Can a “new AI” become a formal science?

Gregor Schöner
RUB Germany

Why become a formal science?

- To what end?
- Is this the bottleneck problem of a new AI?
- Is this a solution in search of a problem?
- How is the problem of the new AI embedding in empirical science?

Disclaimer

- “there is nothing more practical than a good theory”
- => no doubt that elements of the “new AI” are useful when problems close to the formalism are solved
- e.g., in automatic vision

Perspective

- => I'll argue from the perspective of one who is interested in
  - human cognition,
  - in the neuronal foundations of cognition,
  - and thus in the constraints that arise from the embodiment and situatedness of cognitive systems
- because organisms/humans exist in the real world and do display behavior that at least remotely approximates what "new AI" addresses, these are relevant candidates for probing the formal approach

Embodied cognition in humans

- Cognition is linked to the sensory and motor surfaces, constrained by the structure of the nervous system
- Cognition happens while embodied systems are immersed in structured environments and placed in behavioral context
- Cognition happens on a background of behavioral history and experience

playing soccer

- see and recognize the ball and the other players
- select target, track it as well as the other players, all the while controlling gaze
- use working memory when players are out of view to predict where you need to look to update
- control own motion, initiate and control kick
- any time open to update
- get better at it
- background knowledge: goal of game, rules, how hard is the ball, how fast are players

driving

- perceive and estimate ego-motion
- detect and segment the road, segment and categorize cars, estimate car kinematic state
- use working memory to know where to look to update scene representation
- make passing decisions, control car
- adapt to car, to roads, to sight conditions
- get better at driving and seeing
- background knowledge: know typical behavior of other drivers, know geometry of roads

repairing a toaster

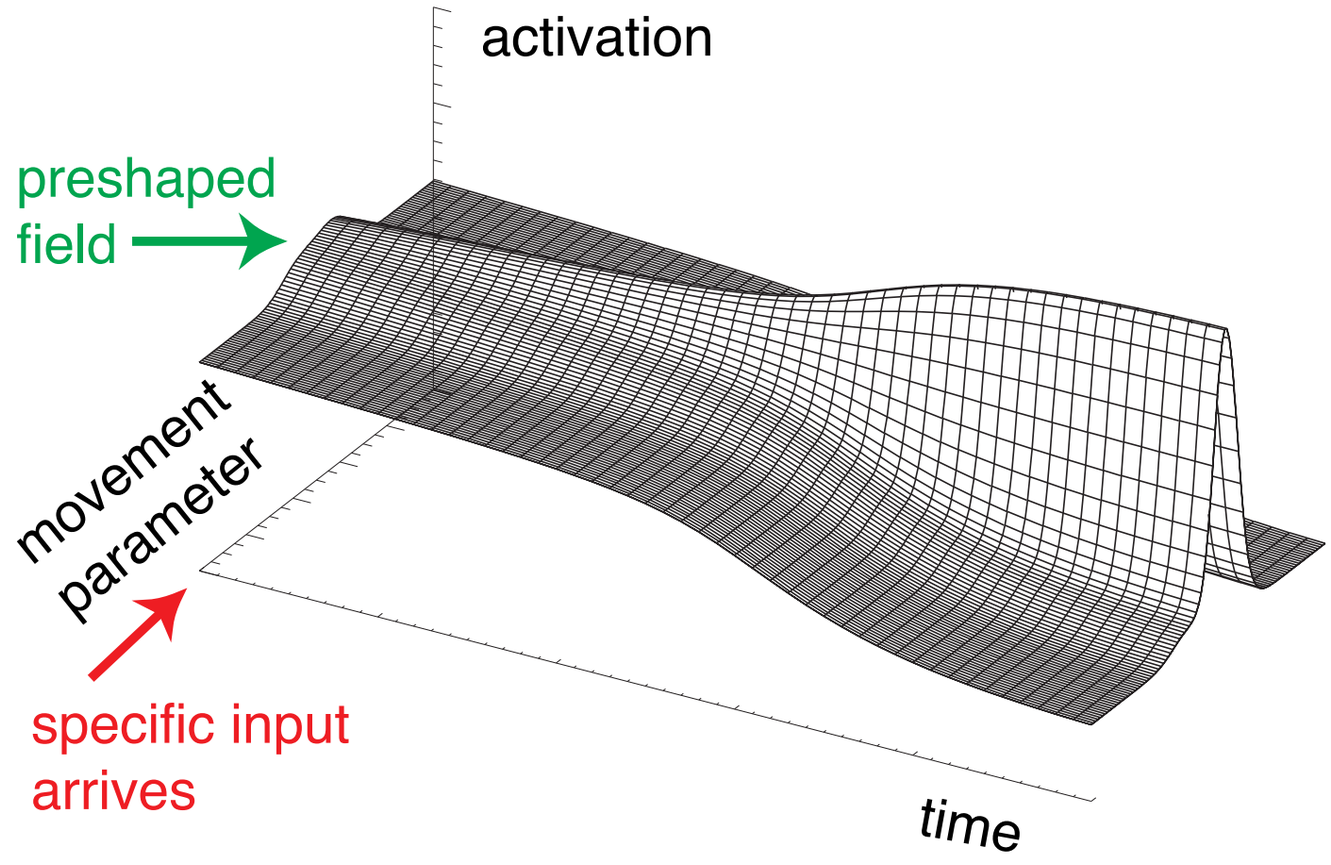
- visual exploration, recognizing screws, while keeping track of spatial arrangement of screws on the toaster (visual cognition, coordinate frames)
- finding tools, applying them to remembered locations, updated by current pose of toaster (working memory, scene represenation)
- manipulating cover, taking it off, recognizing spring, re-attaching it (goal-directed action plan)
- mounting cover back on, generating the correct action sequence (sequence generation)
- background knowledge: cover, screws, how hard to turn screw-driver

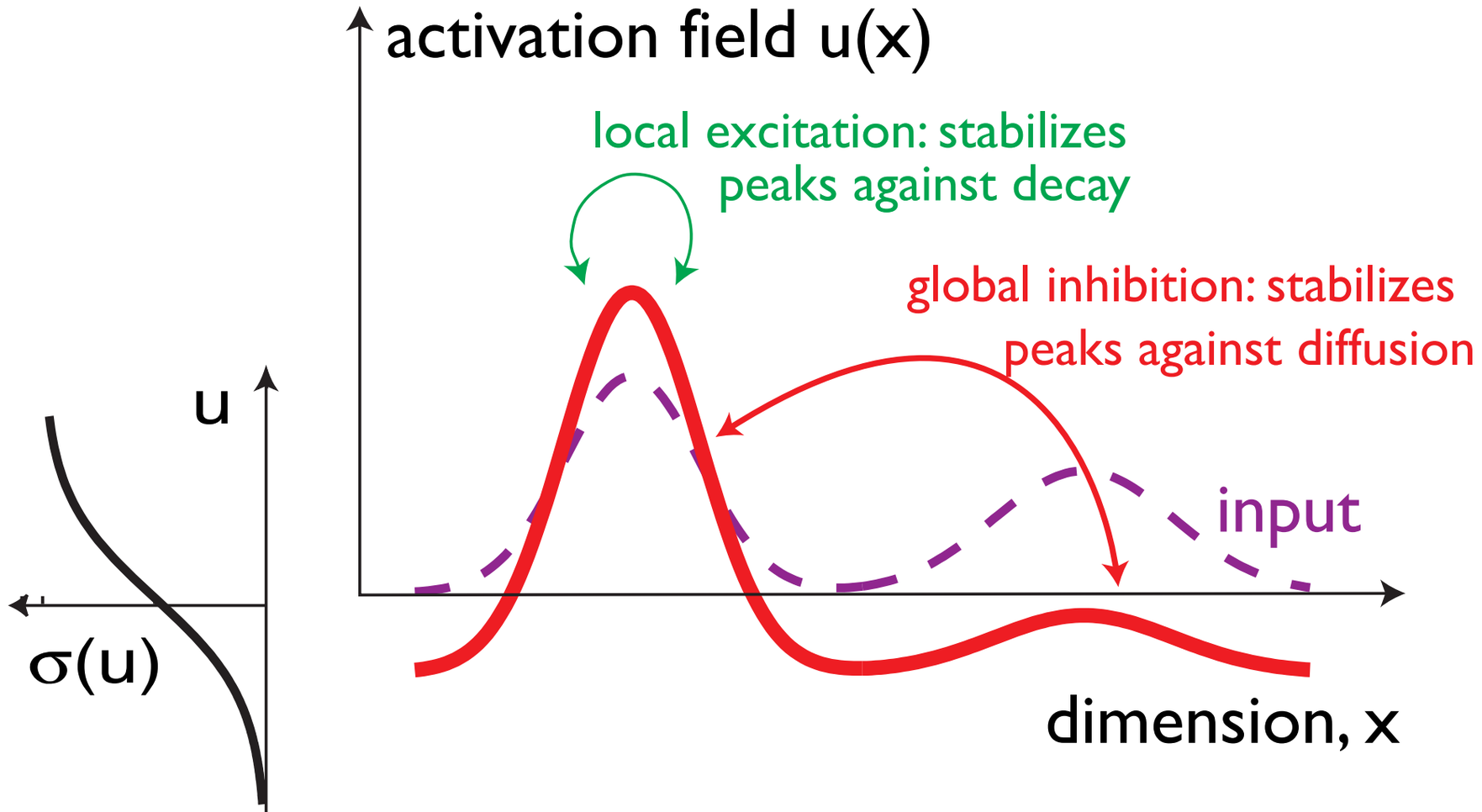
[image: mystery fandom theater 3000]

Neuronal principles

- the neural process of cognition are time continuous and autonomous, not paced by computational steps
- the neuronal populations supporting cognition generate graded activity
- which gains significance only by the connectivity to sensory and motor surfaces and the internal connectivity: continuous space

graded spatial patterns of neuronal activation evolving continuously in time driven by input and interaction





Bayesian thinking

- observations x, e.g., color, shape
- state of the world/system, y, e.g., apple
- observe p(x|y) [how?...]
- learn prior p(y)
- optimally estimate y from p(y|x)~ p(x|y)p(y)

[thanks to Christian Igel]

Jürgen Schmidhuber

- showed how this framework can be used to predict the future from the past
- and how abstract algorithms could make such predictions in an optimal way

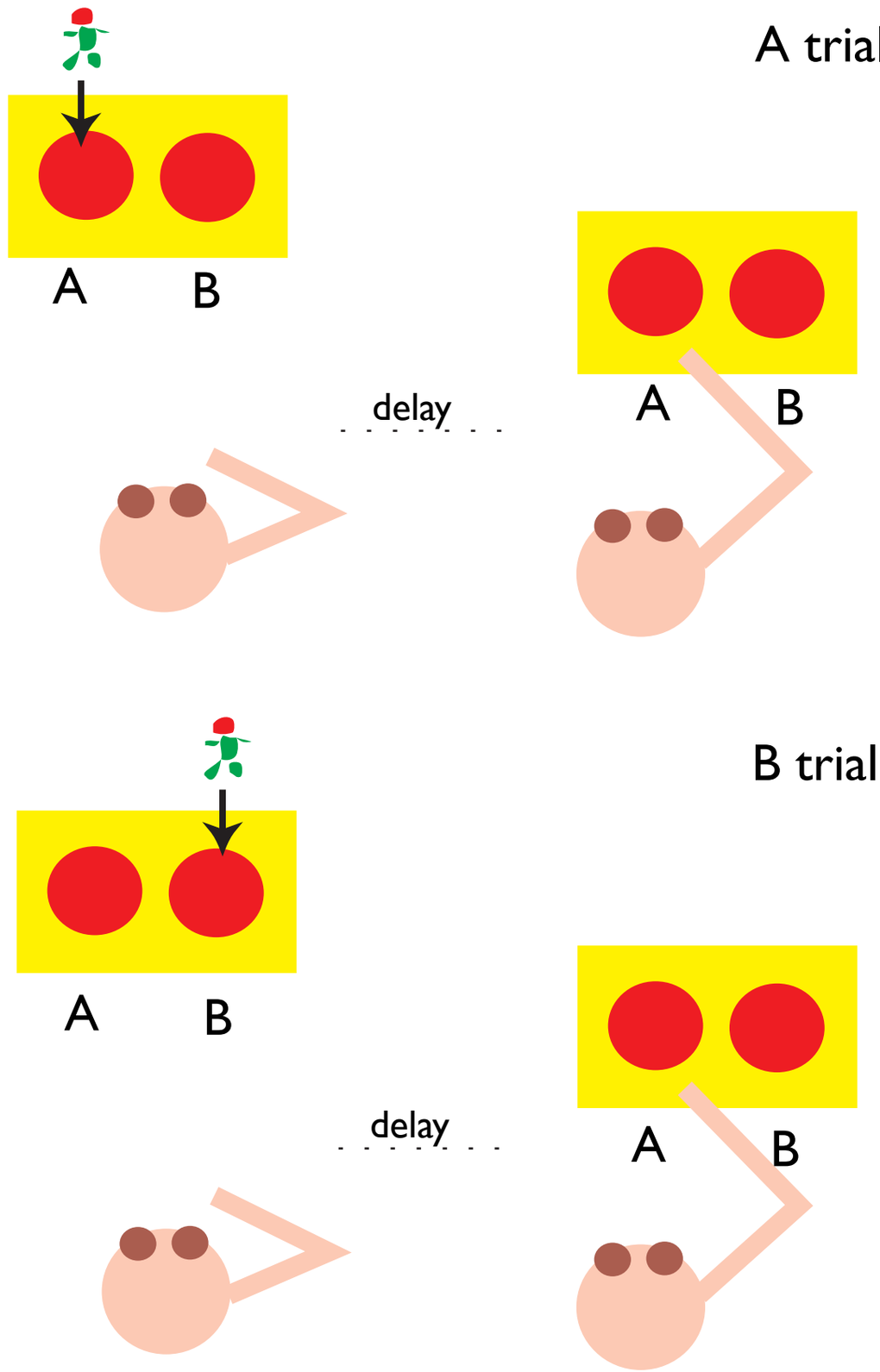
A gap

- (a huge one)
- between the abstraction inherent in these descriptions and the level at which real systems in real time sense and act

