The search for the computational correlates of consciousness

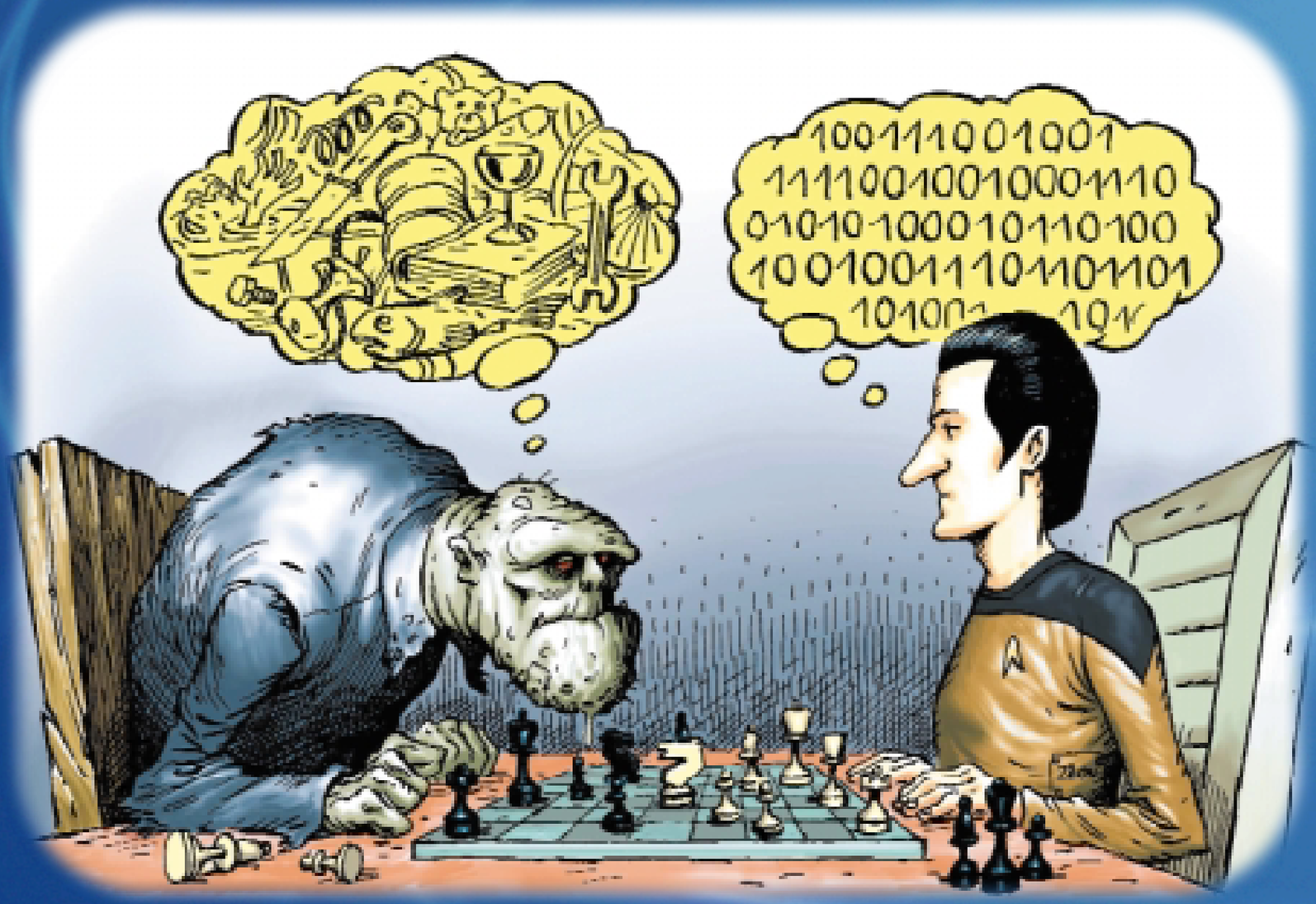
Axel Cleeremans

SRSC

cognitive science research unit

SEMINAIRE DE RECHERCHE EN SCIENCES COGNITIVES
UNIVERSITE LIBRE DE BRUXELLES CP 191
AVENUE F.-D. ROOSEVELT 50
B1050 BRUXELLES

The zombie meets Commander Data



>> Being a zombie

- Two (caricatural) strategies to account for the cognitive unconscious:
  - **Zombie theories** assume total duplication of functions: The cognitive unconscious (the "zombie within") is just like the cognitive conscious, only minus consciousness.
  - **Commander Data** theories assume that mental life is co-extensive with consciousness: Whenever some state is representational, it is also a conscious state
  - Both accounts are rooted in the "classical" notion that cognition consists of symbol manipulation (Problem: Representations are static and remain causally inefficacious unless accessed)

Theories of implicit cognition

- Both perspectives are profoundly unsatisfactory:
  - The notion of a full-blown cognitive unconscious that is just the same as the conscious system only minus consciousness (zombie theories) is as unsatisfactory as the notion that all of cognition involves consciousness (commander data theories): Consciousness is a epiphenomenon in both cases
- The core of the debate is the possibility of unconscious representation
- We need to search for the "computational correlates" of consciousness
  - Identify computational principles that make it possible to characterize the difference between conscious and unconscious states

The search for correlates of C

- **NCC:** "A major part of the programme for studying the neural correlates of consciousness must be to investigate the difference between neural activities that are associated with awareness and those that are not" (Crick & Koch; Frith et al.)

- **BCC:** "A major part of the programme for studying the behavioral correlates of consciousness must be to investigate the difference between behaviors that are associated with awareness and those that are not" (Baars' contrastive approach)

- **CCC:** "A major part of the programme for studying the computational correlates of consciousness must be to investigate the difference between computations that are associated with awareness and those that are not"

Computational correlates of C

- Processes:
  - Resonance (top-down & bottom-up)
  - Synchrony
  - Amplification (through attention)
  - Integration & differentiation
  - Global broadcast

- Representations:
  - Strength (threshold)
  - Stability in time
  - Distinctiveness (explicit representation)
  - Meta-representation (”the front is looking at the back”)

Atkinson, Thomas & Cleeremans, TiCS, 2000

Computational correlates of C

- Convergence towards two aspects:
  - "Quality of representation" as necessary condition for C
  - Meta-representation as sufficient condition for C?

>> Being conscious

- Explore a framework in which:
  - Conscious and unconscious processing are rooted in the same basic mechanisms
  - Consciousness is a graded, continuous, and dynamic process
- Assumptions about:
  - Processing (P1-P4)
  - Representation (R1-R3)
  - Learning (L1-L3)
  - Consciousness (C1-C5)
  - Self (S1-S7)

Assumptions about processing

P1: The cognitive system is best viewed as involving a large set of interconnected processing modules organized in a loose hierarchy. Each module in turn consists of a large number of simple processing units connected together

P2: Long-term knowledge in such systems is embodied in the pattern of connectivity between the processing units of each module and between the modules themselves

P3: Dynamic, transient patterns of activation over the units of each module capture the results of information processing conducted so far

P4: Processing is graded and continuous: Connected modules continuously influence each other's processing in a graded manner that depends on the strength of the connection between them and on the strength of the activation patterns that they contain

Assumptions about representation

R1: Representations consist exclusively of the transient patterns of activation that occur in distributed memory systems

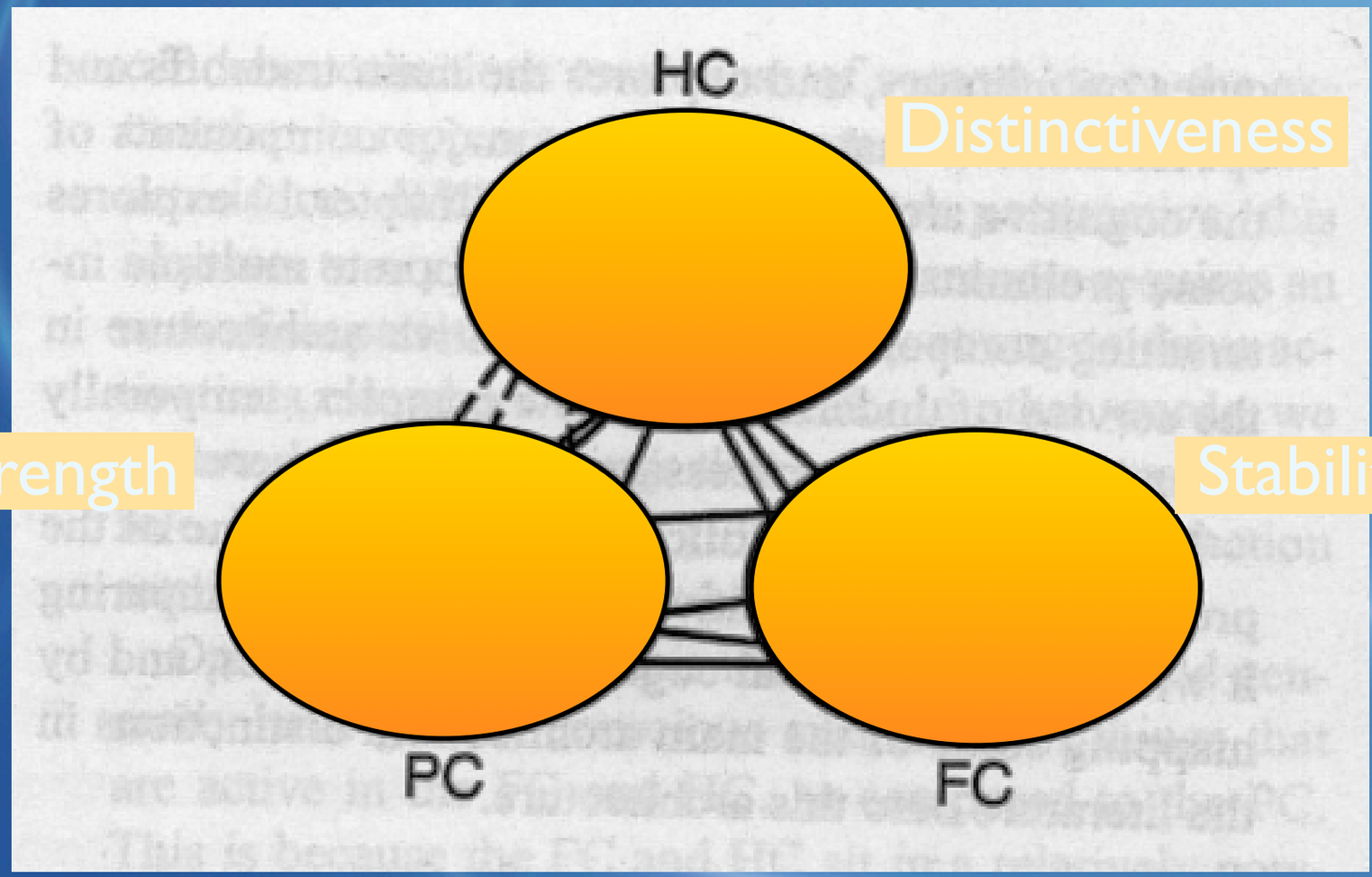
R2: Representations are graded: They vary on several dimensions that include strength, stability in time, and distinctiveness

R3: Representations are dynamic, active, and constantly causally efficacious

Quality of Representation

- **Strength** of representation
  - In connectionist models:
    - Cohen et al. on automaticity
    - Munakata et al. on object permanence
  - In brain imaging studies:
    - Montoussis & Zeki (2002) on invisible stimuli
- **Distinctiveness** of representation
  - Hypothesized as critical in distinguishing FC and HC vs. PC representations (O'Reilly et al.)
- **Stability** of representation
  - As correlate of phenomenal experience in O'Brien & Opie
  - As correlate of consciousness in Mathis & Mozer

Representational systems in the brain



O'Reilly & Munakata, 2000

Assumptions about learning

L1: Adaptation is a mandatory consequence of information processing: We learn all the time

L2: Learning has both direct and indirect effects

Assumptions about consciousness

C1: Consciousness involves three aspects: Access (potency), subjective experience and cognitive control

