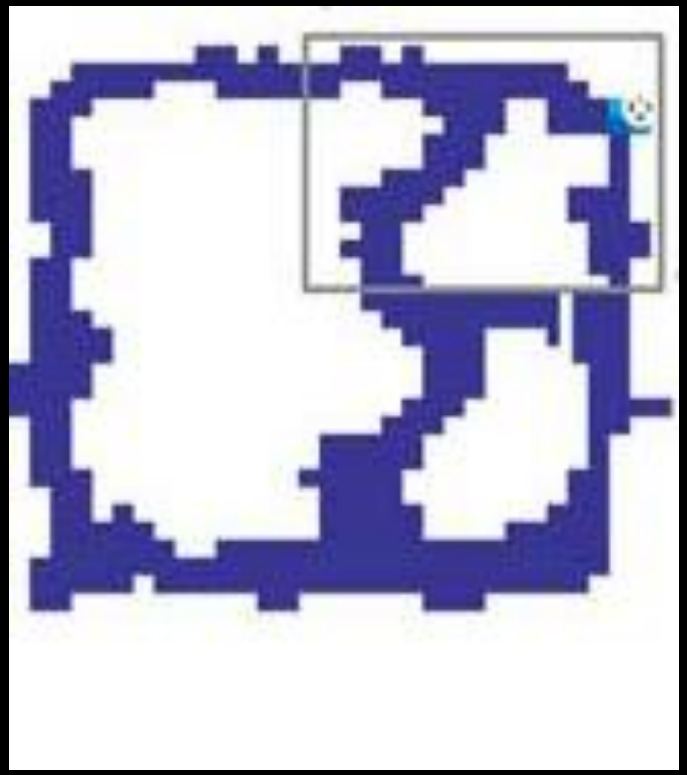
Prospection in action

T-maze with return rails



Cognitive map—hippocampal place cells

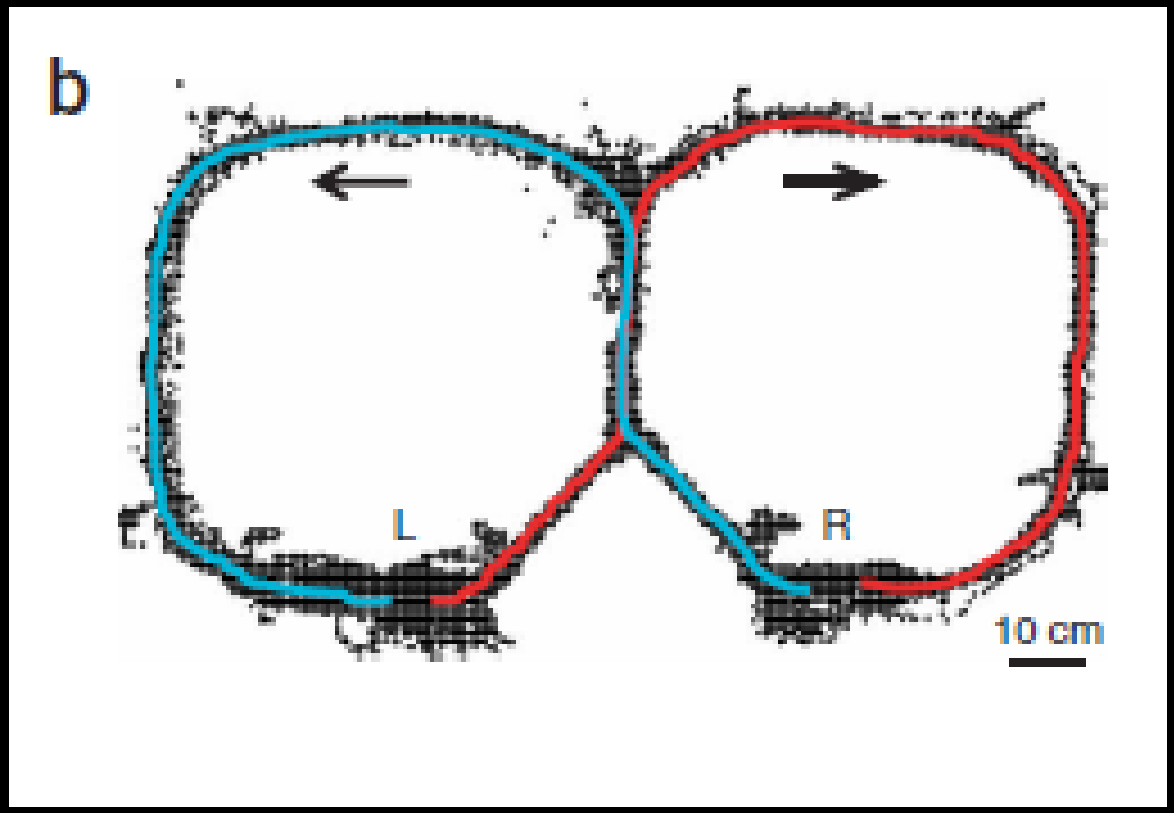
(Johnson & Redish, *J Neurosci* 2007)