Issues

- autonomy
- stability
- integration, behavioral organization
- emergence
- development

Case study: Piaget's "A not B" task

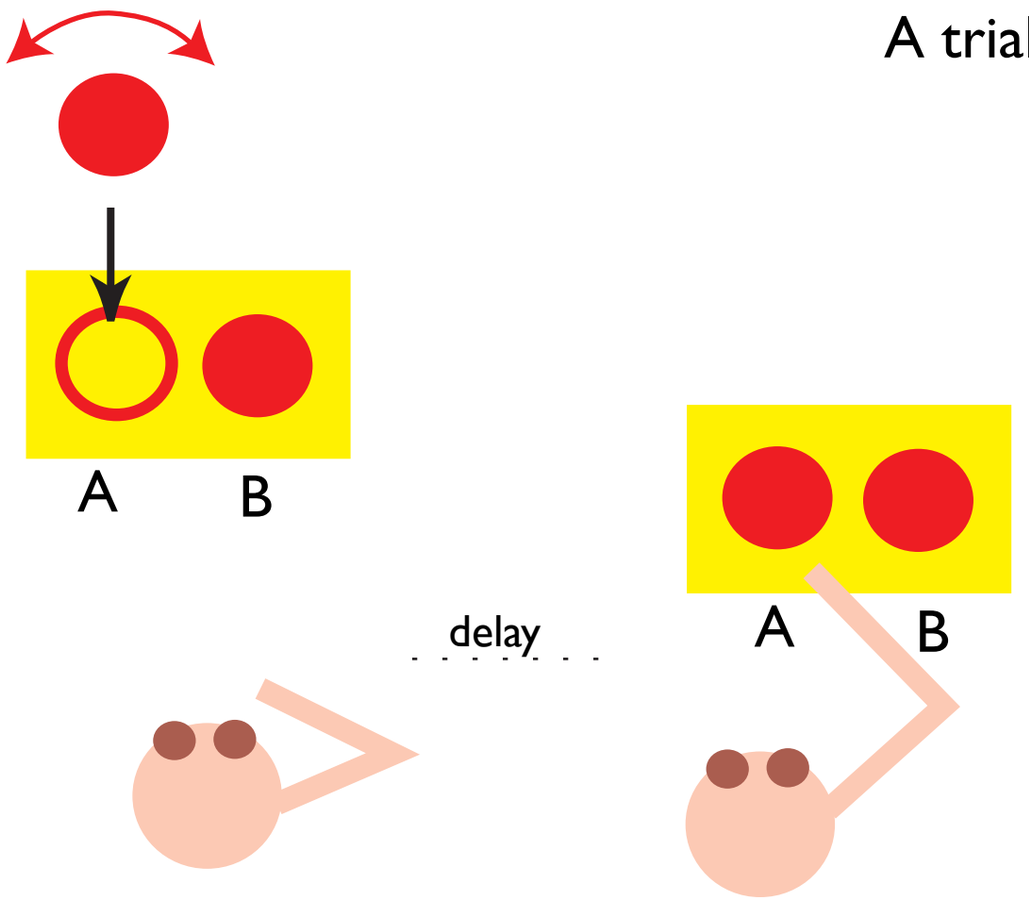


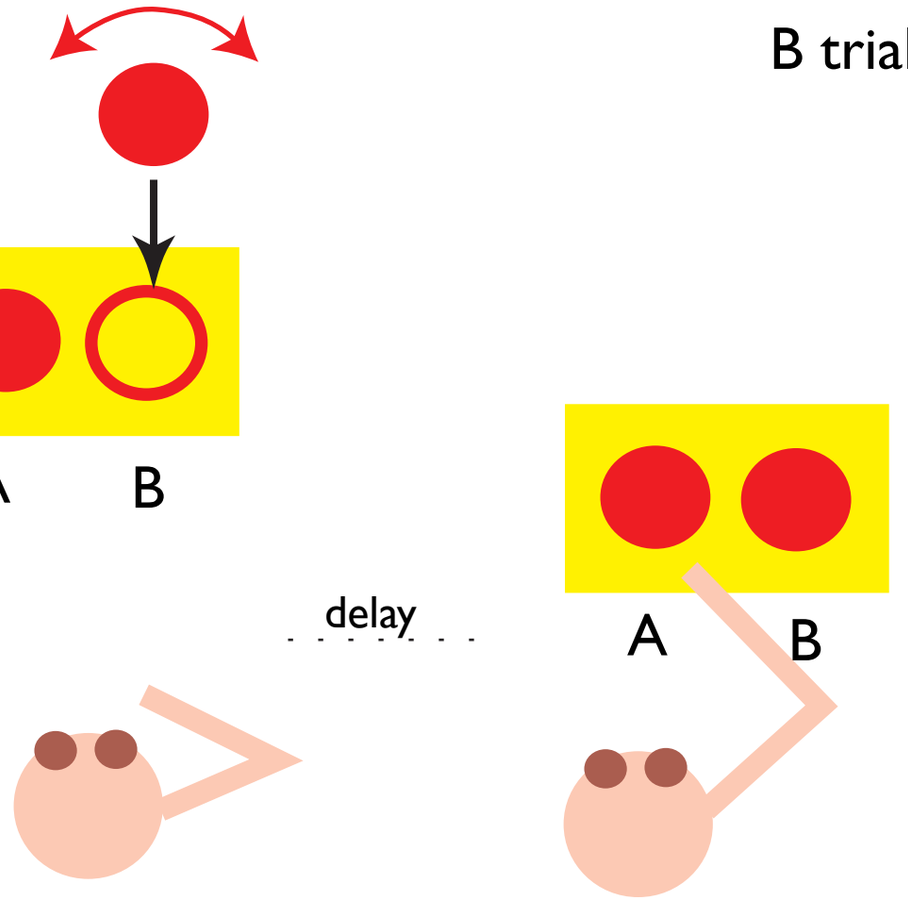
"out of sight - out of mind?"

contains lot's of embodied cognition

- detecting targets/objects
- selecting targets/actions
- stabilizing decisions against distractors
- initiating actions
- learning a habit
- developing

toyless variant of A not B: perseverative reaching





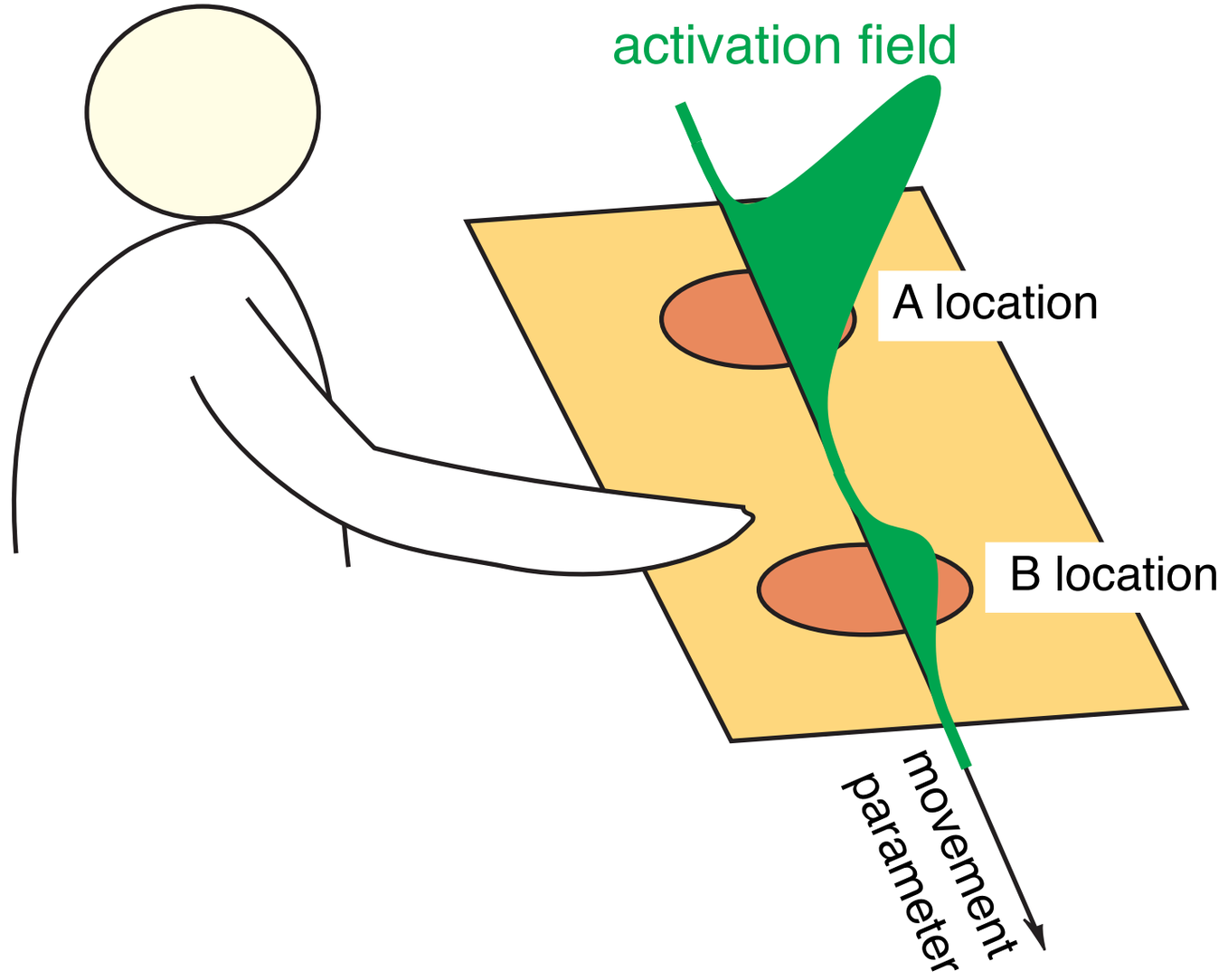
[Smith, Thelen, et al., 1999]

Are young infants optimizing something other than older infants?

- is the language of new AI useful to understand how infants make these reaches?
- is a neuronally mechanistic account useful and possible?