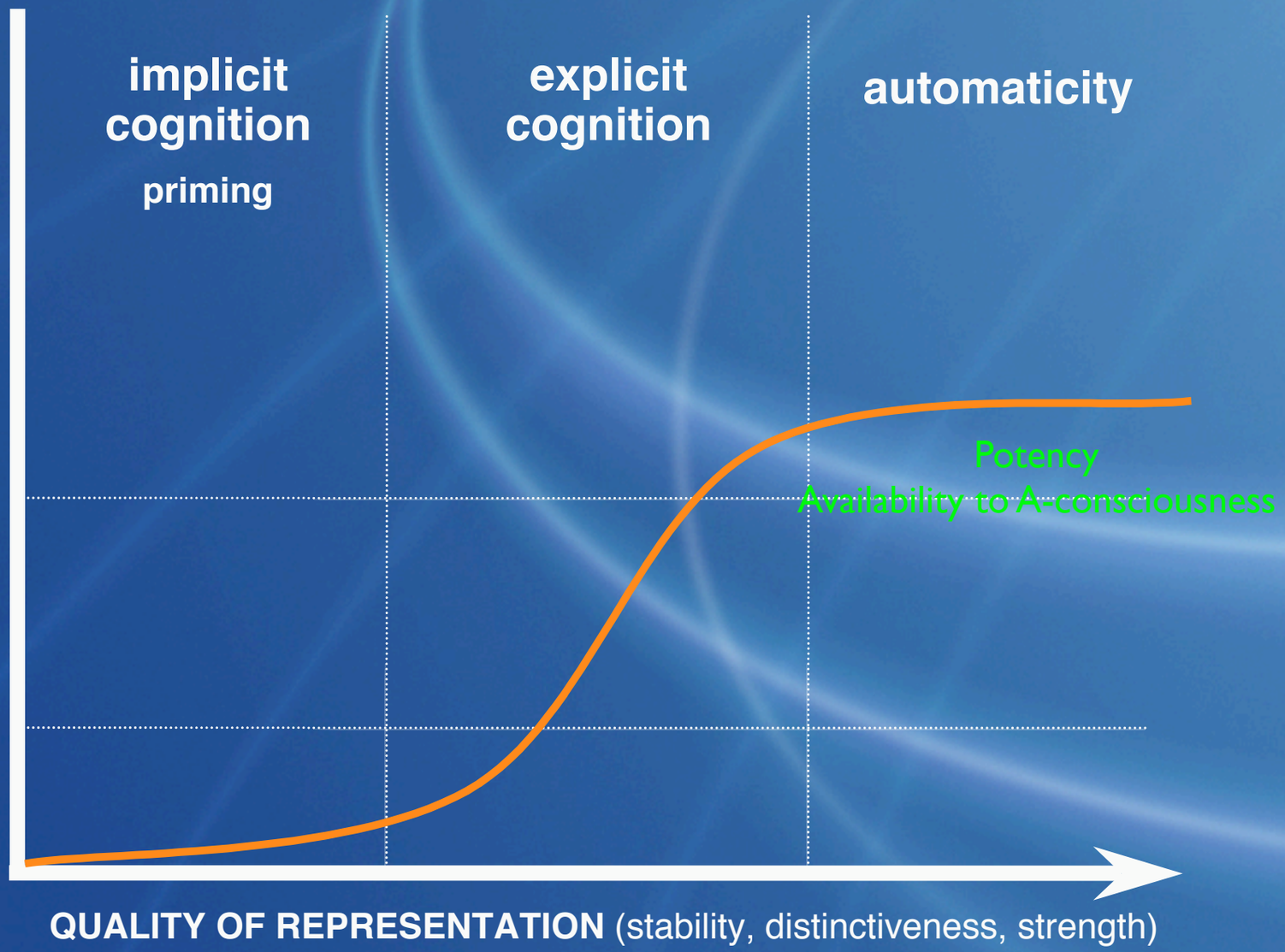
C2: Availability to consciousness correlates with quality of representation

C3: Developing high-quality representations takes time

C4: The function of consciousness is to offer flexible, adaptive control over behavior

C5: Learning shapes conscious experience; conscious experience shapes learning

Processing is graded



- **R2:** Representations are graded: They vary on several dimensions that include strength, stability in time, and distinctiveness

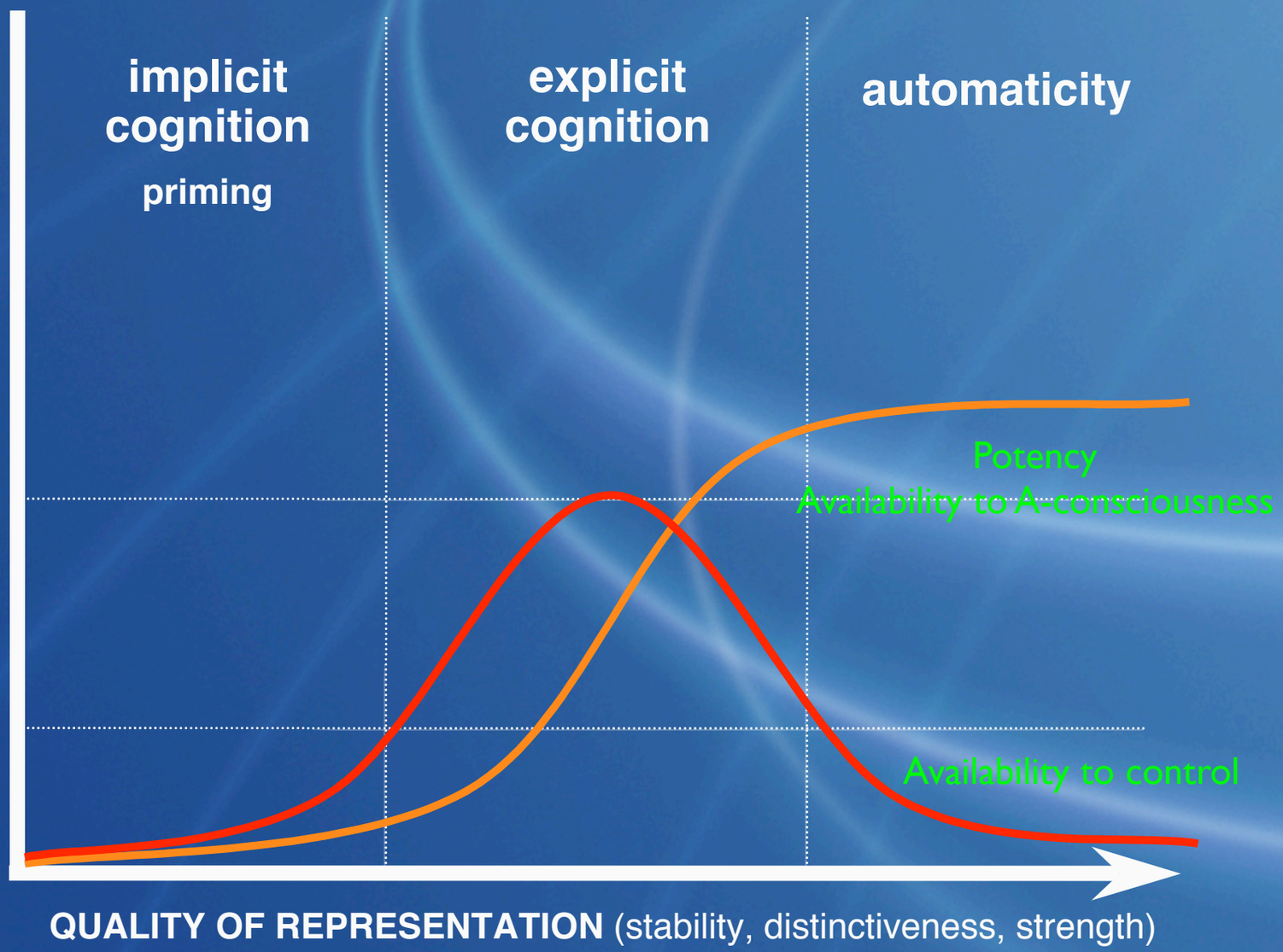
- **R3:** Representations constantly influence information processing regardless of their quality

  Strong representations have more influence on behavior than weaker ones

- **C3:** Developing high-quality representations takes time, both over learning and development, and within a single trial