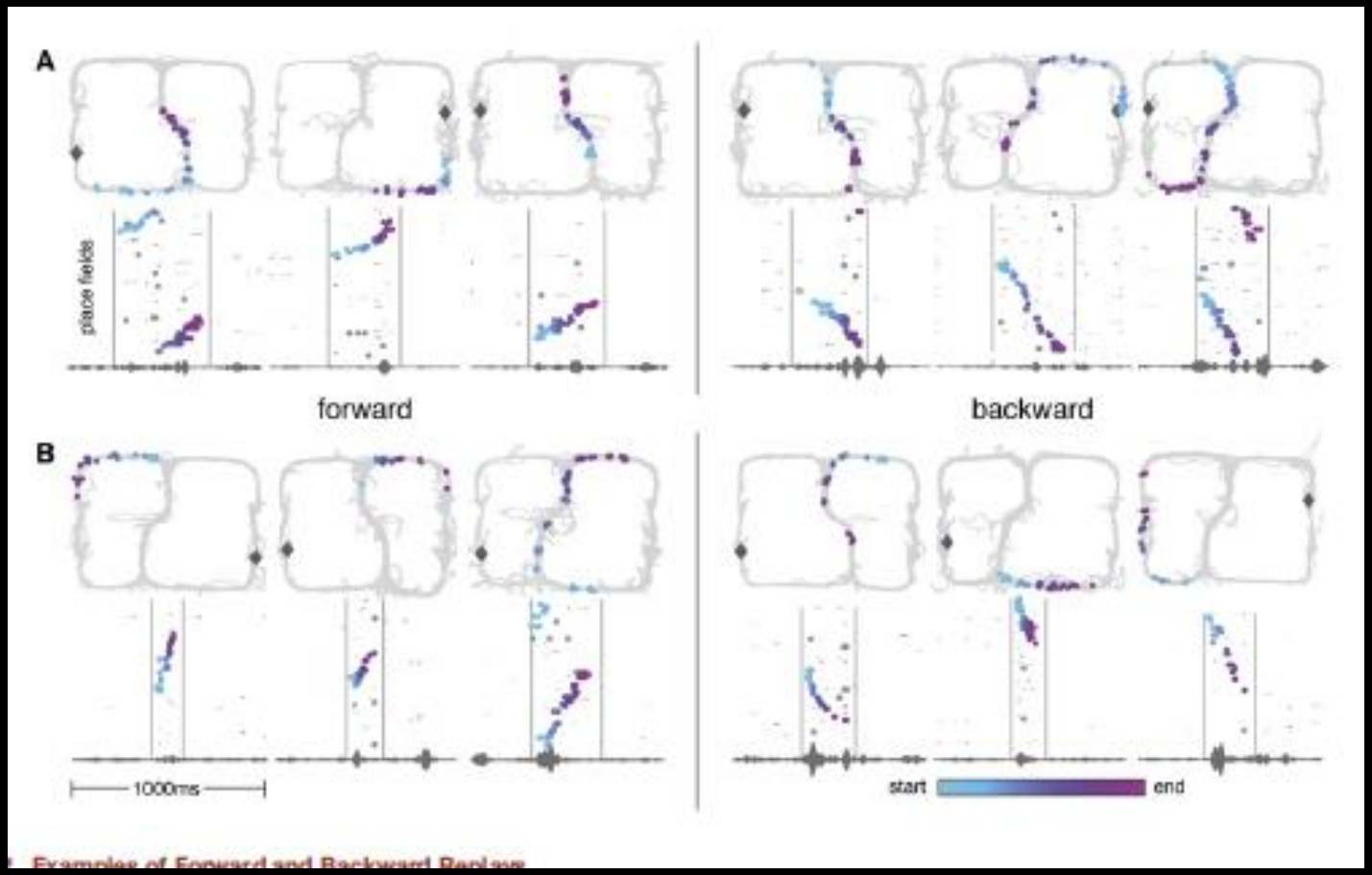
In model-based control …

- … learning can take place by the discrepancy between expectation and observed outcome—no external "reinforcement" is needed for updating to occur.
  - As in the "deep learning" artificial intelligence systems discussed earlier, a simple algorithm can be combined with simulations *generated internally* to improve the quality and informativeness of representations.
  - As in the deep learning systems, too, this sort of *unsupervised* learning can yield novel patterns or features, which then make supervised learning more effective.

Co-ordinated replay of mapping in the rat hippocampus during sleep (Ji & Wilson, 2007)