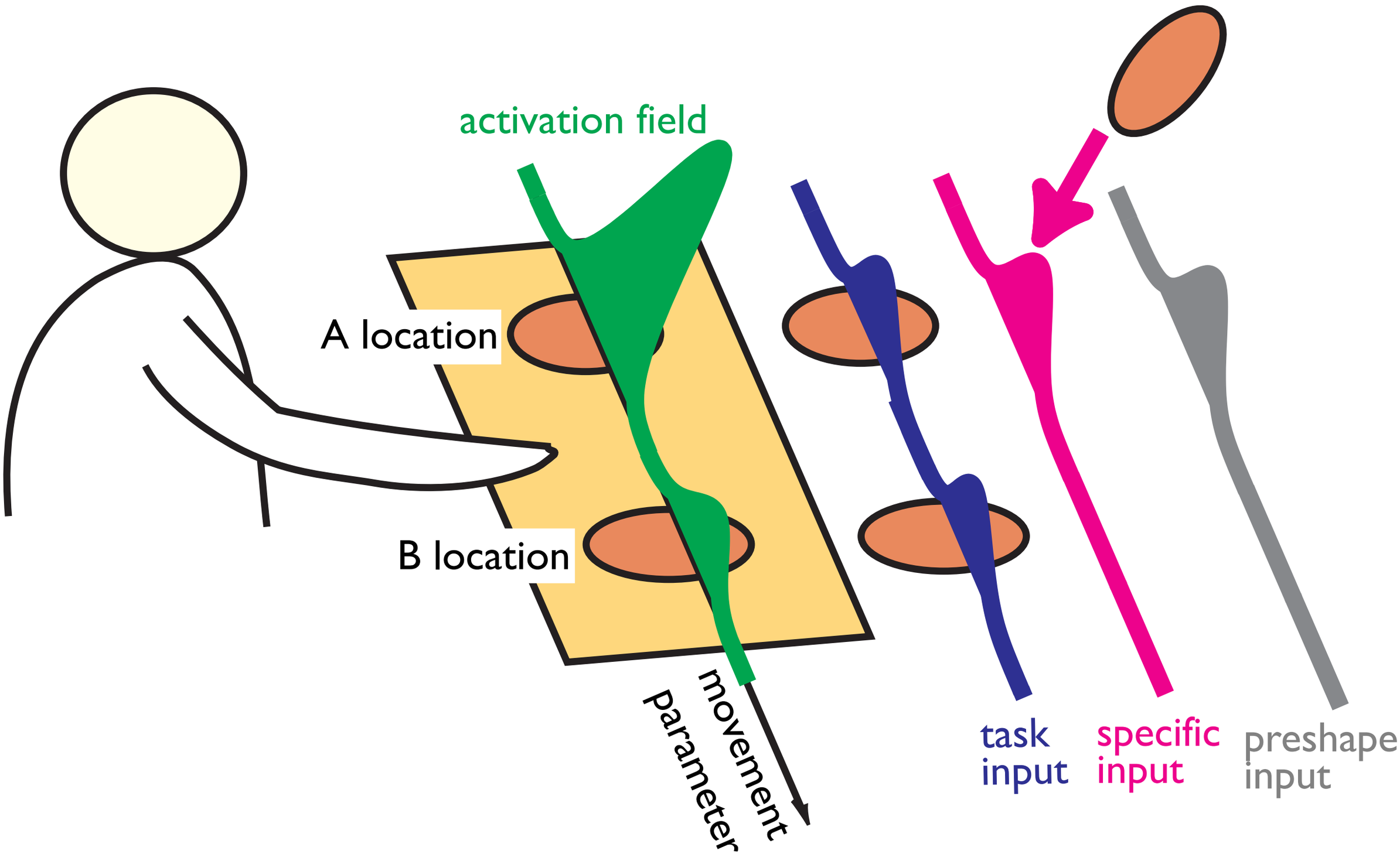
Dynamic Field Theory account

- a field of neuronal activation representing the direction of a targeted directed reaching movement
- a peak of activation represents a motor plan



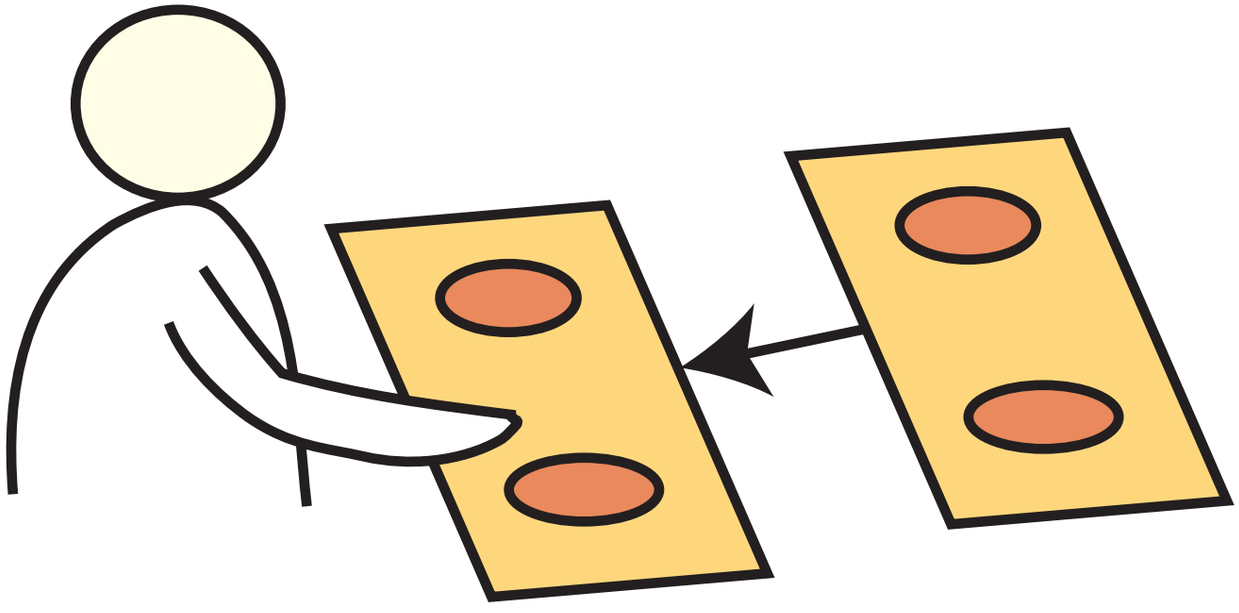
[Thelen, Schöner, et al. , 2001]

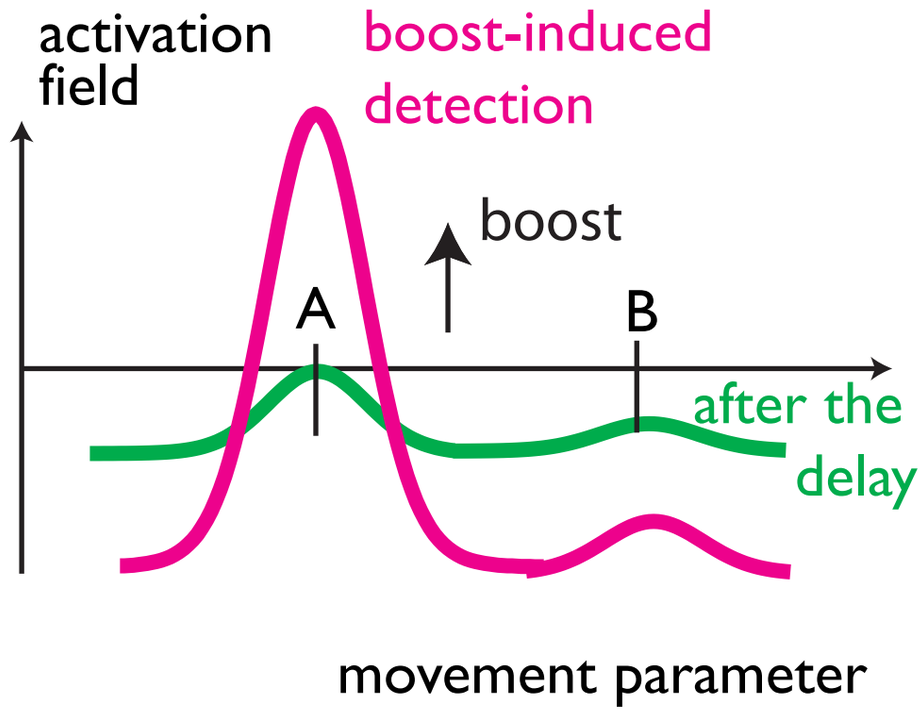
different sources of activation



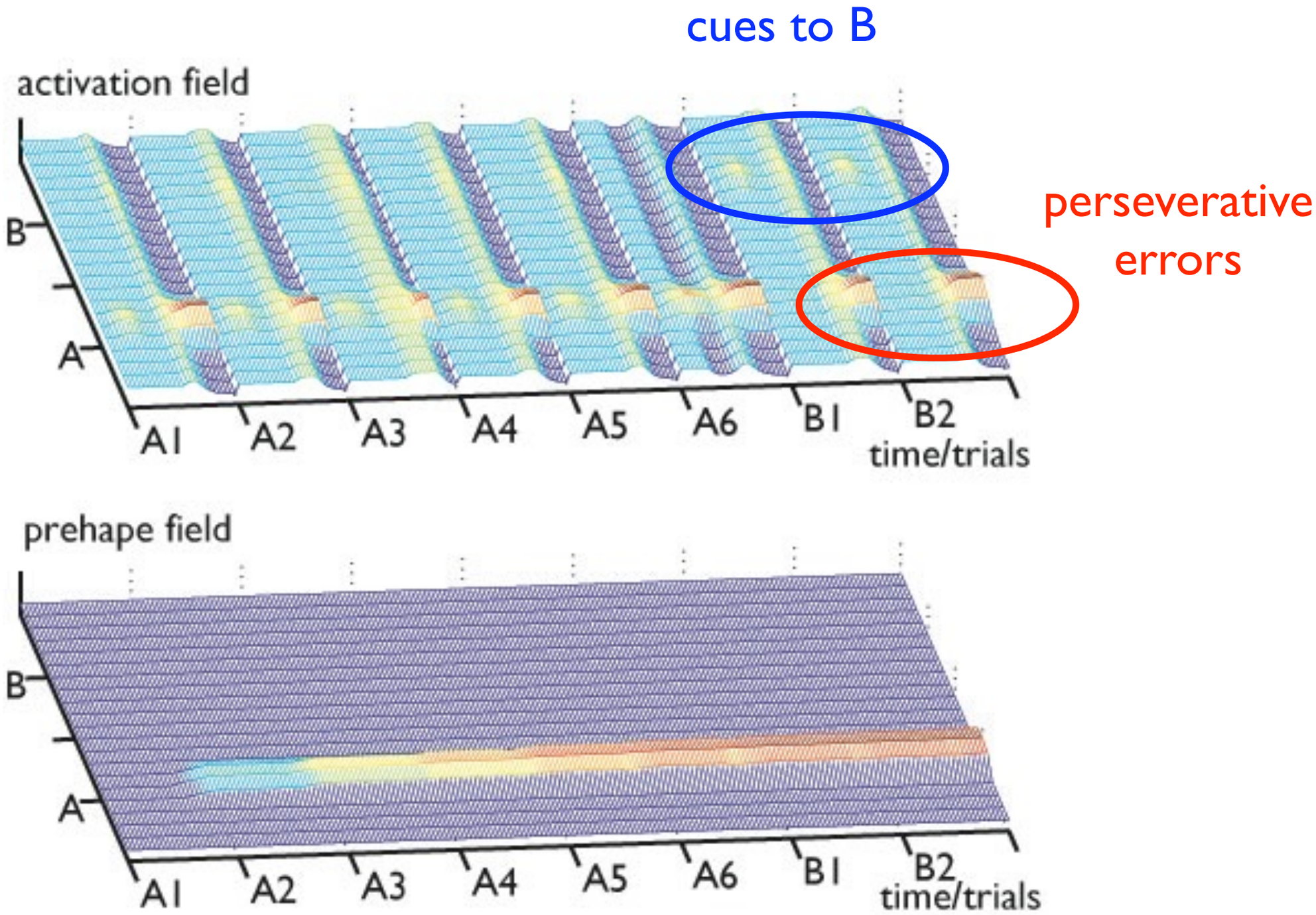
reach initiation

- boost of activation when box enters reaching space => stabilization of a peak and initiation of the reach