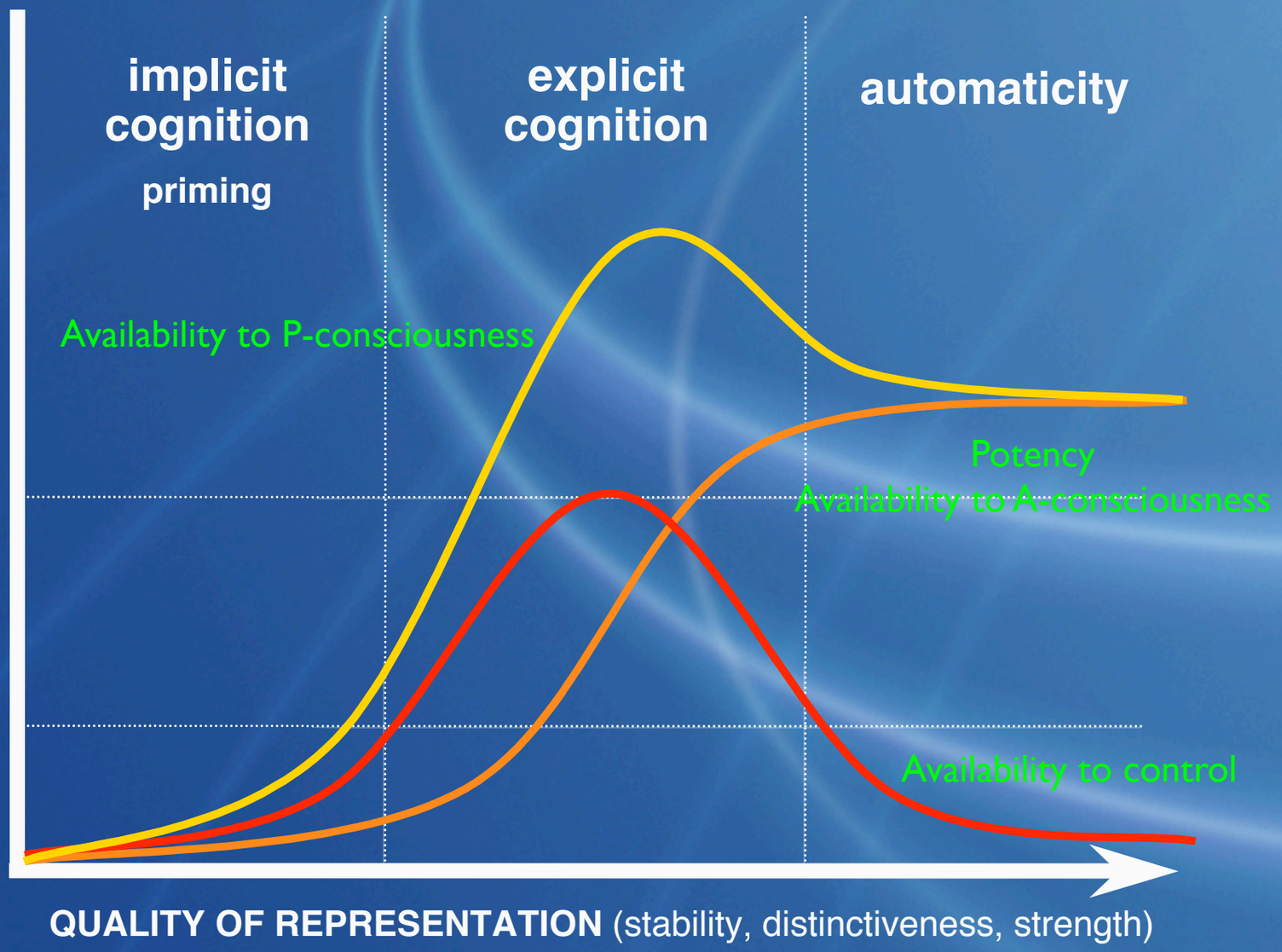
Cleeremans & Jiménez, 2002

Control and the function of C



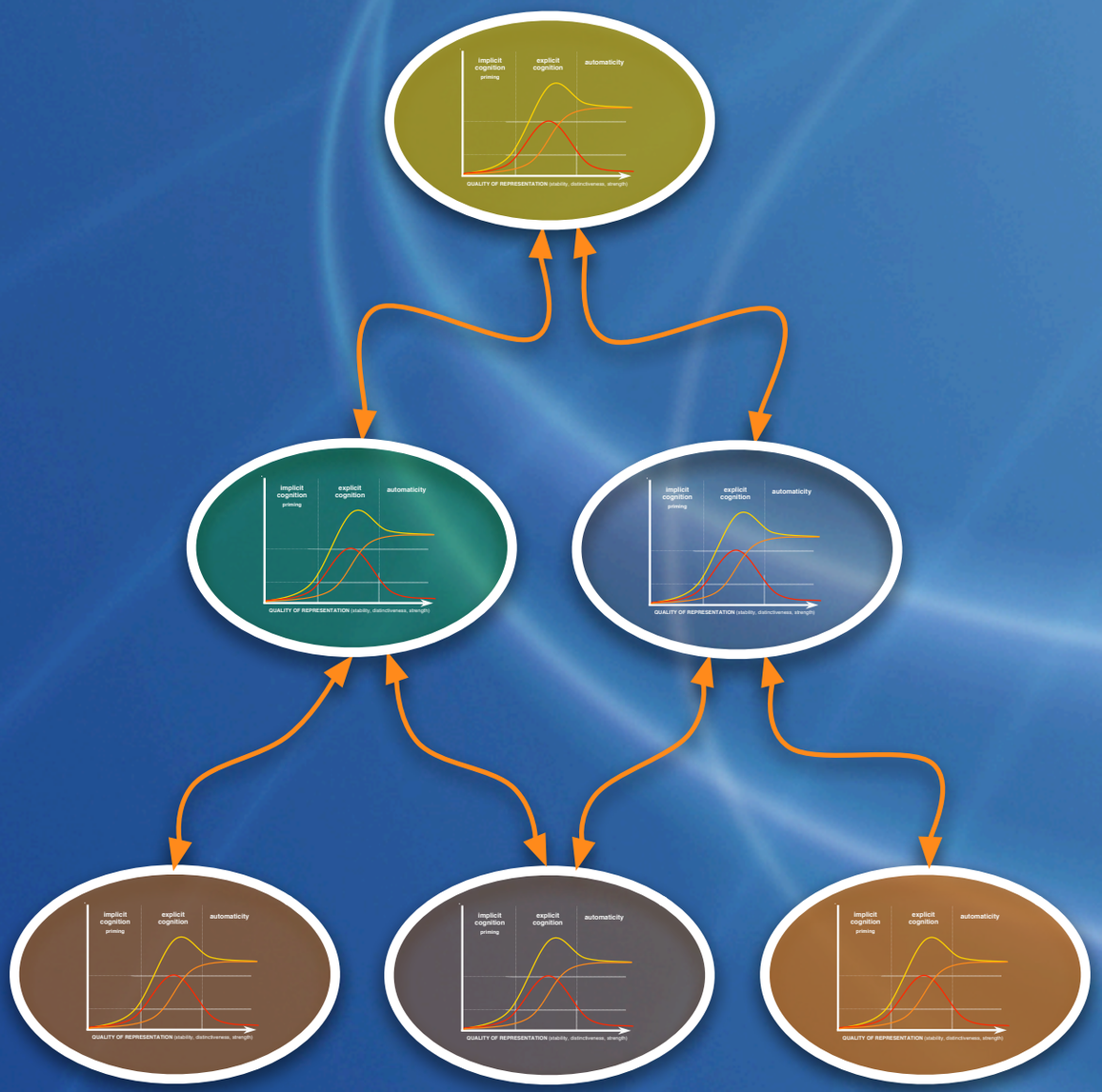
- Weak ("implicit") representations do not need control because they only exert weak effects on behavior (i.e., through priming)
- Strong ("automatic") representations do not need control because they are adapted (**L1:** Adaptation is a mandatory consequence of information processing)
- **C4:** The function of C is to offer flexible, adaptive control over behavior (over those representations that need most control because they drive behavior)

The contents of C



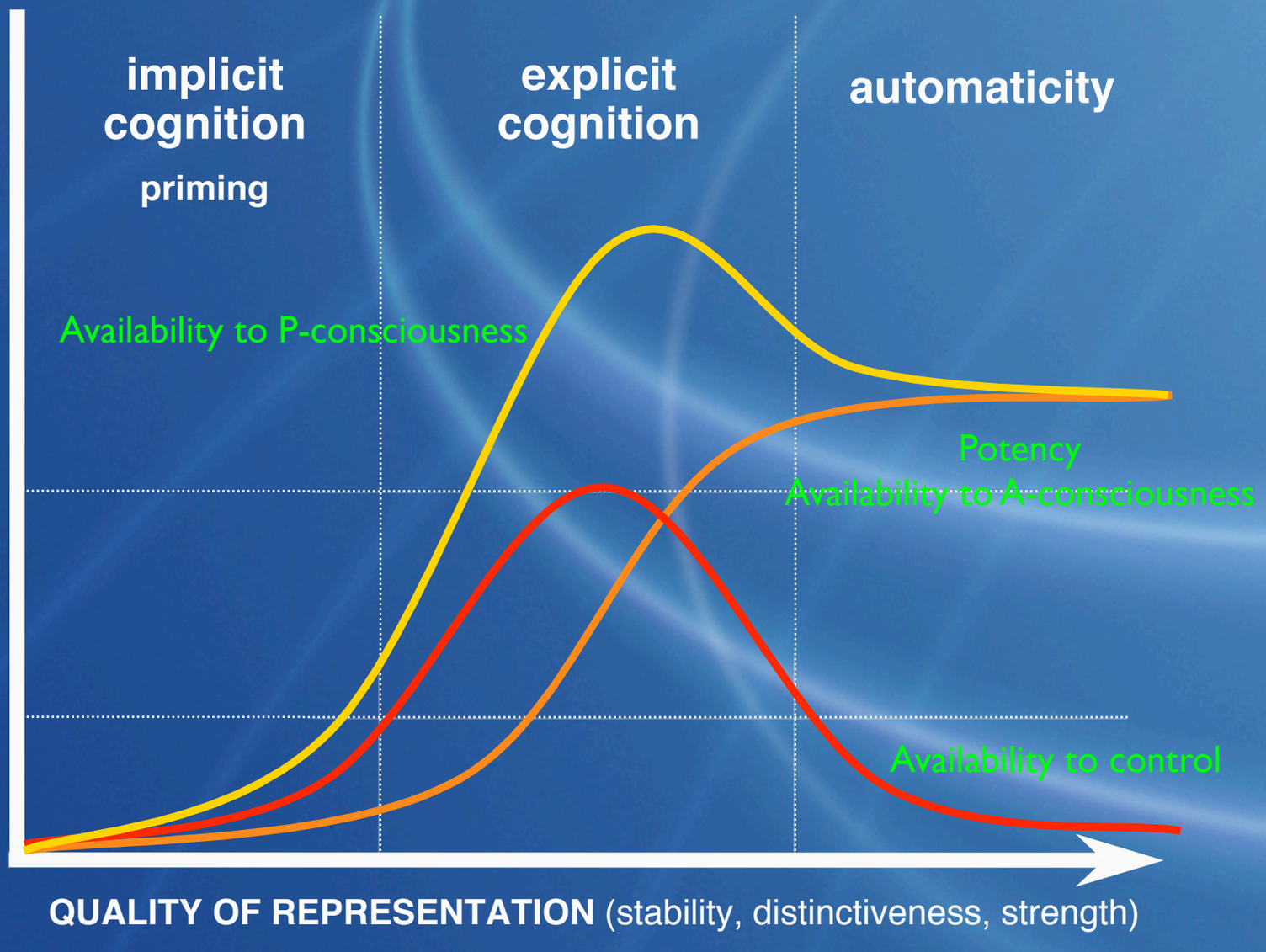
- Availability to phenomenal consciousness depends on both potency & availability to control
- "Implicit" representations are not available to conscious experience
- "Explicit" representations constitute the dominant focus of consciousness
- "Automatic" representations constitute the periphery of consciousness (the fringe)
- **C2:** Availability to C correlates with quality of representation
- **C5:** Learning shapes conscious experience; conscious experiences shapes learning

A cognitive hierarchy



- The brain's functional and anatomical organization involves many interconnected networks sensitive to increasingly abstract dimensions of the stimulus (from PC to HC and FC)
- Skill acquisition and learning involves both moving up in this hierarchy ("the higher levels first") as well as changes within the modules: how GW emerges
- The dominant contents of consciousness consist of the networks in which most change is currently taking place (those that require the most control)

Ways to be (un)conscious

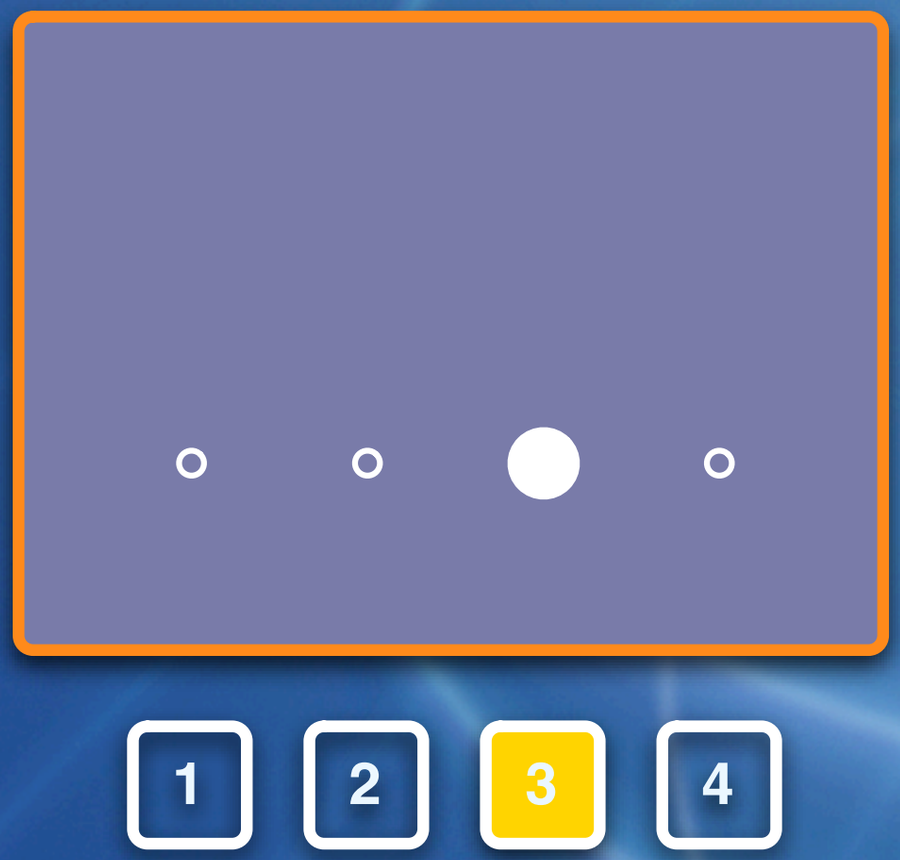


- Weak representations are involved: unconscious learning and priming; implicit knowledge as knowledge without consciousness
- Stronger, conscious representations are not accompanied by relevant metaknowledge: implicit learning as the indirect effects of explicit learning; implicit knowledge as conscious knowledge without metaknowledge
- "Automatic" representations can not be controlled: automatic uses of memory; implicit knowledge as conscious knowledge without control

Ways to measure (un)consciousness

Consciousness has different aspects:	Consciousness can be measured through different tasks:
Consciousness allows verbal access to the acquired knowledge	Verbal reports, questionnaires
Consciousness allows recollection of the acquired knowledge	Recognition & generation tasks
Consciousness allows control on the relevant knowledge	Inclusion & exclusion tasks
Consciousness is associated with metaknowledge	Confidence judgment tasks