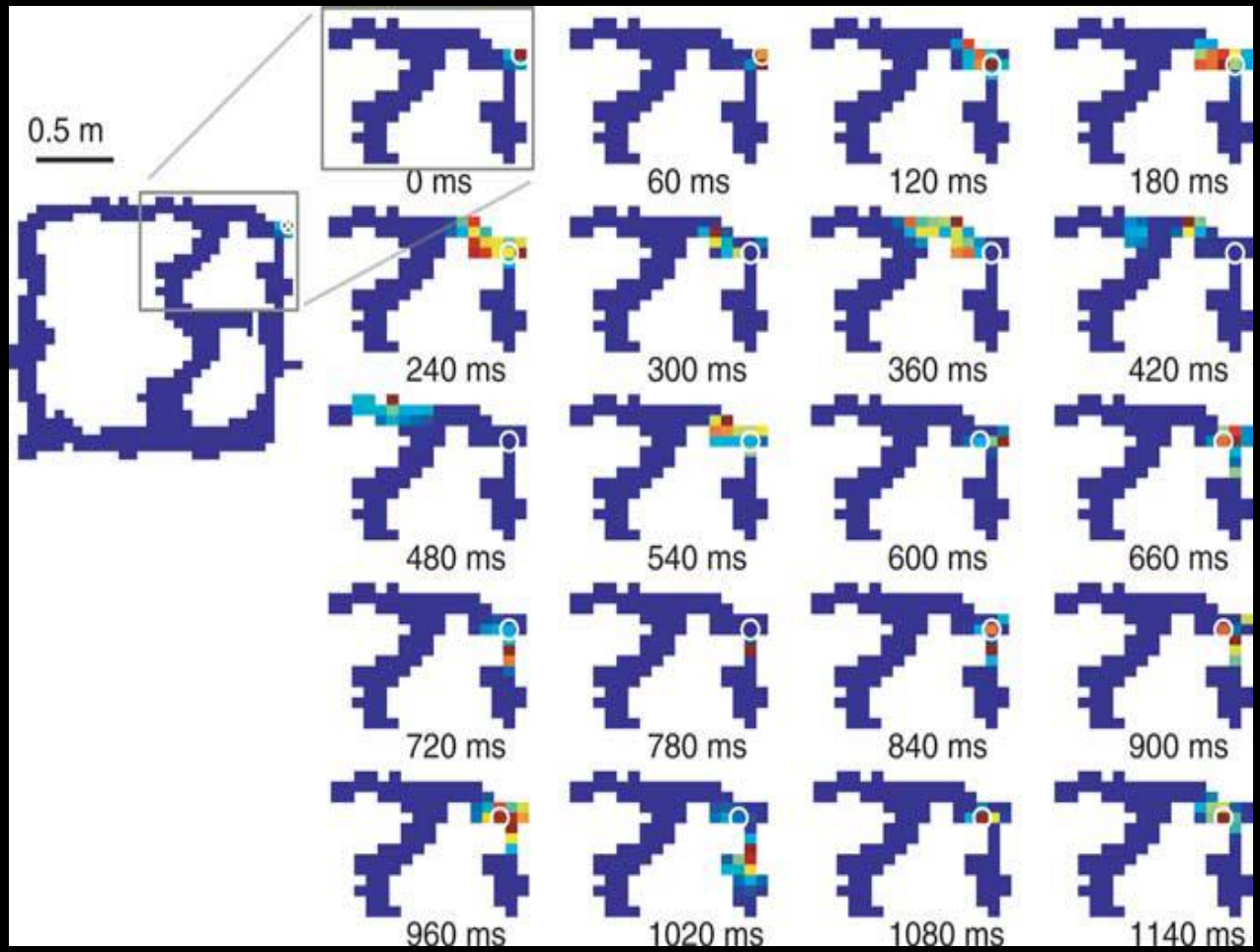


Hippocampal construction of novel paths in sleep (Gupta *et al.*, 2010)