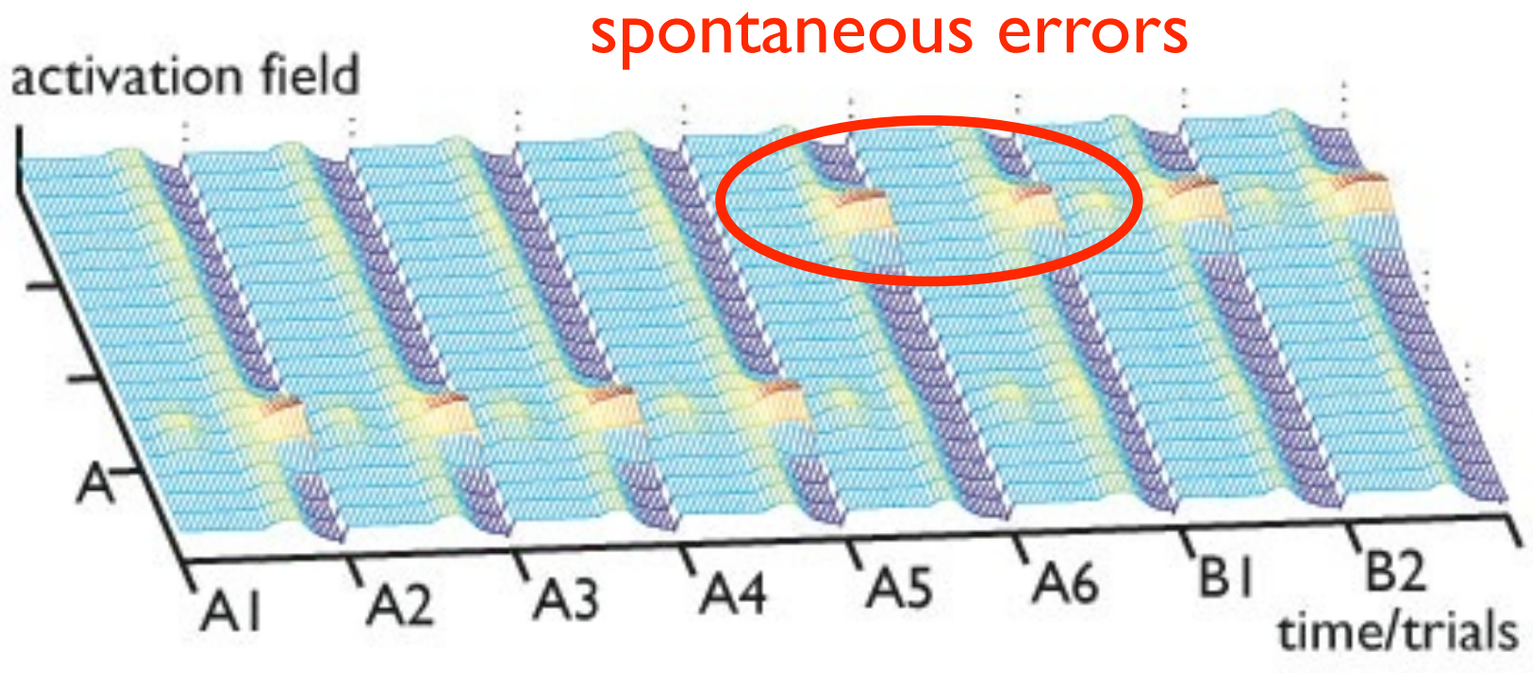
simulation

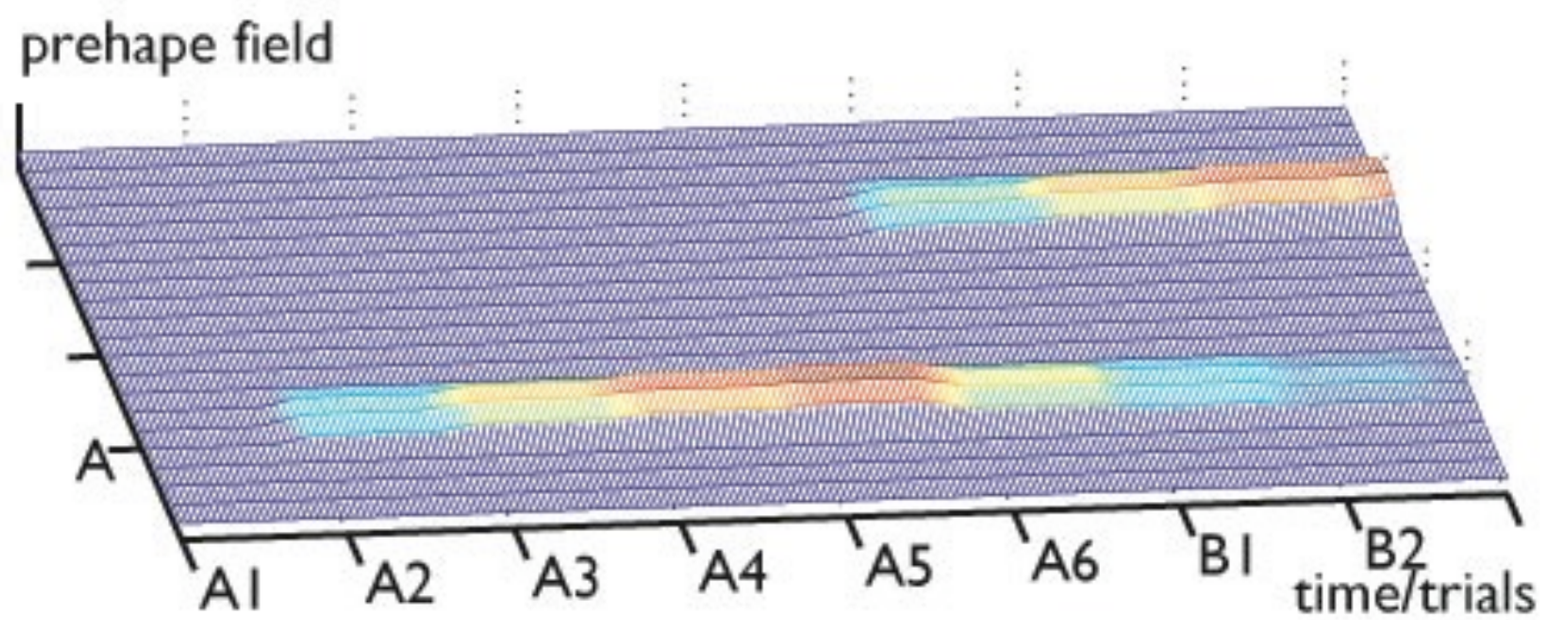


[Dinveva, Schöner, 2007]

behavioral history matters

- spontaneous errors=reaches to B on A trials
- leave a memory trace at B
- which reduces the A not B error



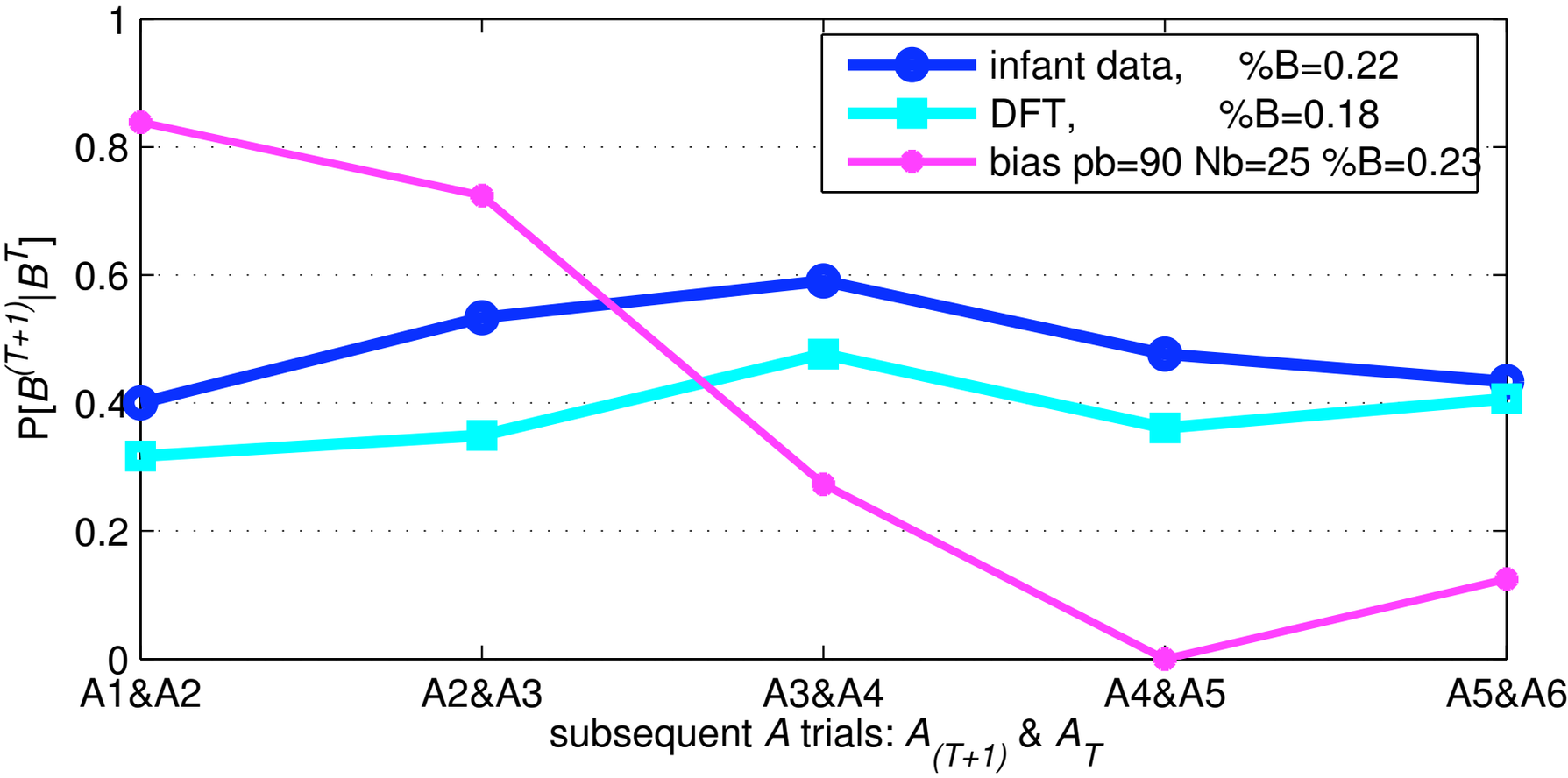


[Dinveva, Schöner, 2007]

behavioral history matters

- spontaneous errors promote more spontaneous errors

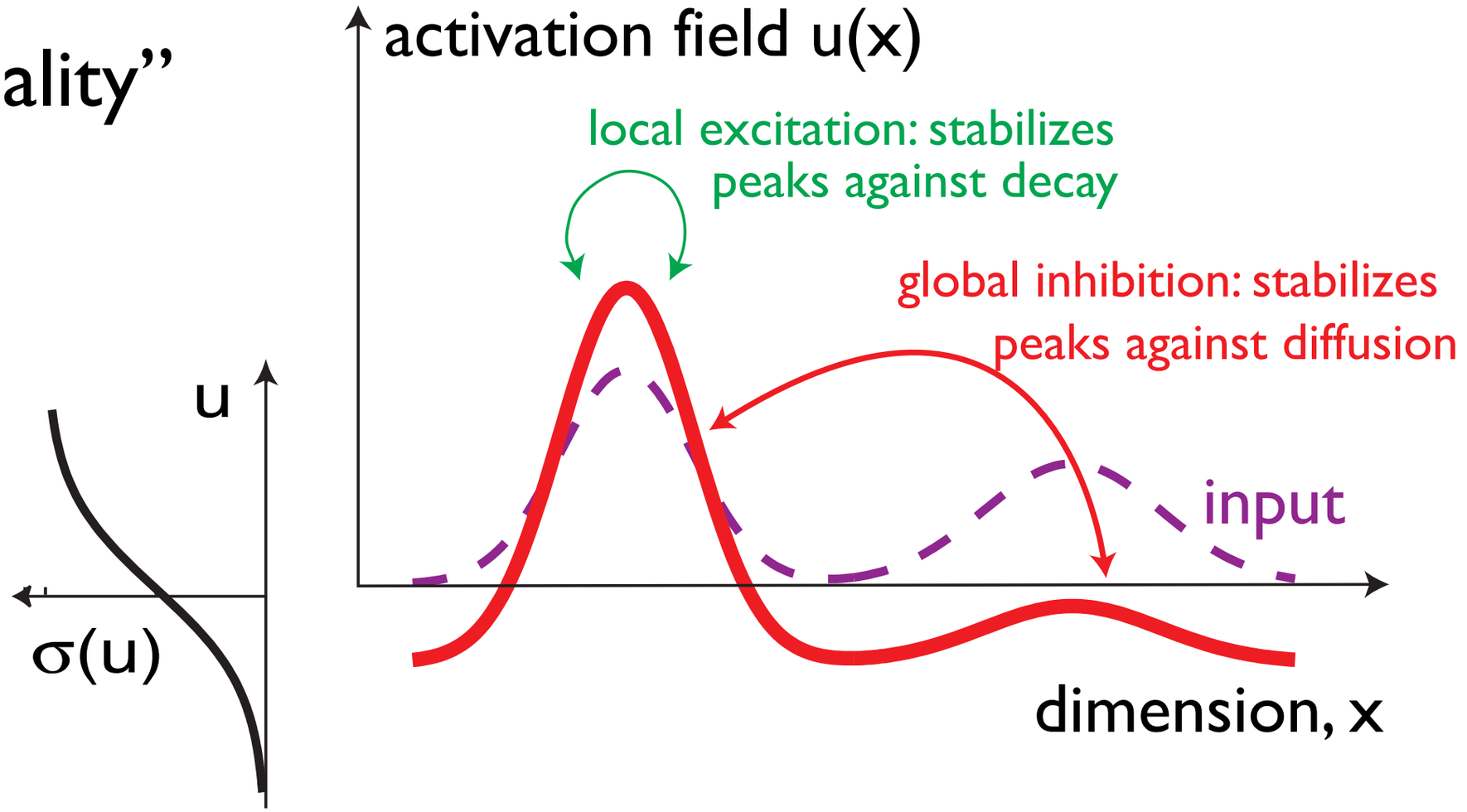
first and **second** reaches to *B* are on two subsequent *A* trials



[Schoner, Dineva, Developmental Science, 2007]