The sequence learning paradigm

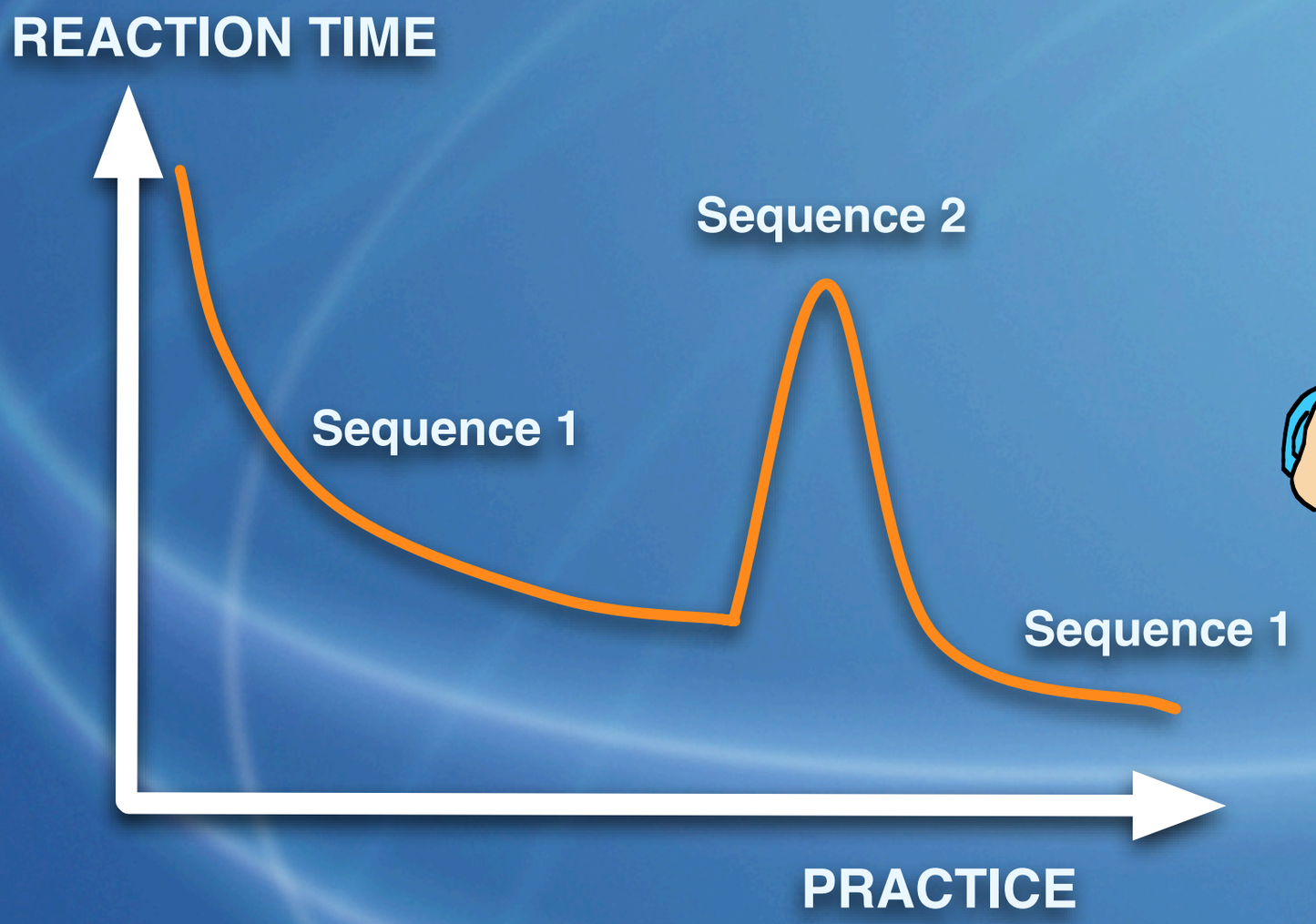


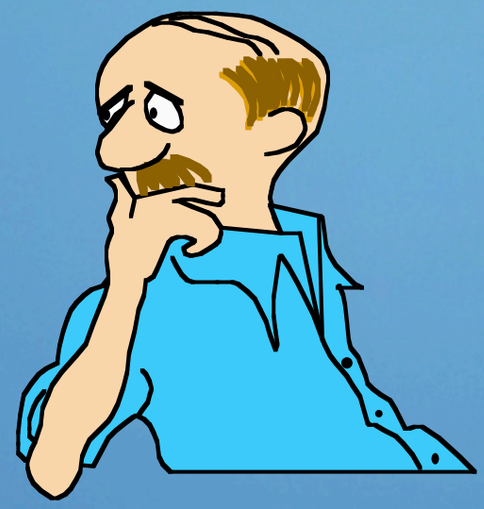
- Task is choice reaction
- Unknown to subjects, stimuli follow a repeating sequence
- Learning is assessed by switching to a different sequence during an unannounced transfer block

342312143241 342312143241 … **(training)**

341243142132 341243142132 … **(transfer)**

Typical Results





IMPLICIT LEARNING:

A change in performance that is not accompanied by a corresponding change in the ability to describe the acquired knowledge

The Process Dissociation Procedure
(Jacoby, 1991)

- Any task always involves both implicit and explicit components
- After training on the serial reaction time task, participants perform two direct tests that differ with respect to the instructions:
-

➡ **the inclusion condition:**

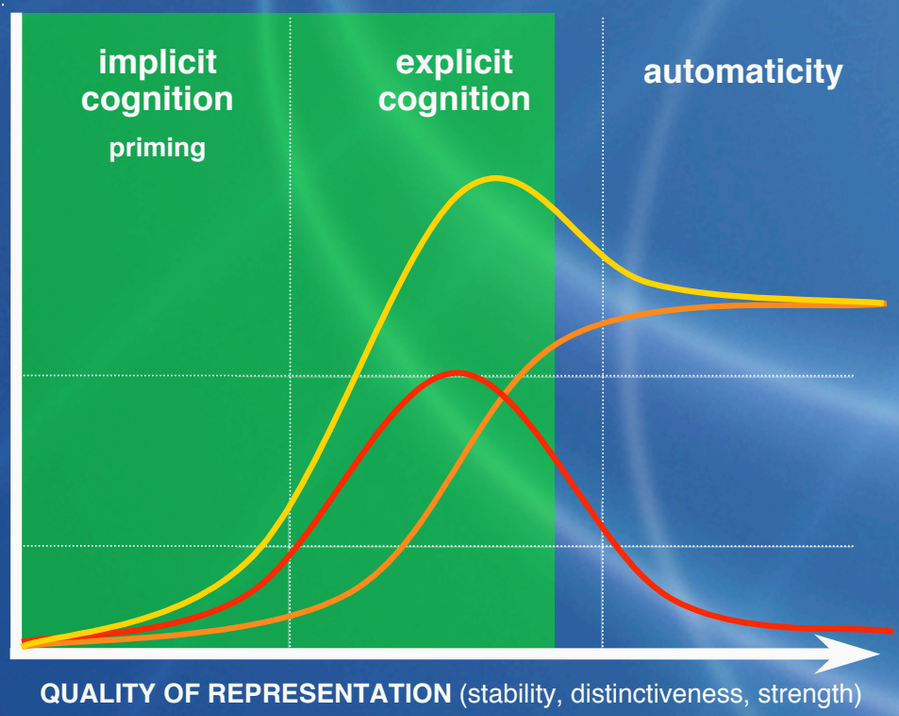
- Participants are asked to recollect and reproduce the training sequence. If they cannot recollect the location of a stimulus, they are told to use their intuition and to guess
- Explicit and implicit influences can both contribute to performance improvement

➡ **the exclusion condition:**

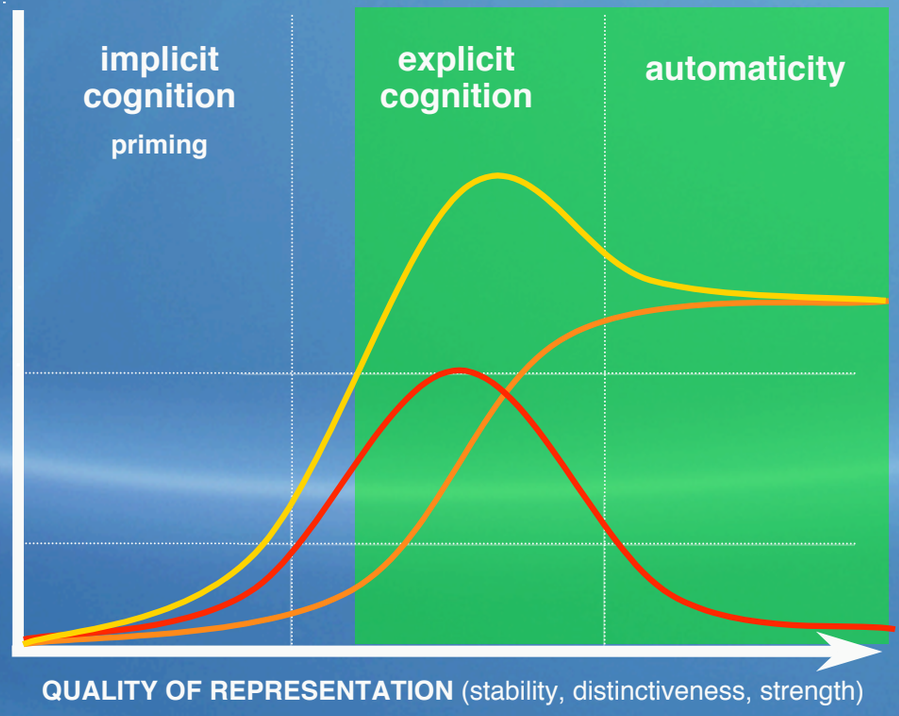
- Participants are told to generate a sequence of stimuli that differs from the training sequence: They must try to avoid reproducing the training sequence
- Explicit and implicit influences are set in opposition

>> Testing the framework

- The framework predicts increased availability to consciousness with stronger representations, at different time scales: time course within a single trial, training & development



Time course within a single trial



Training

From implicit to explicit :
Associations between availability to consciousness and availability to control

From explicit to automatic:
Dissociations between availability to consciousness and availability to control

>> Testing the framework: Temporal effects

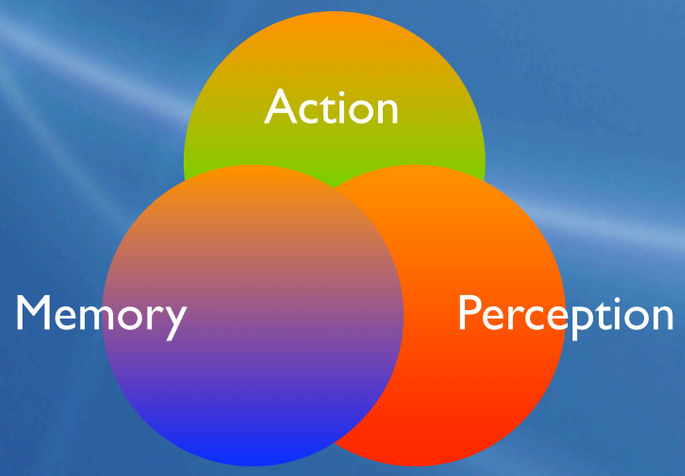
- In a first series of experiments, we manipulated the extent to which learning is implicit or explicit by varying the response-stimulus interval (RSI)
  - Preparation for the next event in choice reaction time tasks involve both (unconscious) priming and conscious preparation
  - Reducing the RSI to zero might prevent the development of conscious expectations about the next stimulus, and hence selectively impair explicit sequence learning (see also Squire et al. on conditioning)
  - Increasing the RSI might promote the development of strong, conscious representations