A rat following an evaluative representation

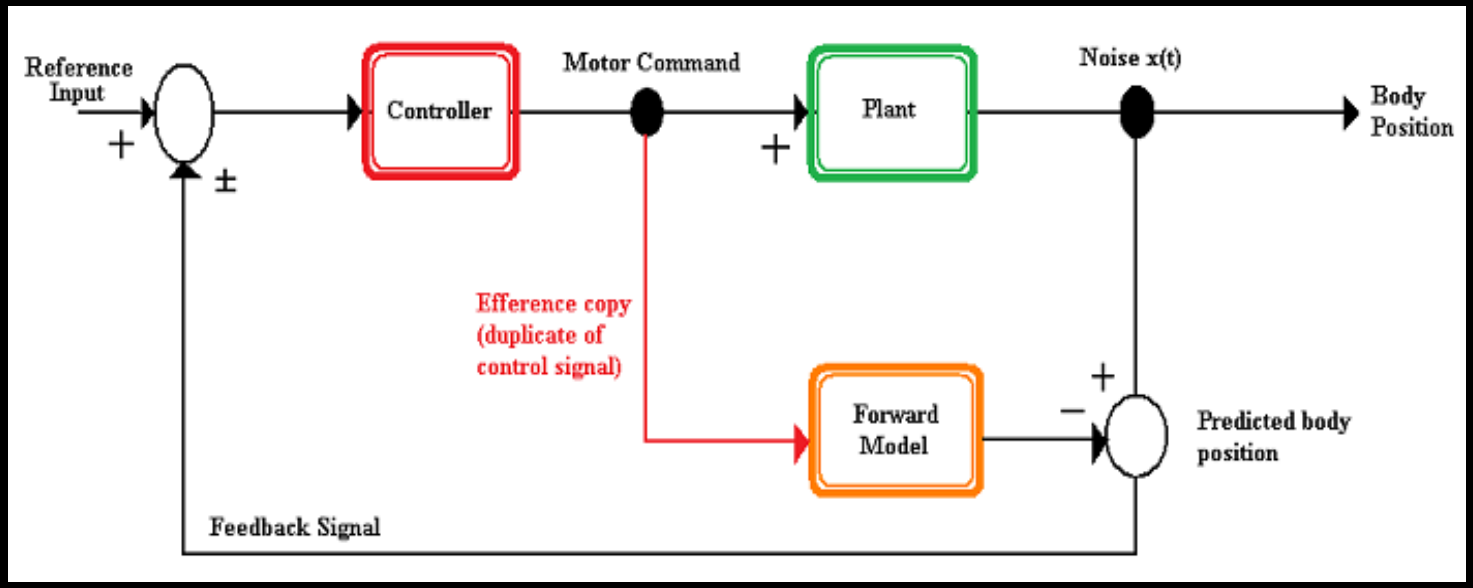
(Johnson & Redish, *J Neurosci* 2007)