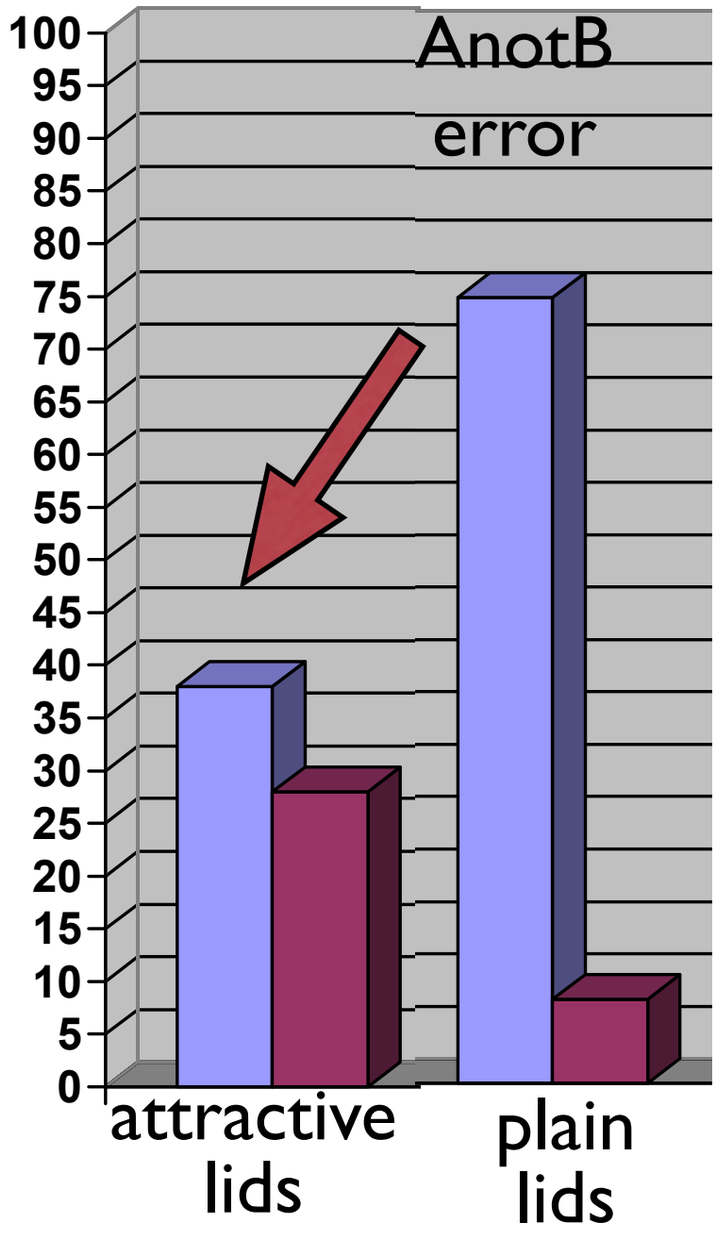
underlying neural principle

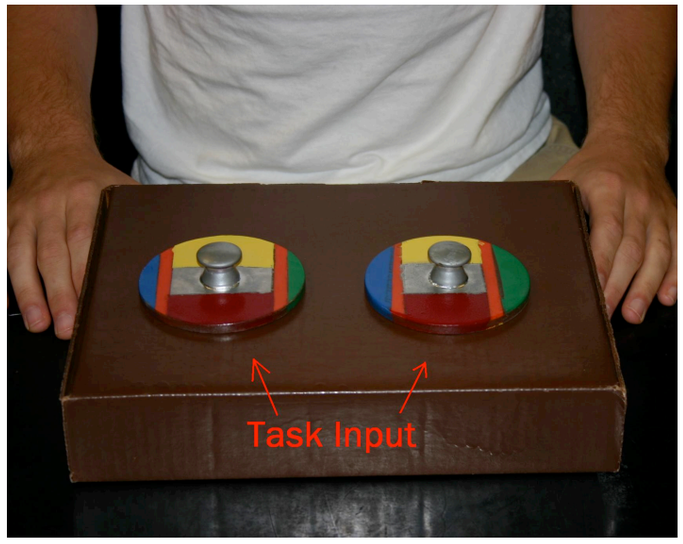
- bistability
- "circular causality"
- emergence

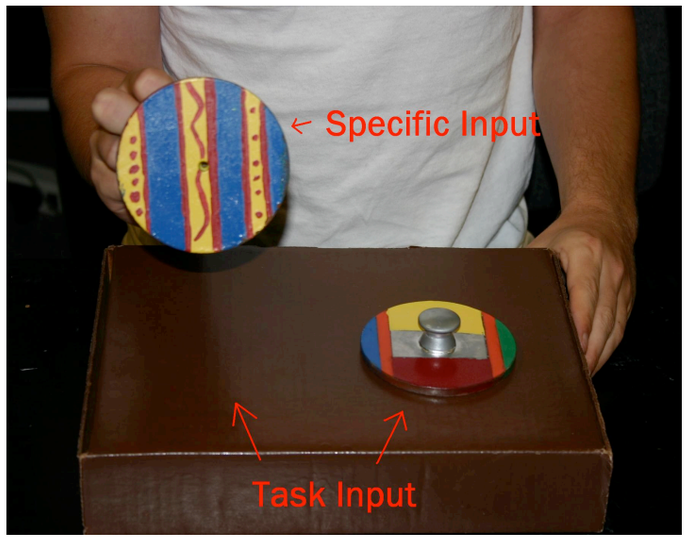


emergence: suppressing the A not B error by "pumping up neural energy"

- making both locations more attractive reduces the A not B error
- as predicted by DFT



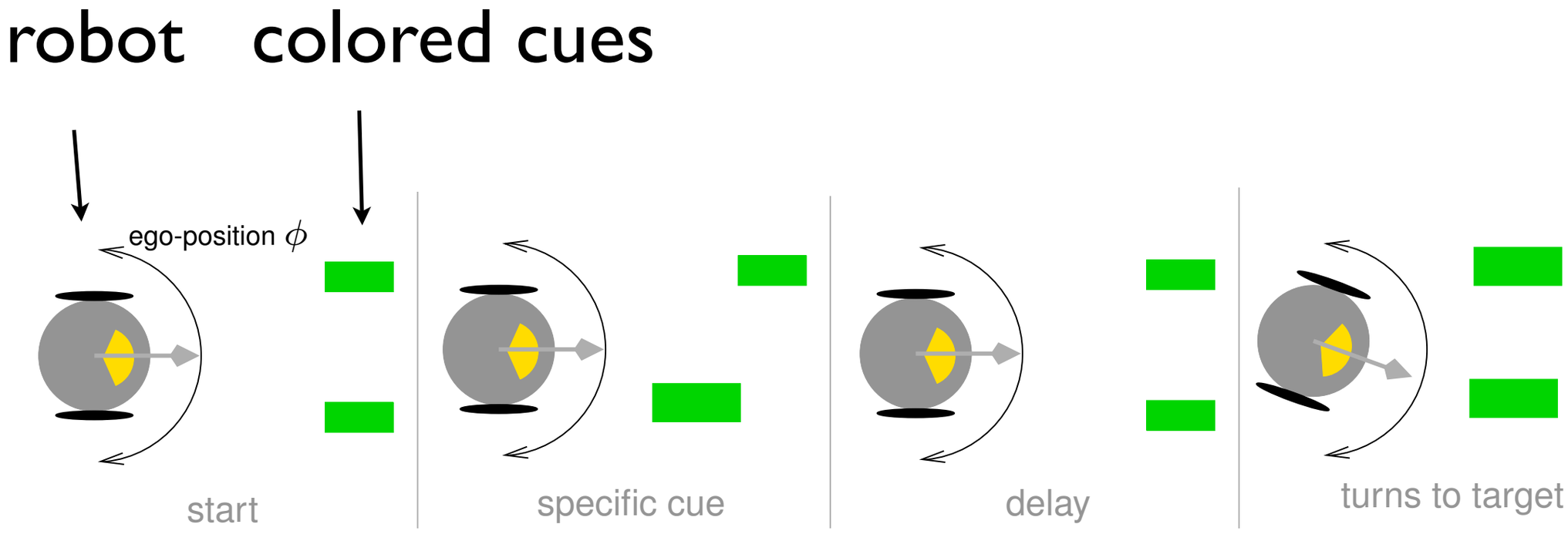


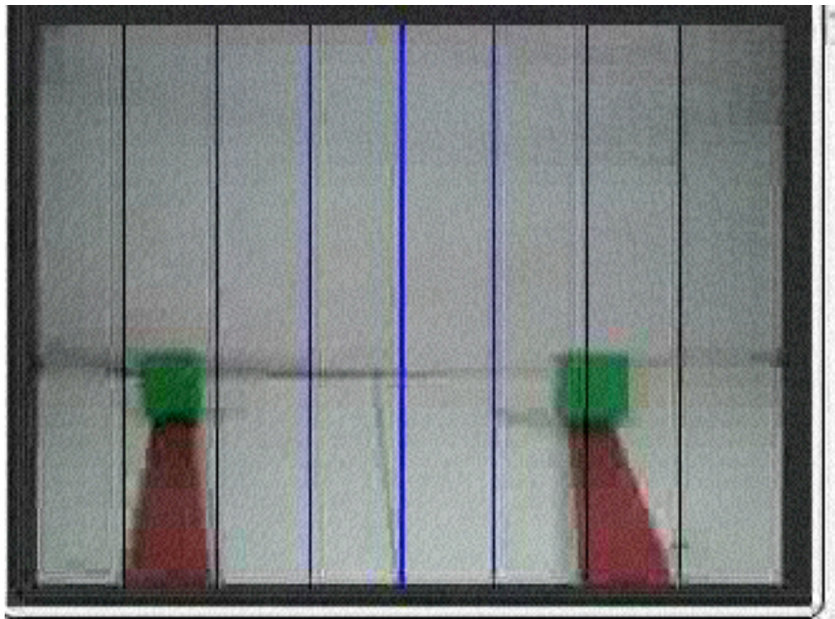


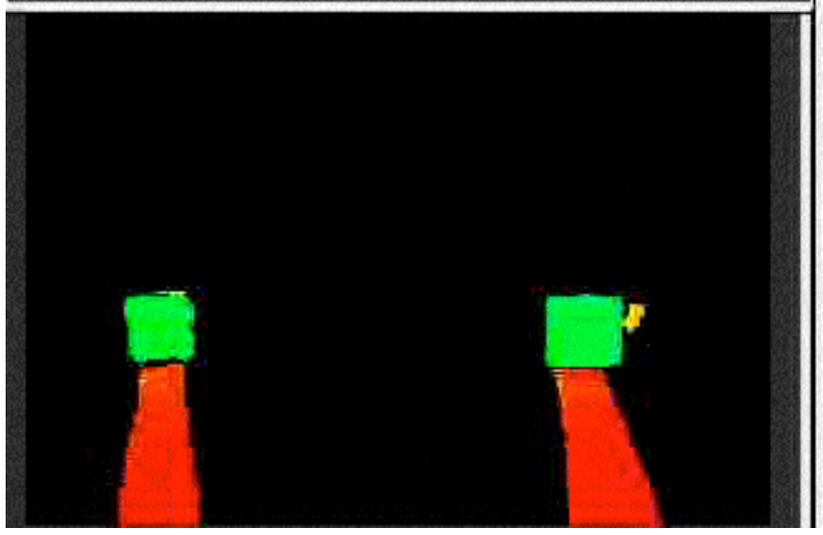
[Anderson, Dineva, Schöner, Thelen (2009)]