Arnaud Destrebecqz

Destrebecqz & Cleeremans, PBR, 2001
Destrebecqz & Cleeremans, 2003

Time course of a single trial

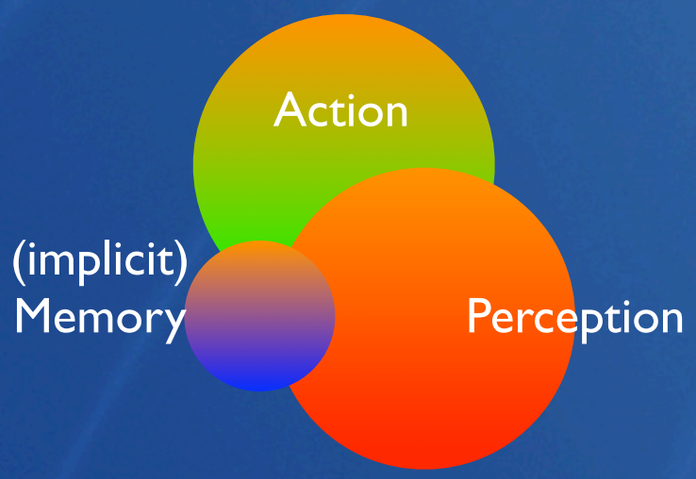
RSI = 250ms
Standard condition



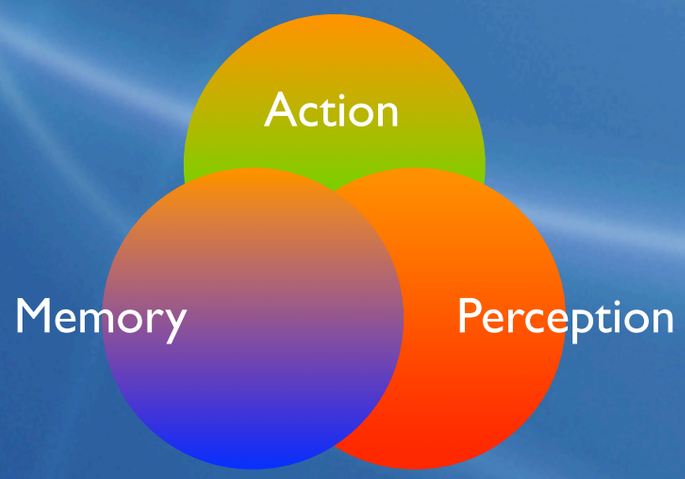
RSI 250

Time course of a single trial

RSI = 0ms
"No RSI" condition



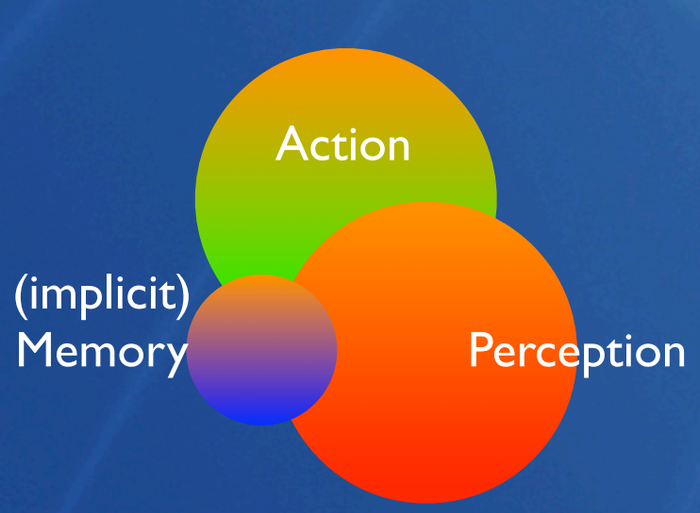
RSI 0



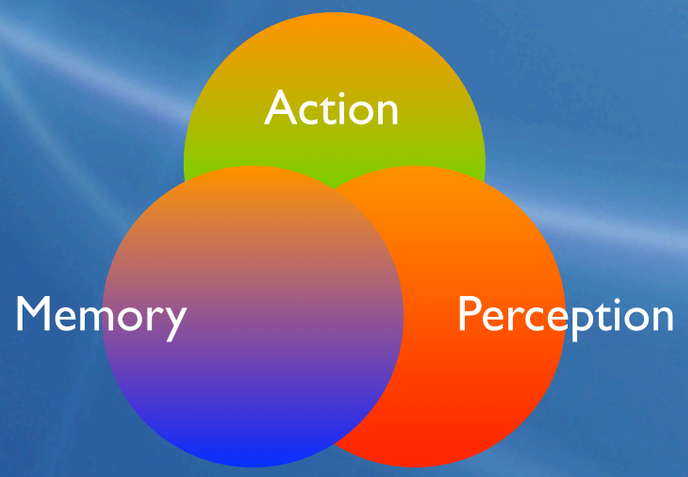
RSI 250

Time course of a single trial

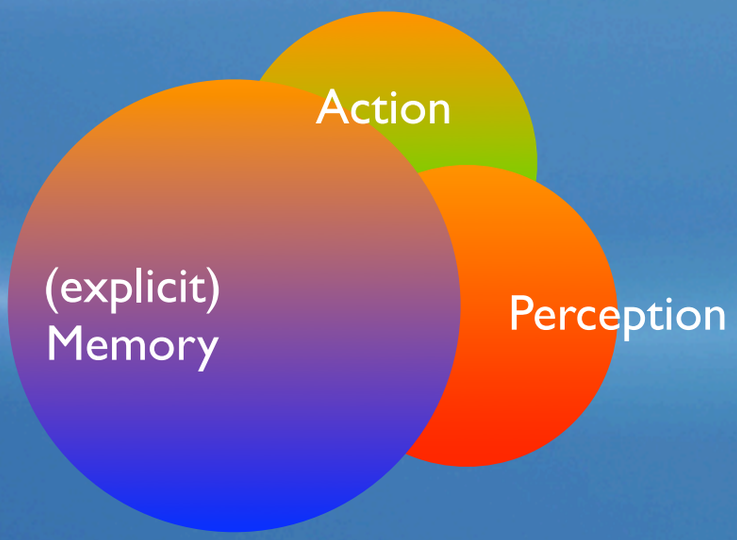
RSI = 1500ms
"Long" condition



RSI 0

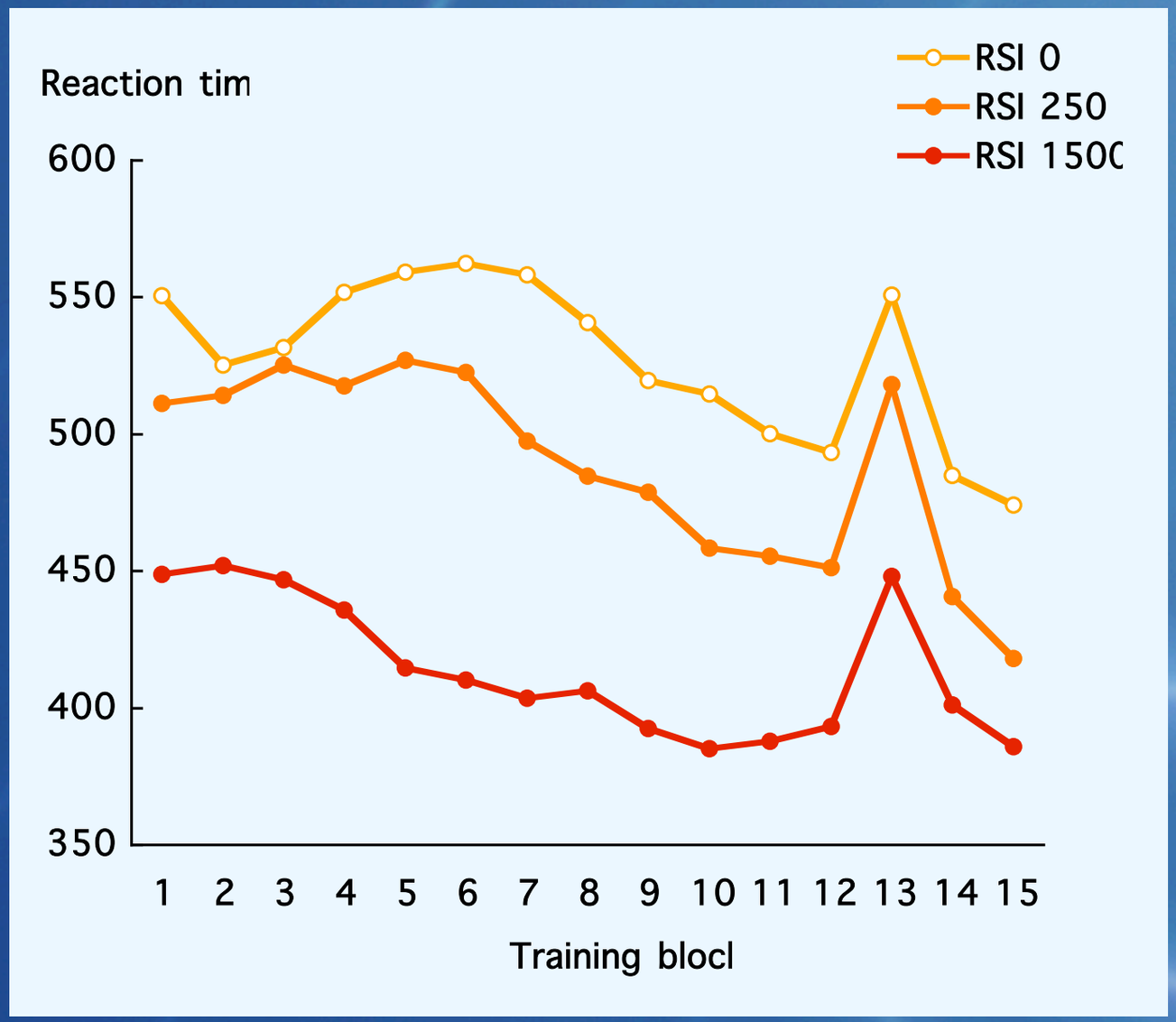


RSI 250



RSI 1500

Reaction time results



- The same 12-element sequence is presented on blocks 1-12 and 14-15
- A different 12-element sequence is presented on Block 13
- Higher values of RSI are associated with faster reaction times
- Subjects learn in all three conditions

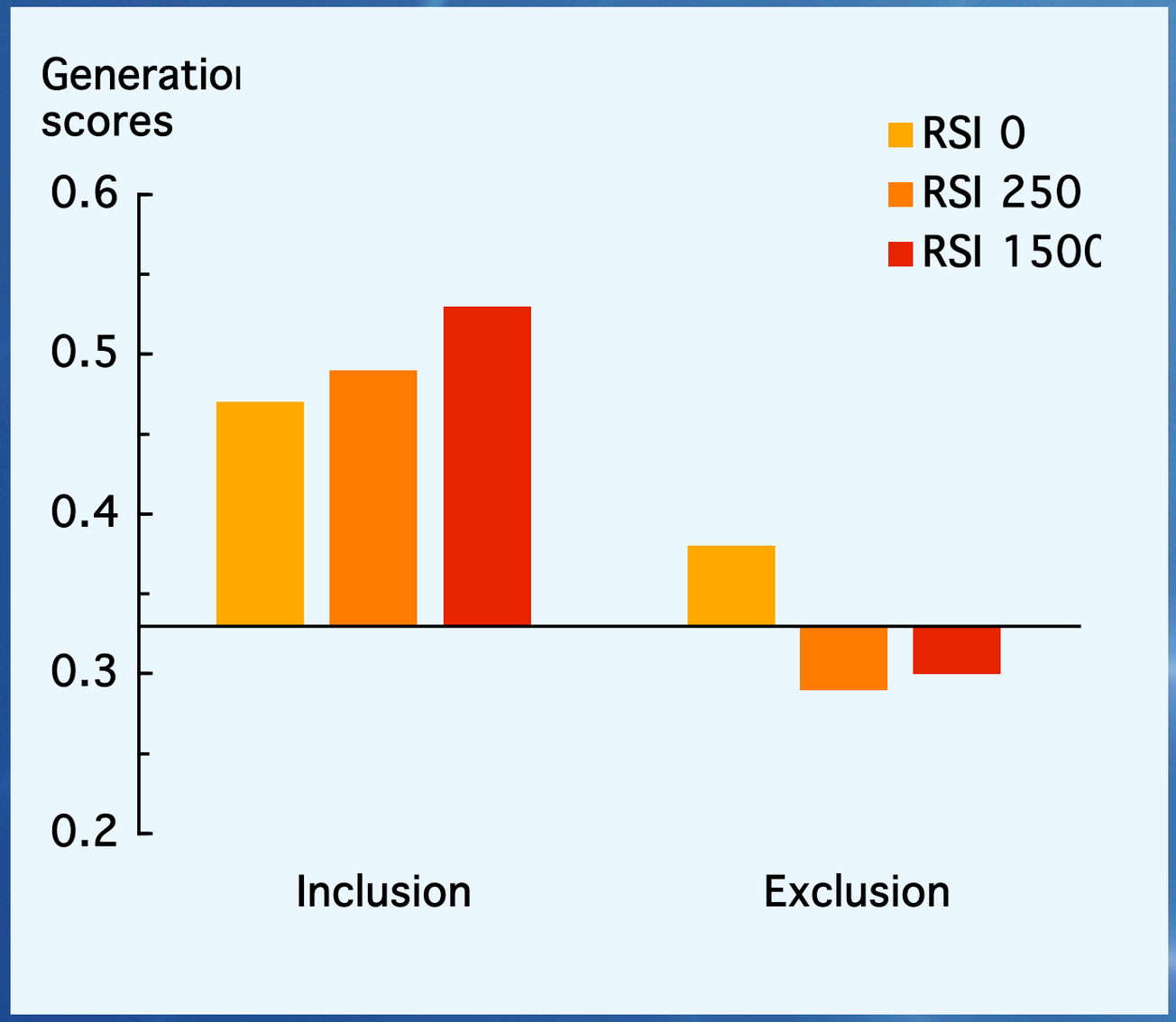
Questionnaire results

- After the SRT task, subjects were presented with the following 5 propositions:

  1. The sequence of stimuli was random
  2. Some positions occurred more often than others
  3. The movement was often predictable
  4. The same sequence of movements would often appear
  5. The same sequence of movements occurred throughout the experiment
  5.