74

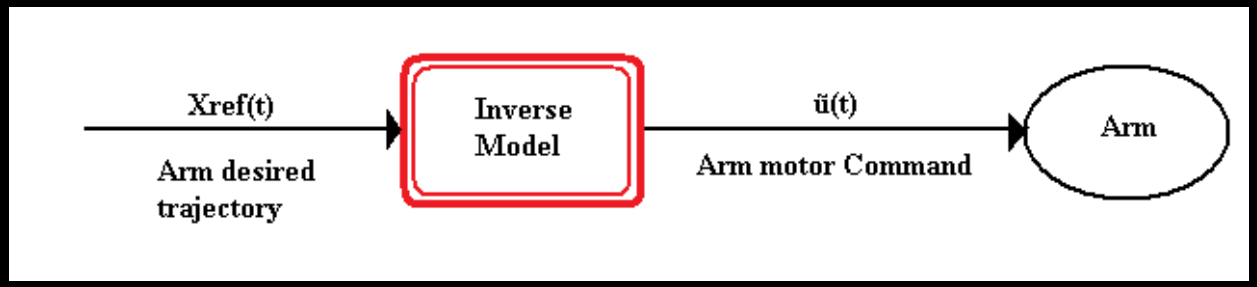
Elements of model-based control

- Representations of desired or target states of the system
- Representations of the actual or possible states of the system and its environment
- Representations of the actual or possible motor commands for the system in a given state
- A capacity to generate expectations for the outcome of a motor command in a given state and environment ("forward model")