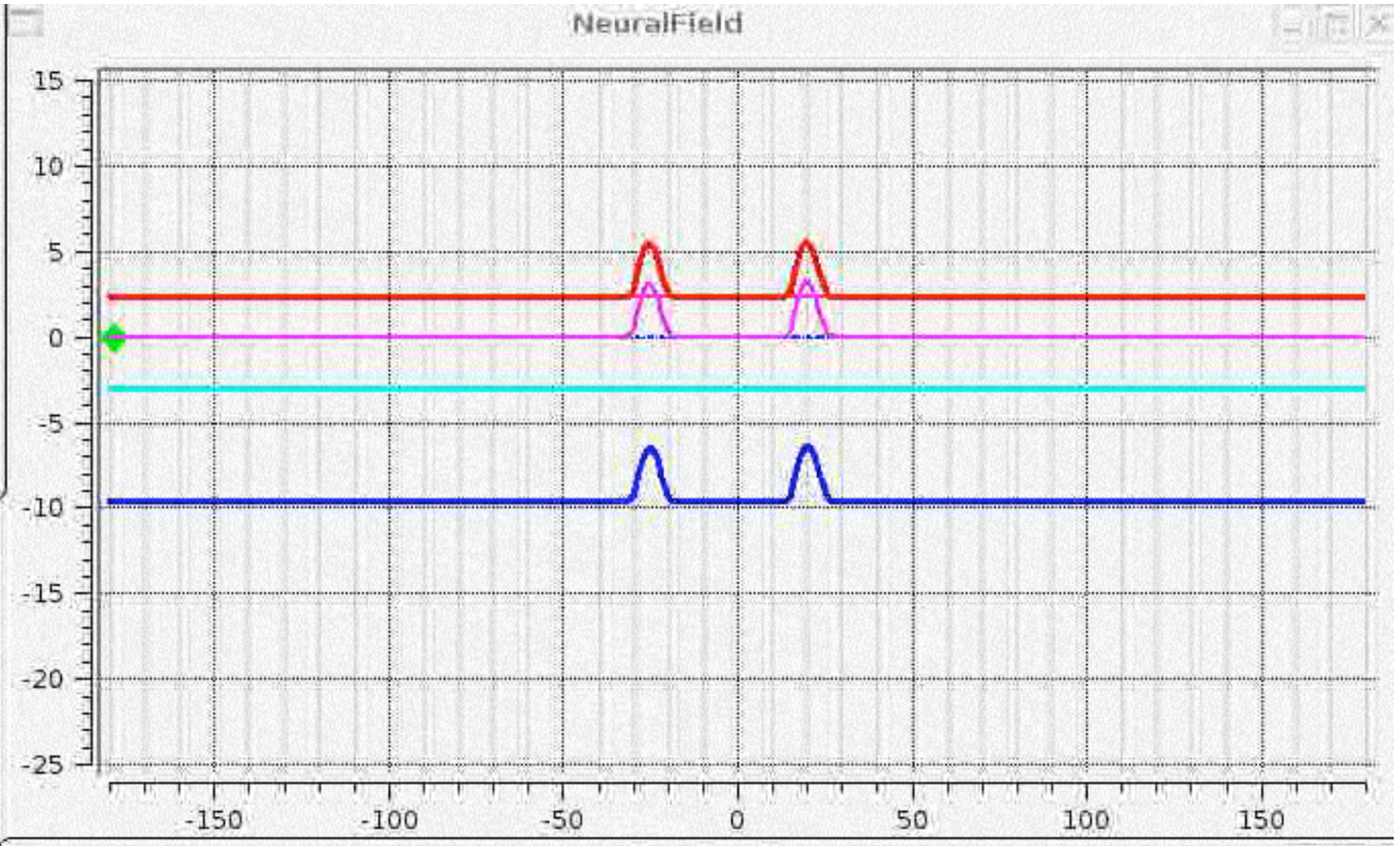
is this account actually embodied?

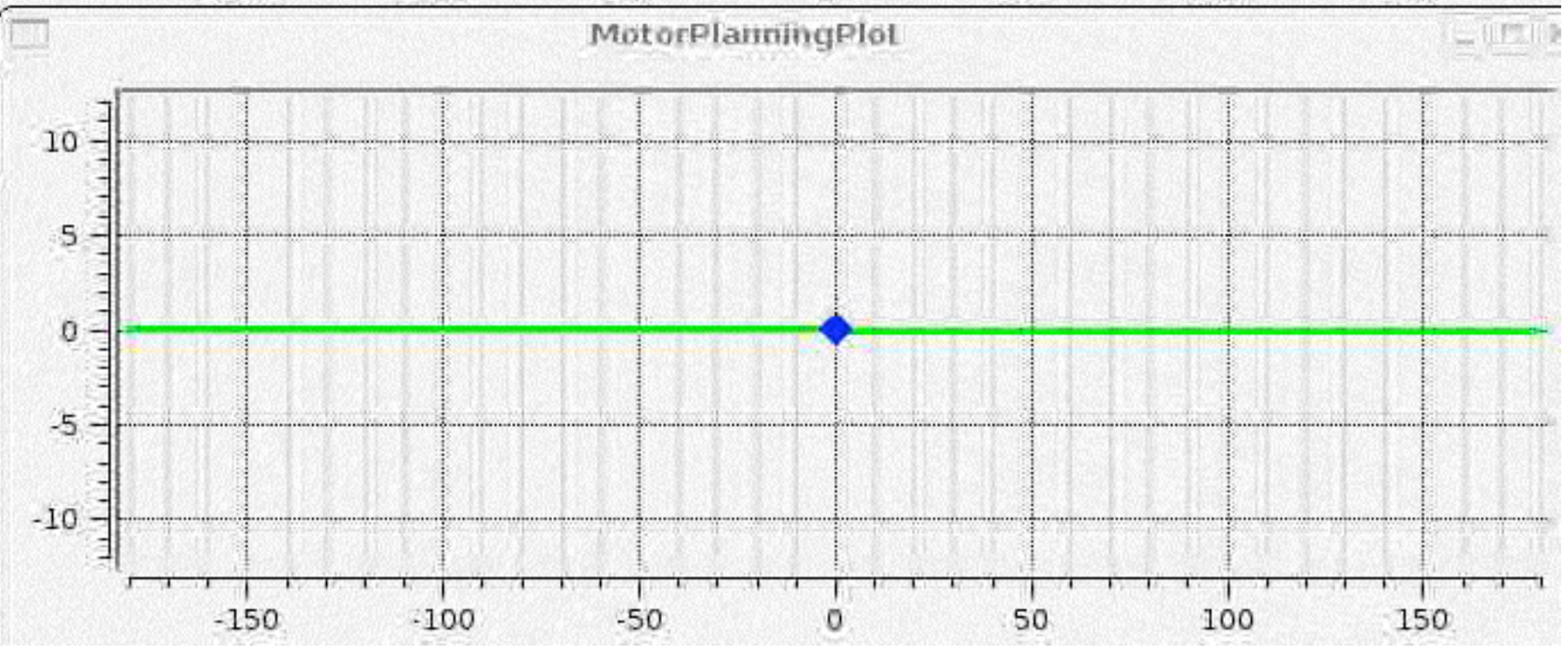
- proof by building a robot the instantiates the field dynamics
- and behaves like the infants do

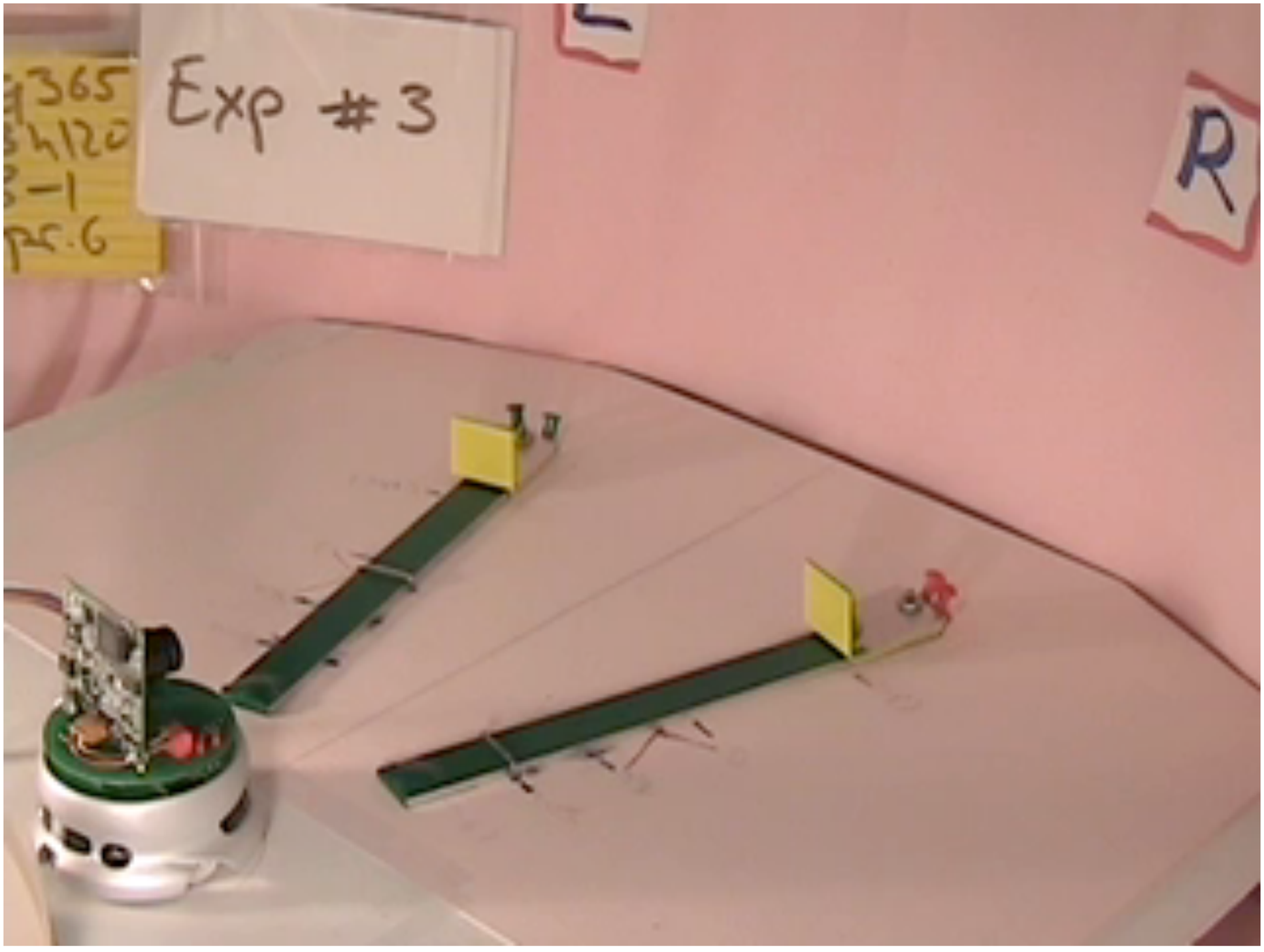






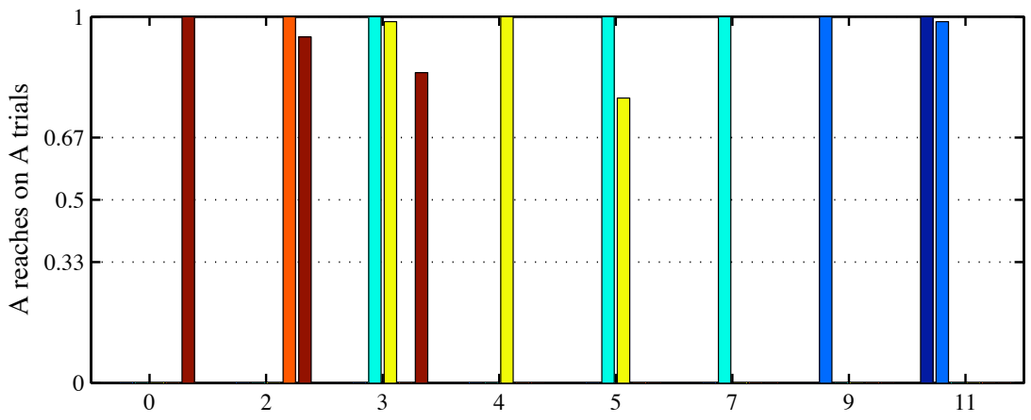

NeuralField
15
10
5
0
-5
-10
-15
-20
-25
-150
-100
-50
0
50
100
150


MotorPlanningPlot
10
5
0
-5
-10
-150
-100
-50
0
50
100
150


Exp #3
R

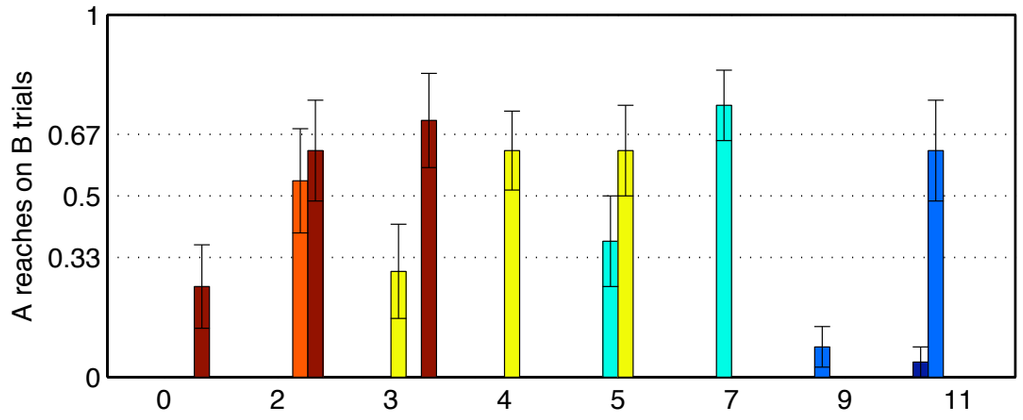
result: reproduce fundamental age-delay trade-off in A not B