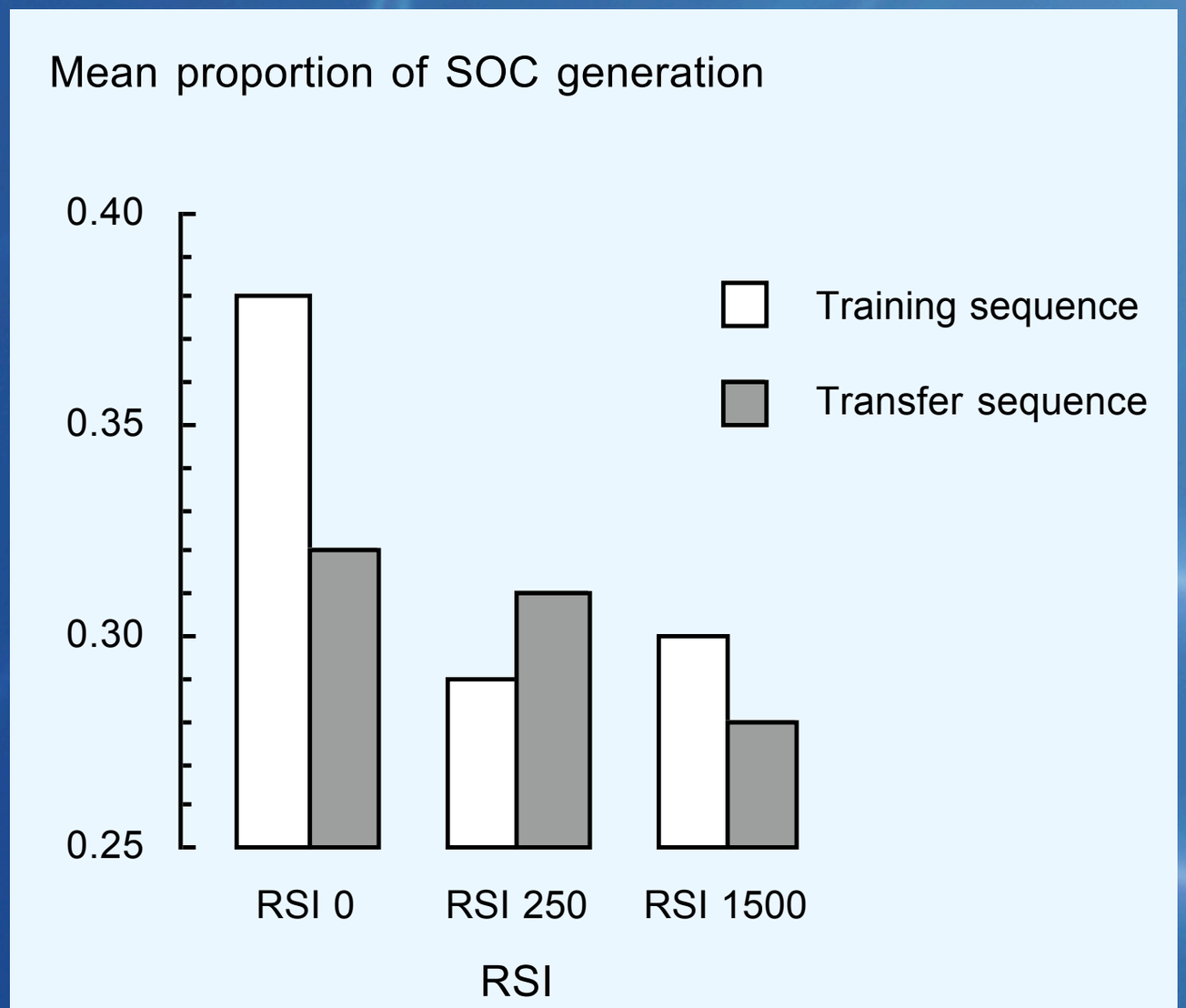
- The five propositions (1 to 5) describe increasing degrees of sequential structure
- Most subjects consider that "the same sequence of movements would often appear"
- There is no difference between conditions

Generation results



- Generation scores represent the proportion of generated triplets that are part of the training sequence
- Inclusion scores do not differ from each other and are significantly above chance level in all three conditions
- Exclusion scores are above chance level in the RSI 0 condition only
- Inclusion scores are higher than exclusion scores but this difference is only marginally significant in the RSI 0 condition

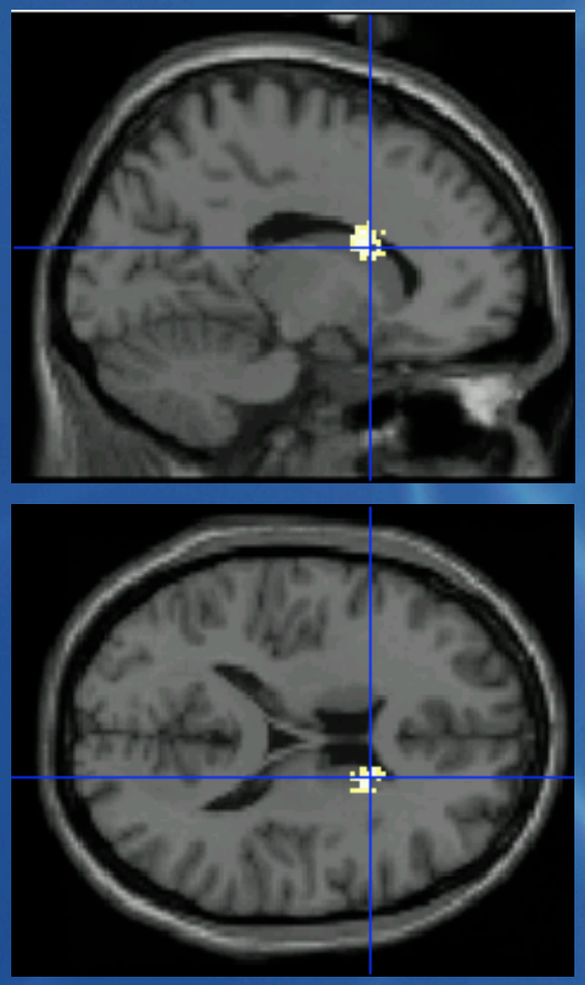
Exclusion results

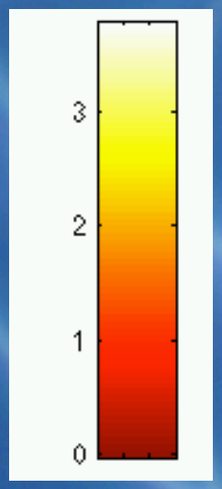


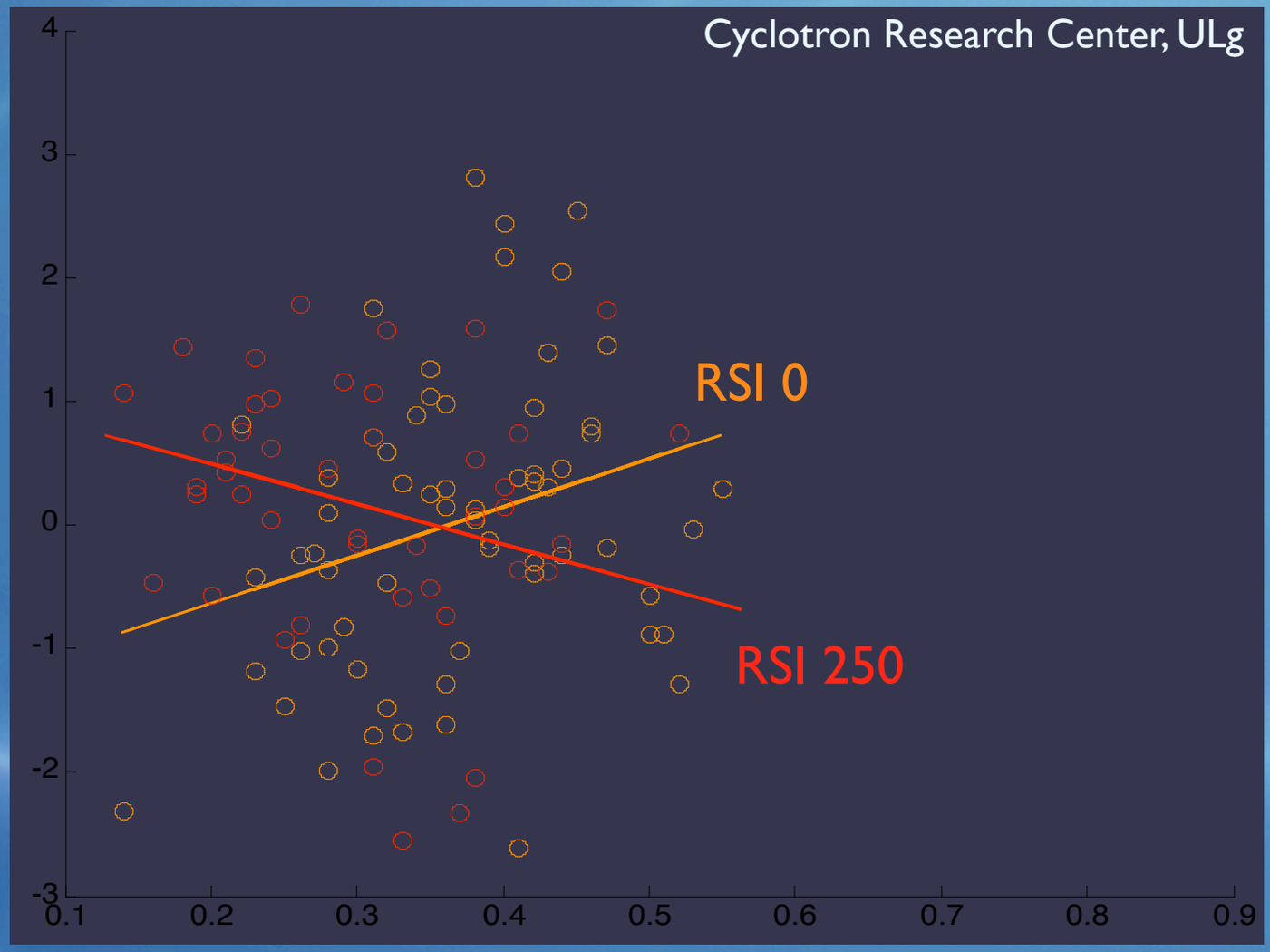
- RSI 0 participants produced more training than transfer triplets in the exclusion task

➡ Weak traces are difficult to control because "you don't know you have them"

Imaging implicit knowledge

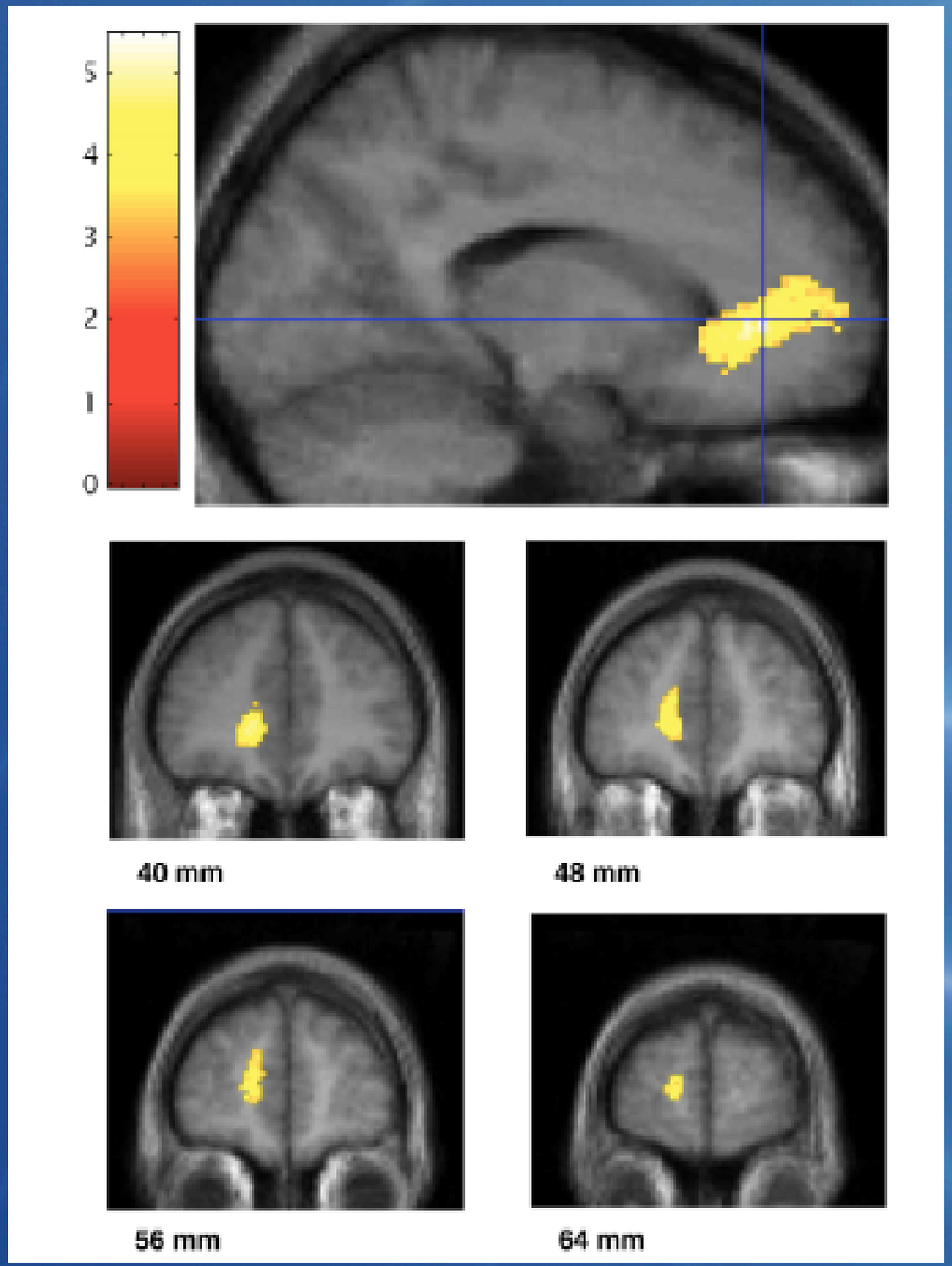






- rCBF in the right caudate nucleus correlates significantly more with the generation score obtained when exclusion is performed after training with RSI=0 than after training with RSI = 250
- Striatum is specifically involved in implicit sequence learning
- Confirms that exclusion scores depend essentially on implicit knowledge