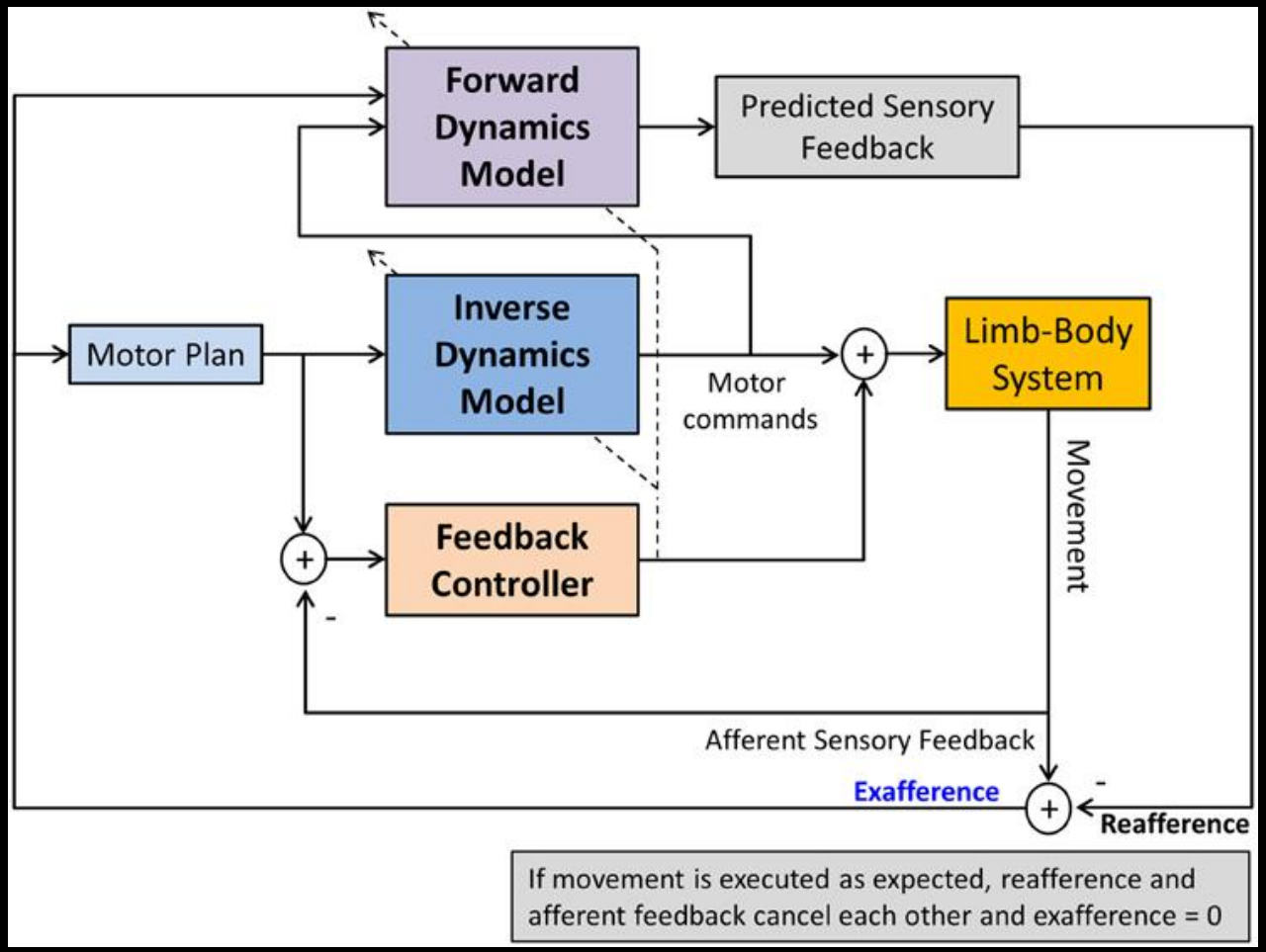


Elements of model-based control

- A capacity to receive input from internal and external sensors indicating the state resulting from a command
- A capacity to compare this result to expectations
- A capacity to use any discrepancy to modulate the next motor command ("inverse model")



Putting it together



A corollary of the "Good Regulator Theorem"

- Conant & Ashby (1970) also conclude:
  - "the living brain, so far as it is to be successful and efficient as a regulator for survival, must proceed, in learning, by the formation of a model (or models) of its environment."

An increasing body of evidence …

- … supports the idea that human skilled movement is in fact the result of model-based control of this kind.
  - Studies of elite athletes found evidence that both forward and inverse models play an essential role—in learning as well as movement (Yarrow *et al.*, 2009)
- Rather than program detailed joint angles, etc., robotic systems were given a physical model and allowed to use simulation of possible movements to *learn* optimal movements relative to constraints and target outcomes.
  - Model-based optimal motor control theory predicts fine-grained features of human motor control (Liu & Todorov, 2007; Scott, 2009).

Even "habit" …

- … isn't what it used to be. Except for cases of "overtraining", behavior learned through reinforcement learning
  - … does not consist in the ingraining of "fixed motor programs"
  - … but remains flexible, and responsive to changes in information and reward
  - … up to the point of "overtraining" (Smith & Graybiel, 2013, 2014).

Converging lines of evidence—we are designed to build, and use, prospective, evaluative models

- Neuroscience
- Behavioral science
- Systems theory
- Motor control theory
- Learning theory

- Action will be more successful and more efficient if it accurately *anticipates* in the here and now what will happen in the then and there—investigating possibilities mentally beofre we need to investigate them physically.

Anticipation—a first clue as to the nature of desire

- Efficiency considerations alone, not to mention competitive pressures, mean that **natural selection will have relentlessly favored organisms that do better at anticipation**.
- Anticipating not only what will happen, but what it will be like, and whether this is a good or bad thing for the organism, to be sought or avoided.
- At the most basic level, *approach* vs. *avoidance motivation*.
- We should expect that the architecture of the mind and brain is built around the problem of successful anticipation, and the use of such anticipatory representations to motivate and guide action.

Desire as evaluative anticipation

- What to want and when to want it seems as much a locus for selection as what to think and when to think it.
- Indeed, the two cannot be pulled apart.
- Anticipation in belief and desire alike is a key enabler of learning.