A reaches on A trials



A reaches on B trials

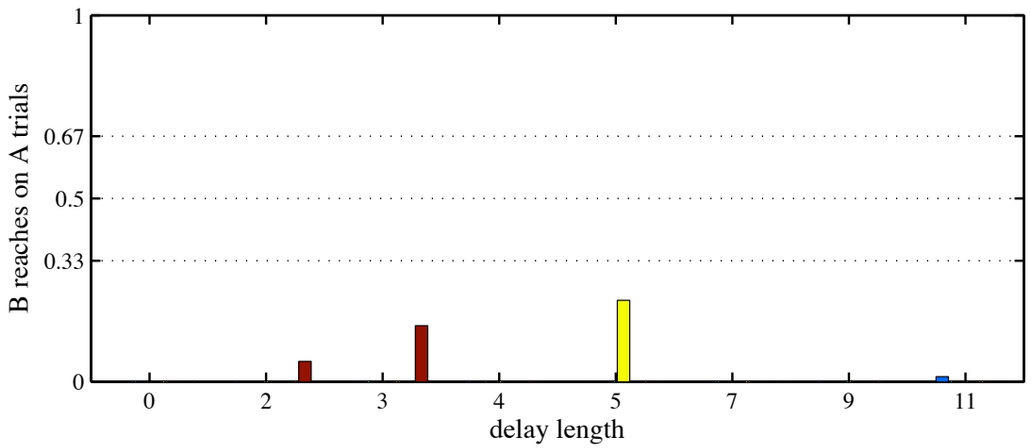
old



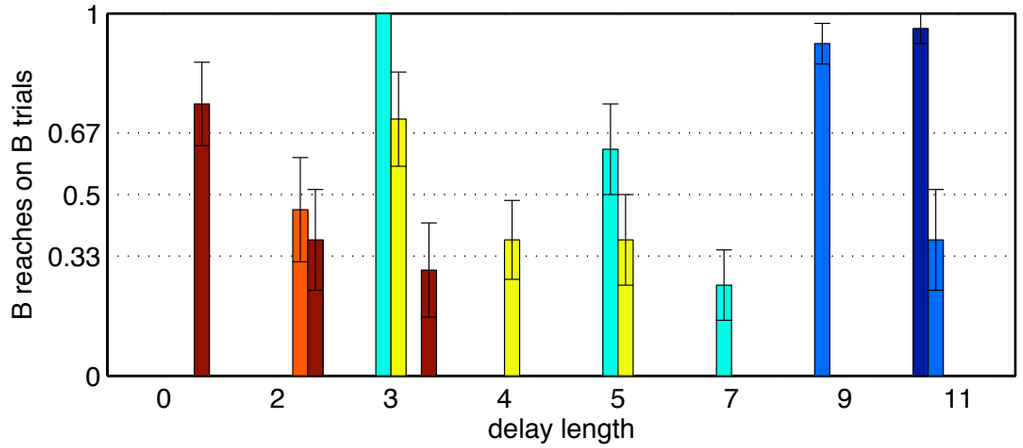
young

h_{rest}

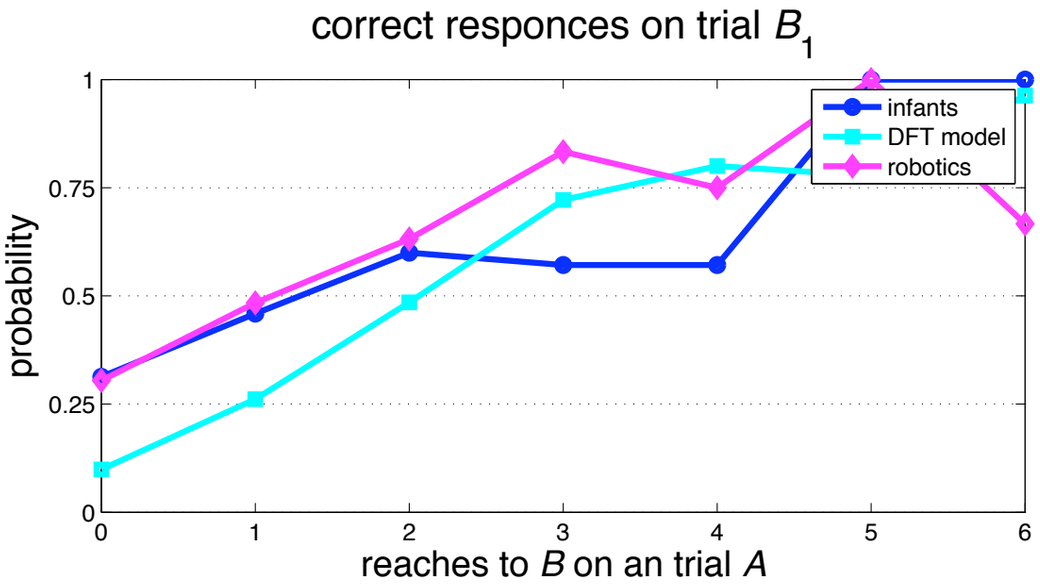
B reaches on A trials

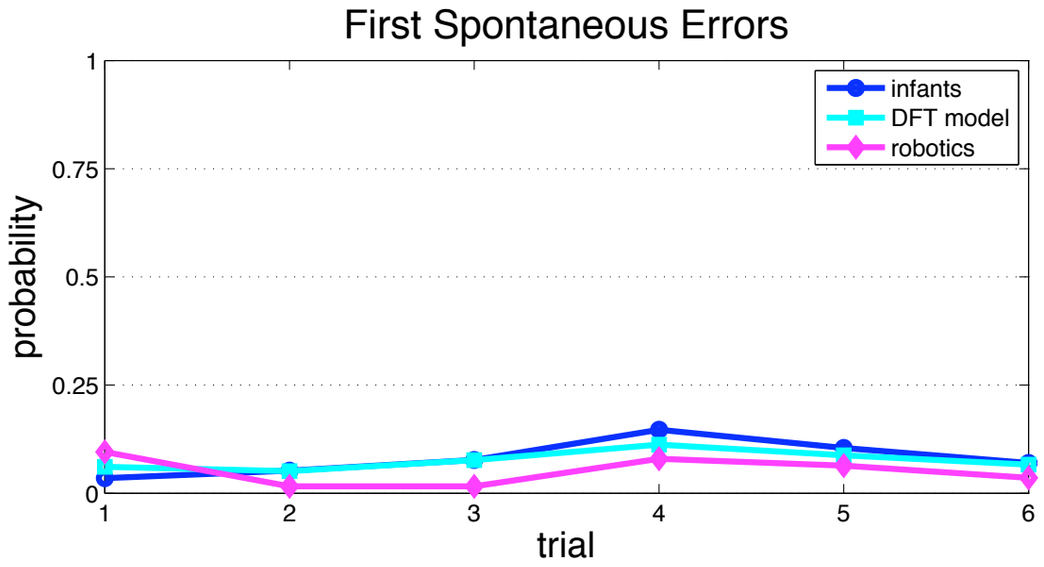


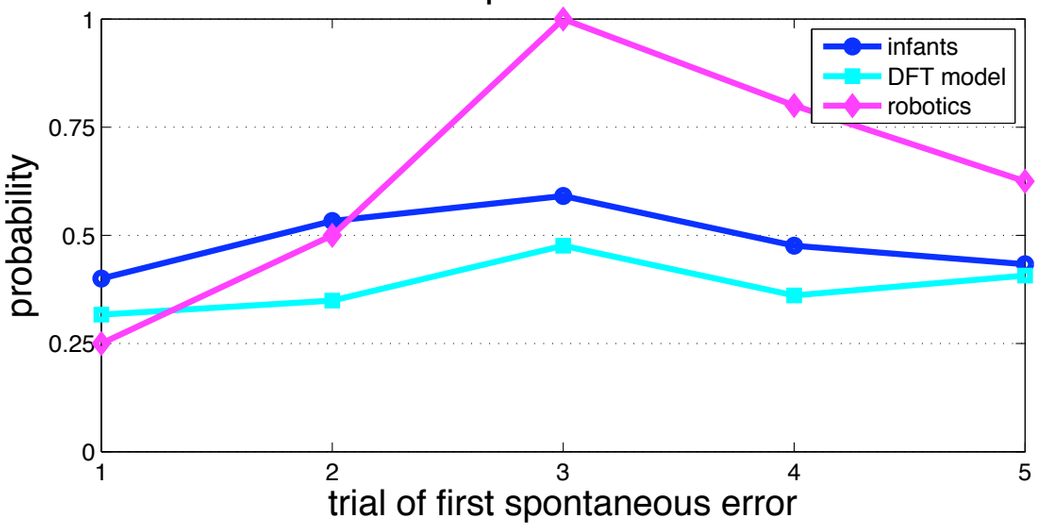
B reaches on B trials

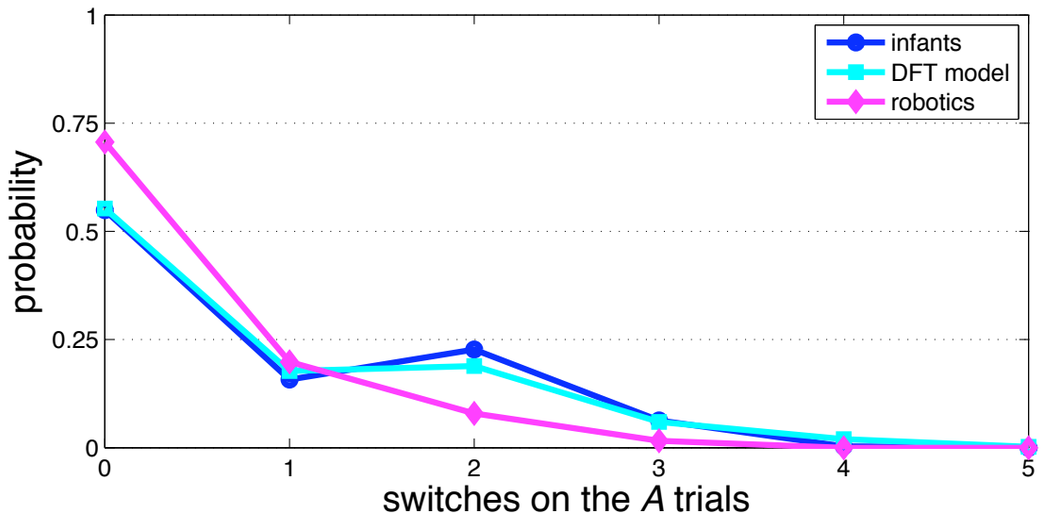


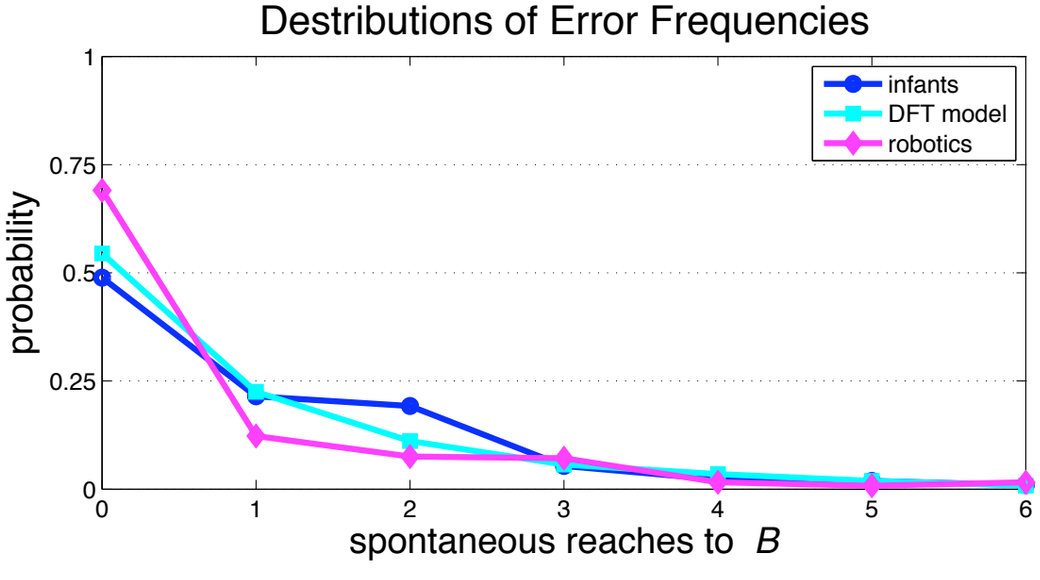
h_{rest}


correct responces on trial B_1
probability
reaches to B on an trial A
infants
DFT model
robotics


First Spontaneous Errors
probability
trial
infants
DFT model
robotics


Second Spontaneous Errors
probability
trial of first spontaneous error
infants
DFT model
robotics


Destributions of Switches
probability
switches on the A trials
infants
DFT model
robotics