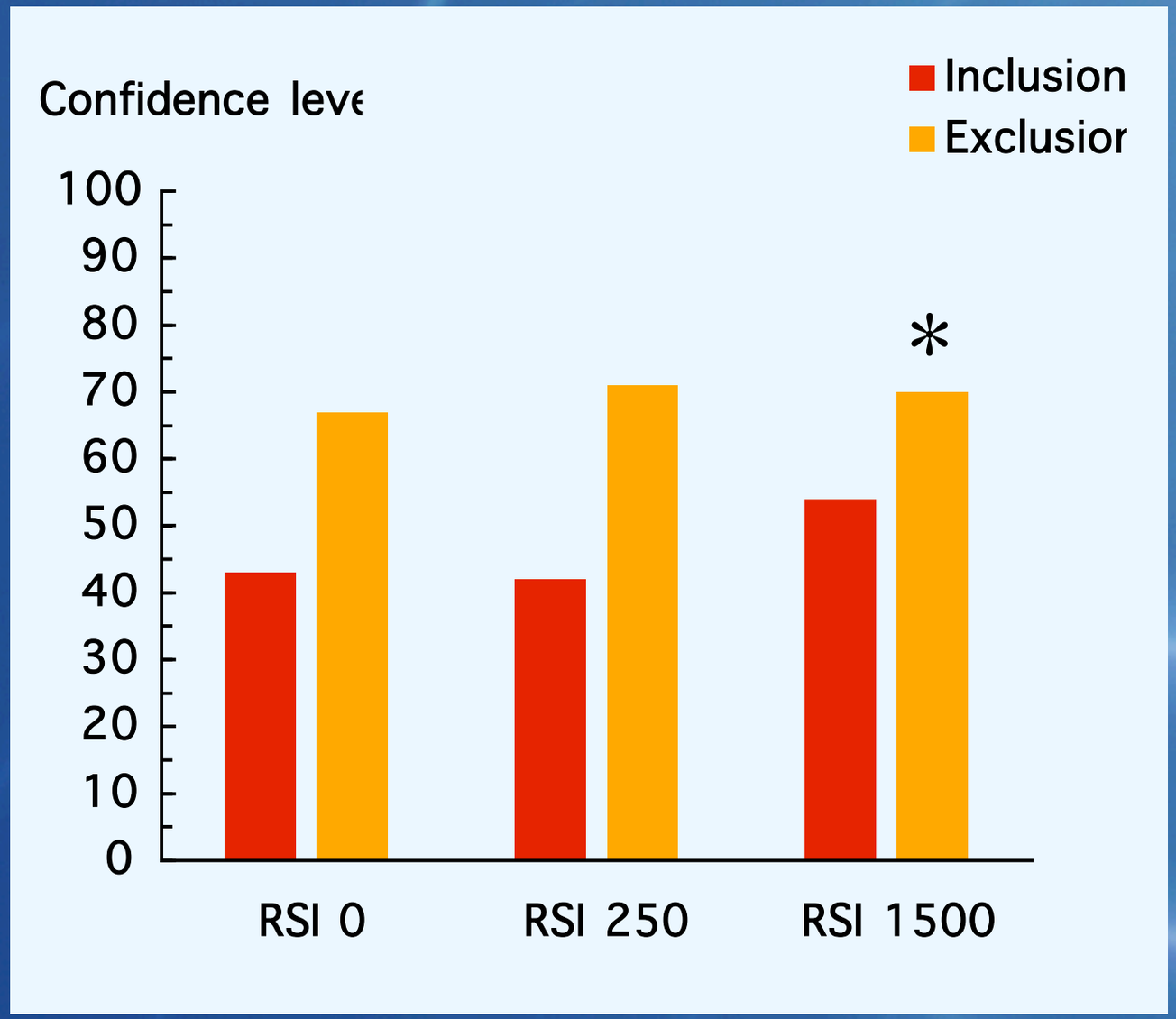
Imaging explicit knowledge



- Interaction between condition (inclusion vs. exclusion) and generation score after training with RSI = 250ms
- Inclusion = C + U
- Exclusion = U
- I - E = C
- ACC/MPFC (BA 32/10) rCBF correlates specifically with C

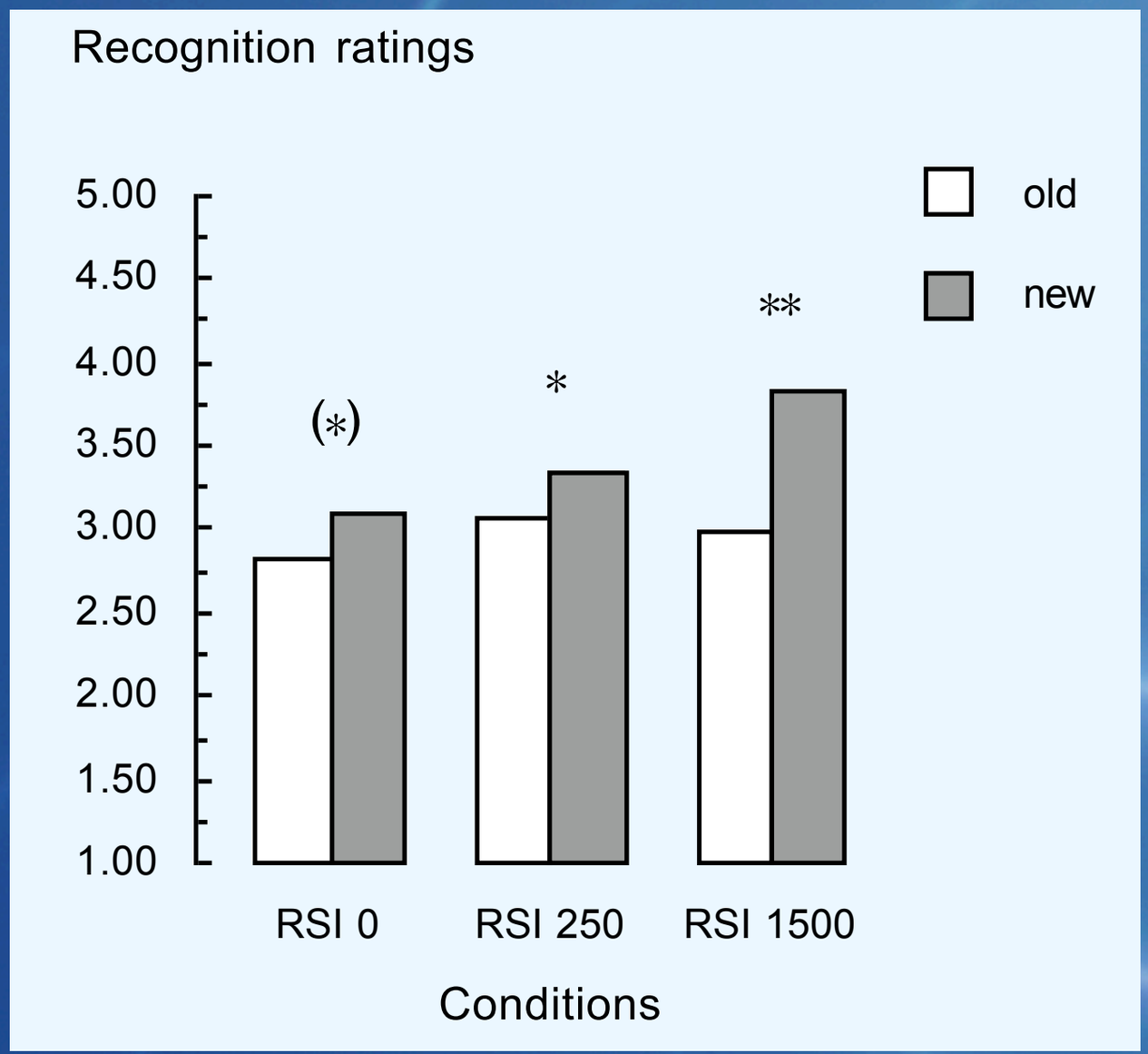
Destrebecqz et al., CBR, 2003

Confidence judgments results



- After each generation task, participants had to rate how confident they were in their performance on a scale ranging from 0 to 100
- Participants are more confident in their exclusion than in their inclusion performance
- There is no effect of the RSI on confidence ratings
- Significant correlation between inclusion performance and confidence judgments only in the RSI 1500 condition: metaknowledge

Recognition results

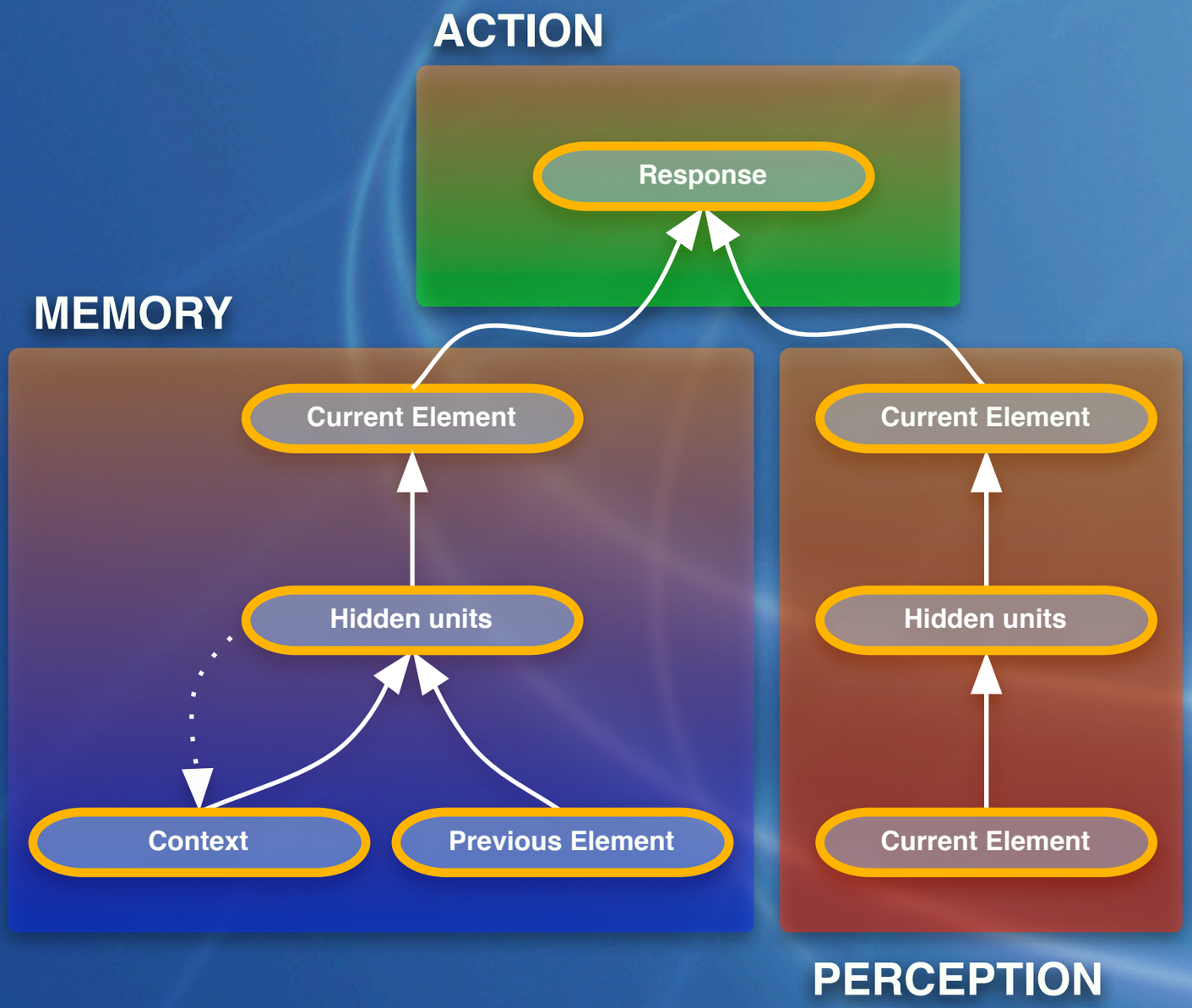


- Subjects were presented with 24 sequences of three elements. Half of these triplets were part of the training sequence and the other half were part of the transfer sequence
- They were asked to rate (between 1 and 6) the extent to which they felt the triplet was part of the training sequence or not
- Ratings differ between old and new triplets but this difference is only marginally significant in the RSI 0 condition (it is not significant in another replication)
- The mean difference between old and new triplets is higher in the RSI 1500 condition than in the RSI 250 condition (ps < .05)