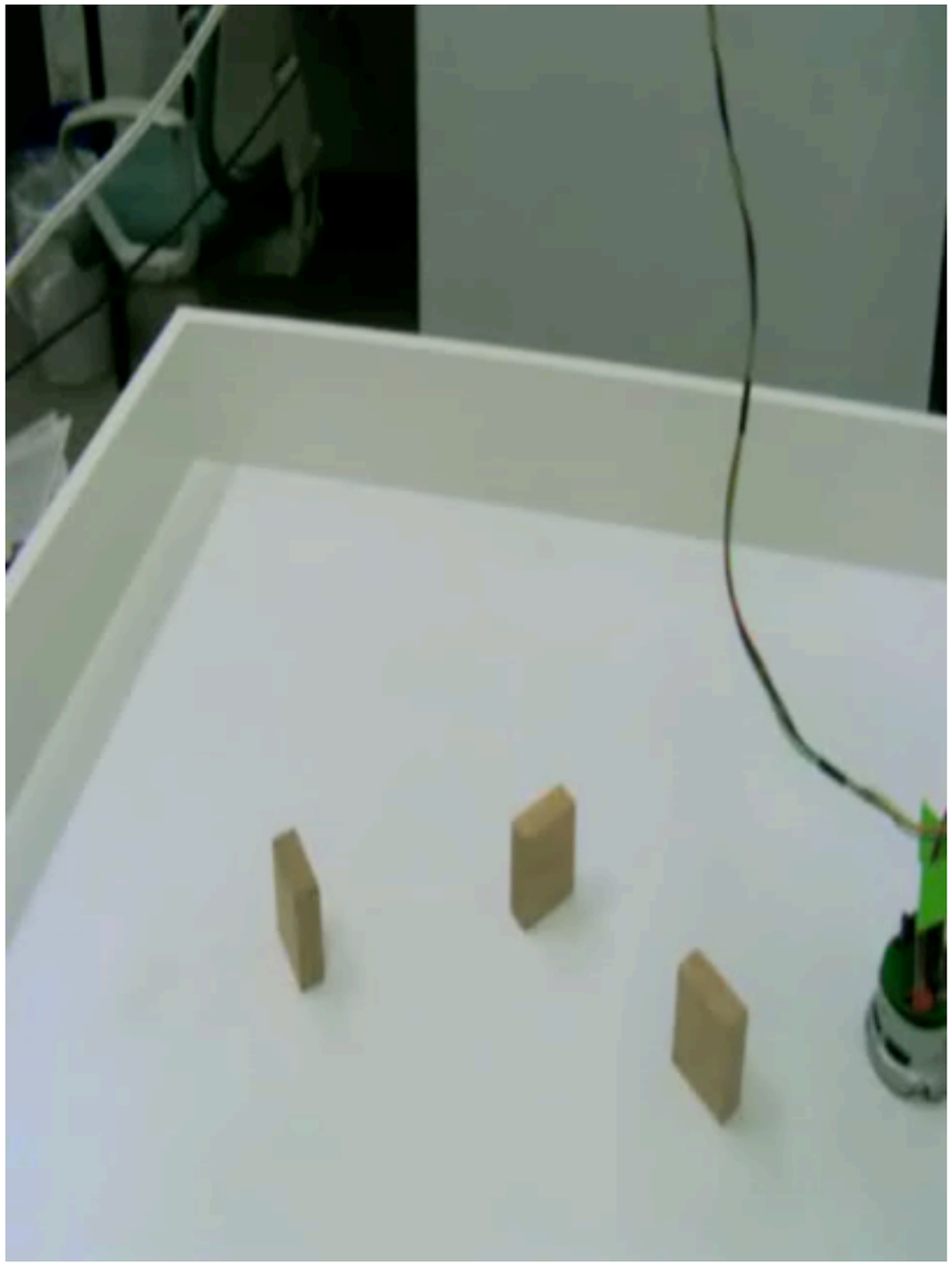

Destributions of Error Frequencies
probability
spontaneous reaches to B
infants
DFT model
robotics

heuristics

- this implementation uncovered a conceptual error in an earlier Dynamic Field Theory model
  - in which the selection decision was done by read-out of maximally activated action at each time step (selecting the optimal action?)
  - this lead to random fluctuations between the two targets, leading to averaging
- need to stabilize decisions when system in continuously linked to sensory input

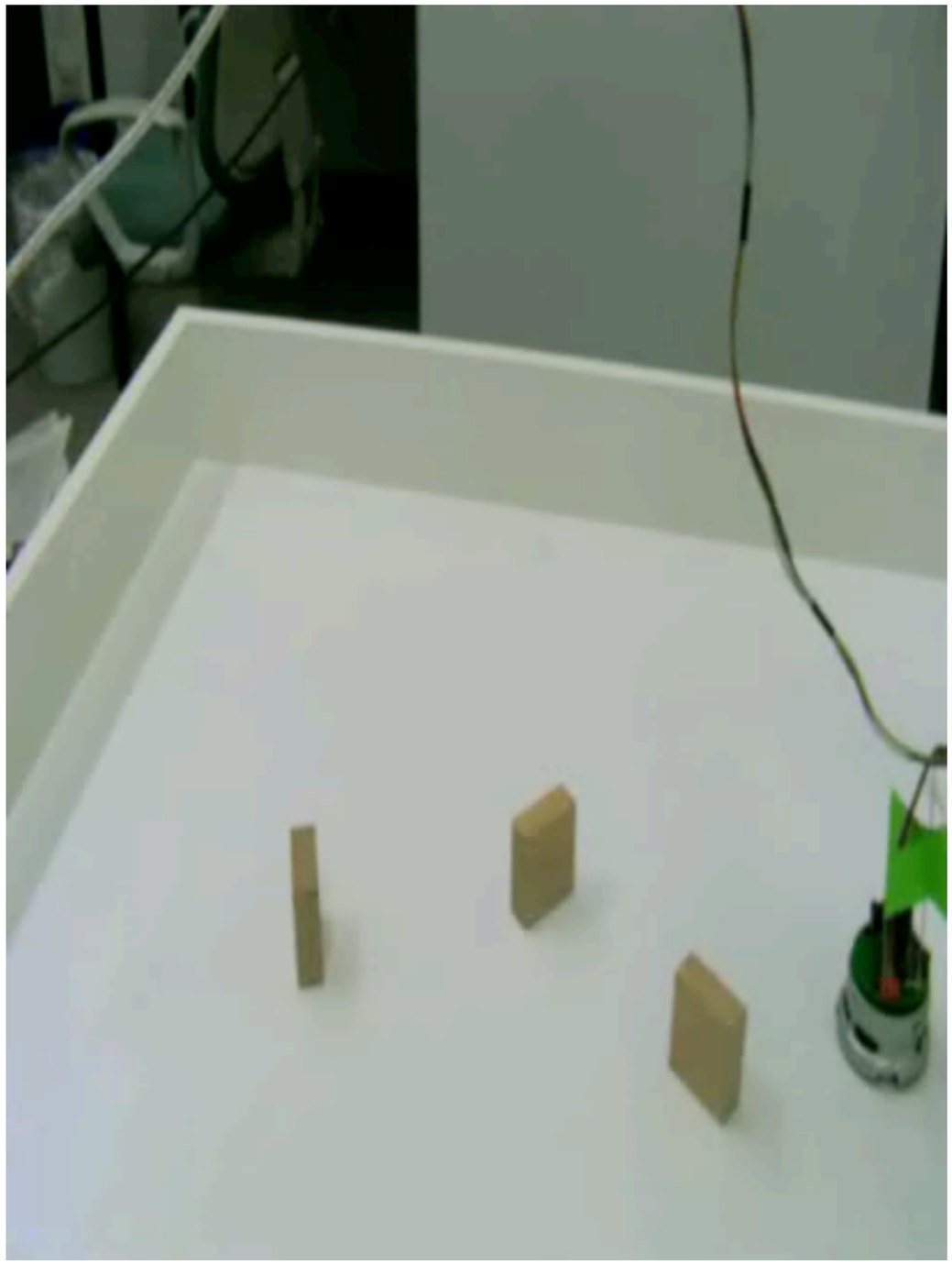
“young” robot

target



“old” robot

target

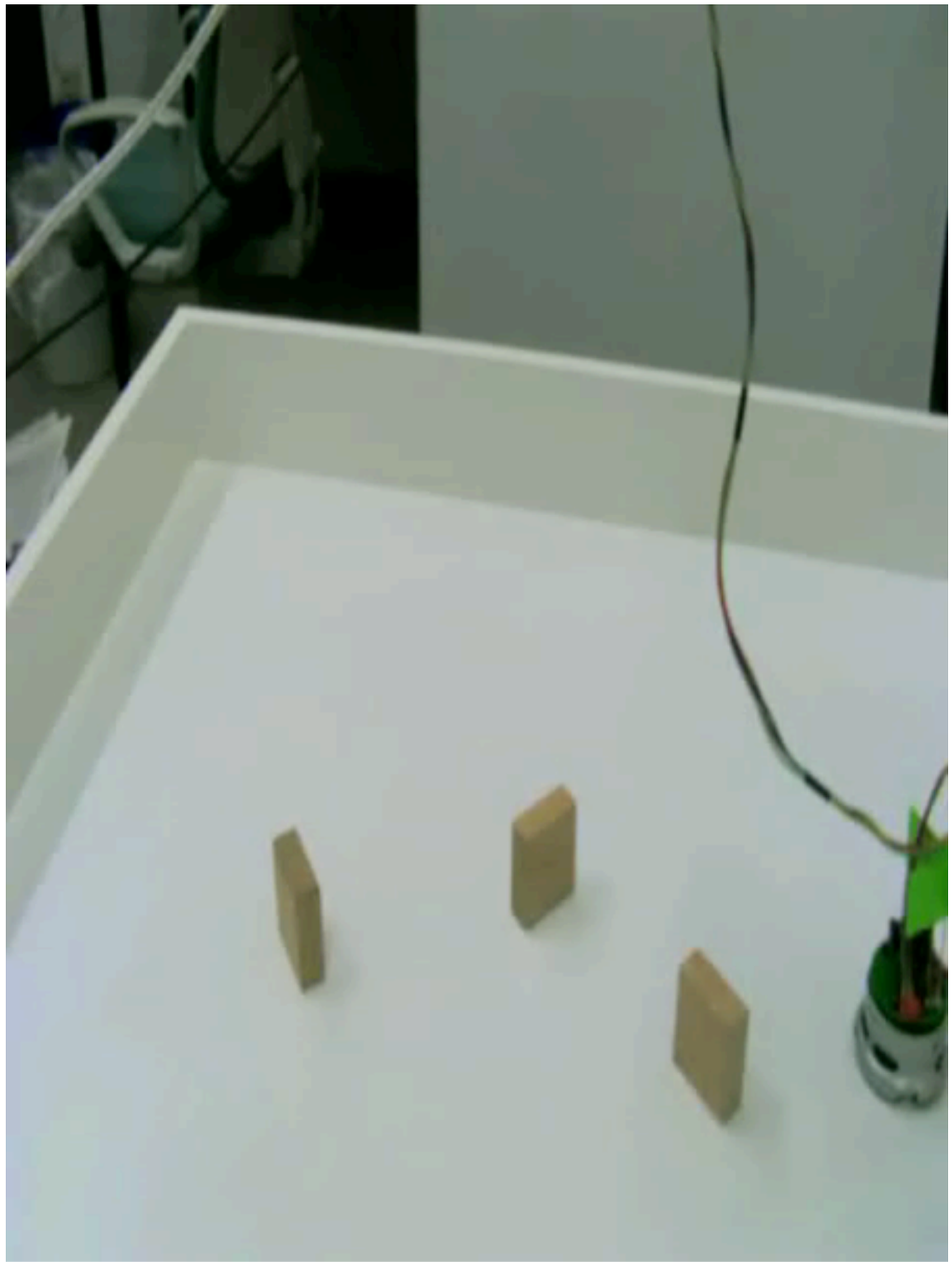


no memory trace

habit formation stabilizes behavior

"young" robot with memory trace

target



preliminary conclusion

- ... this example illustrates
  - how theoretical thinking can be used to investigate the emergence of cognition dependent on context, history, etc.
  - how this has a very different quality that talking about something abstract being optimed
- more and deeper examples will be available later
  - ...Yulia Sandamirskaya on sequence generation
  - ... John Spencer's Horizon Lecture

Is the new AI about human cognition?

- officially not
- but why is human cognition used so extensively to motivate and interpret the ideas?
  - "the brain is performing about 10 trillion instructions per second"
- is new AI a solution in search of a relevant problem?

Conclusion

- real, human, embodied cognition offers rich heuristics of what are relevant problems that a general new AI might want to solve
  - solving these may require addressing issues like autonomy, stability, integration, and development
- these seem not very closely related to the problem of making AI formal
  - in fact, making AI formal may be the exact opposite direction of what would make AI relevant