Summary

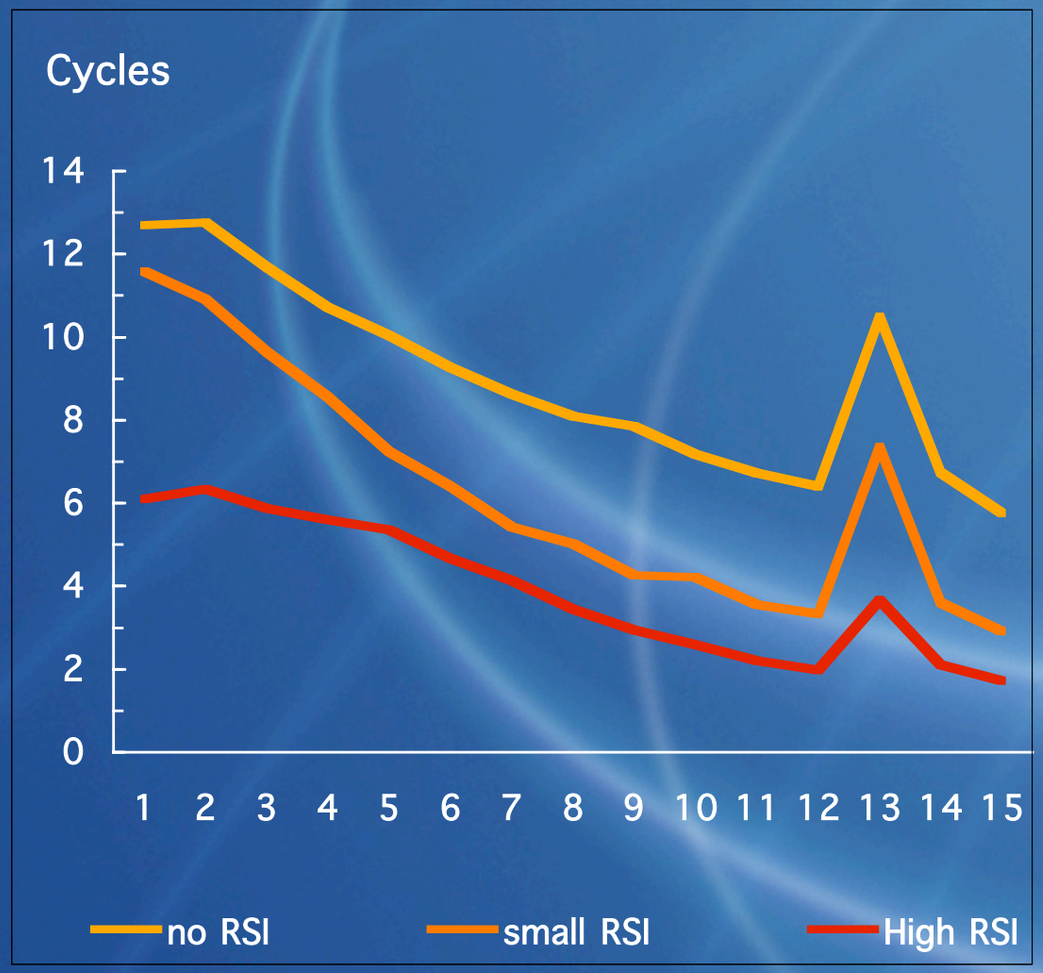
- Time available to process each trial in the training task results in stronger representations of the sequential material:
  - Subjects trained with an RSI=0 lack control over their knowledge and do not differentiate between old and new sequence fragments
  - Subjects trained with an RSI=250 acquired explicit sequence knowledge that they could both control in the exclusion task as well as recollect in the recognition task
  - Increasing the RSI to 1500 ms further allowed subjects to acquire meta-knowledge, as suggested by the pattern of correlations between inclusion performance and confidence judgments
- The acquired knowledge is associated with different (graded) qualitative behaviors — from weak, implicit influence to controlled responding and meta-knowledge

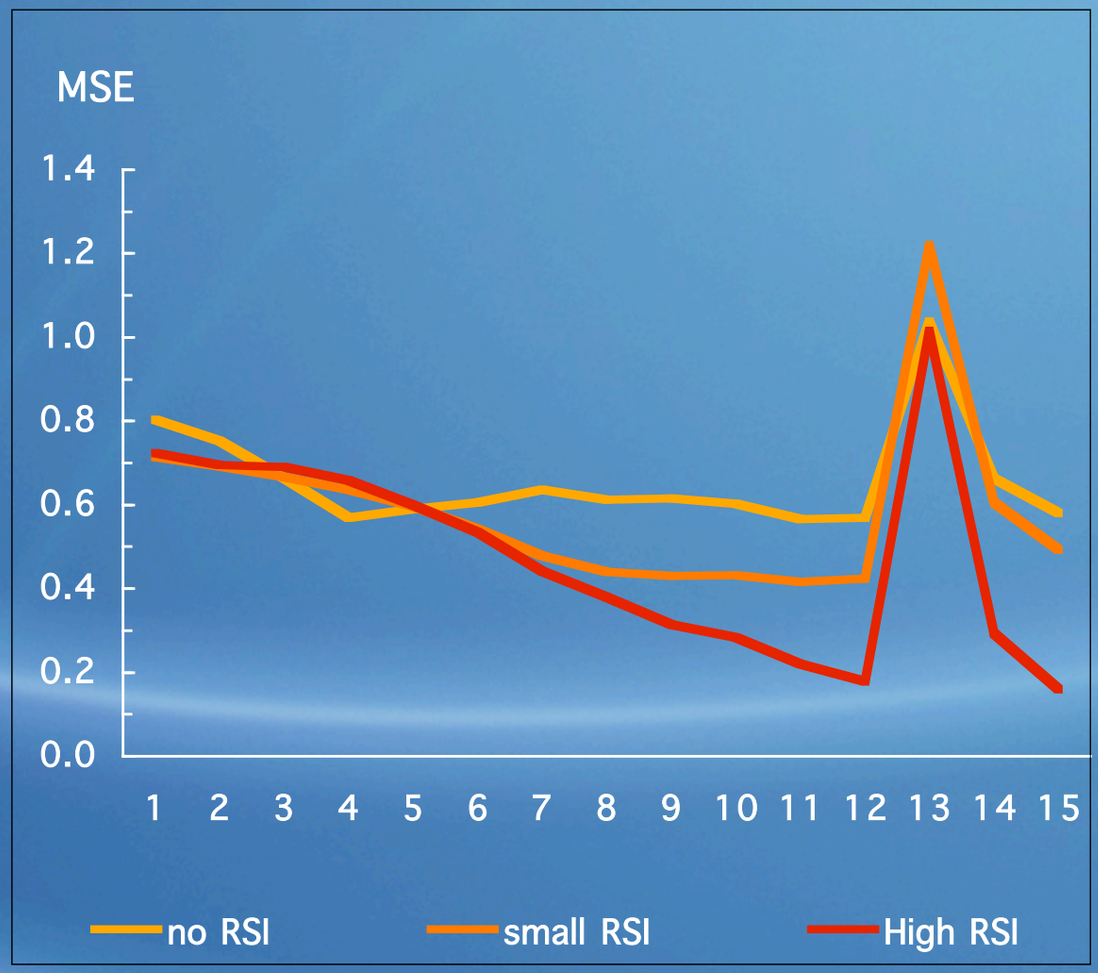
Simulations



- Three components:
  - **MEMORY:** an SRN trained to predict stimulus t based on stimulus t-1 and on a representation of the temporal context that the network has developed itself
  - **PERCEPTION:** an auto-associator that identifies the current stimulus (t)
  - **ACTION:** a set of response units that integrate SRN and auto-associator outputs through a series of fixed "one-to-one" connections
- Processing is **cascaded**

Simulation of the SRT task



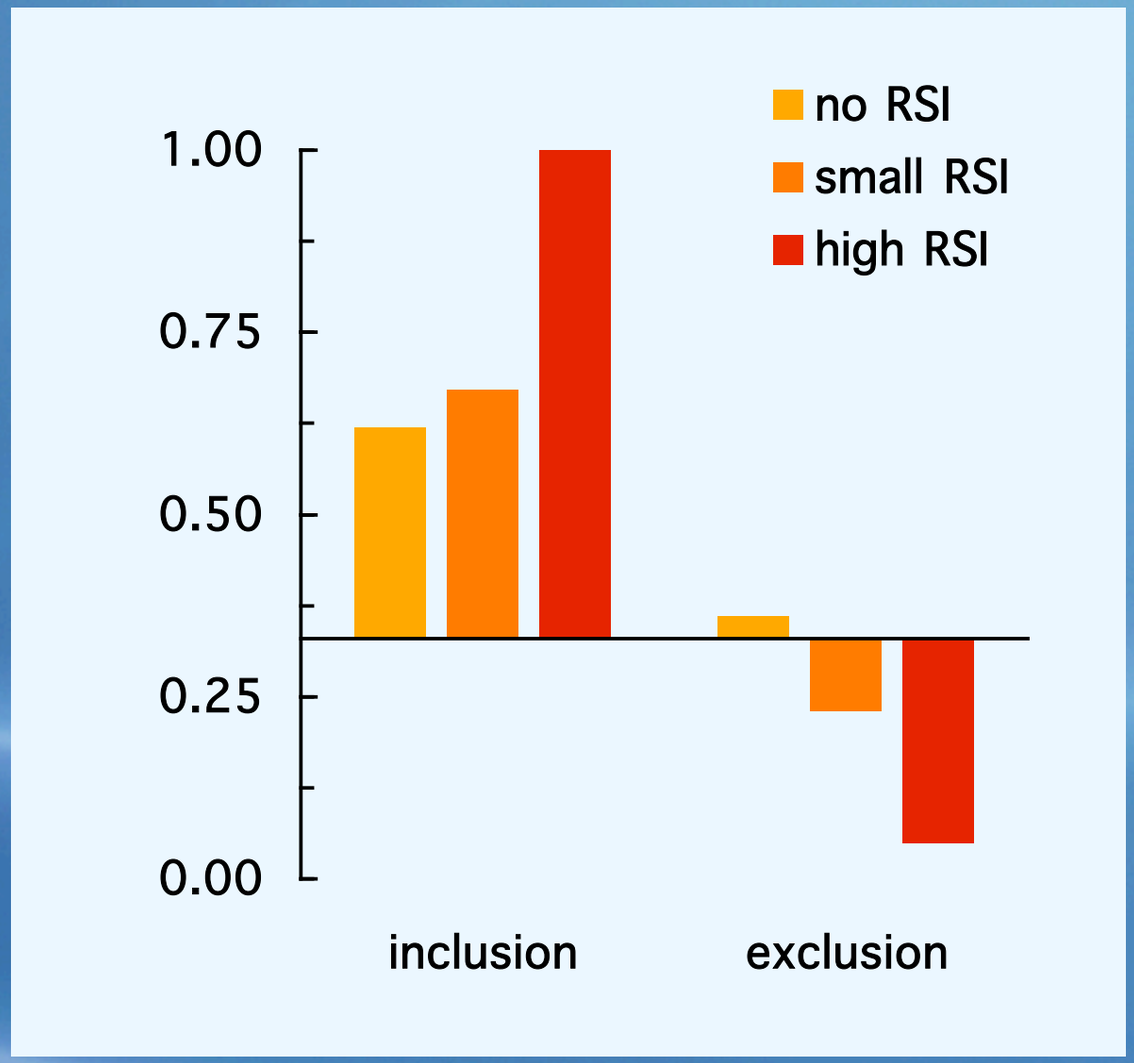


The network captures the effects of the RSI on choice reaction performance

Simulations of generation

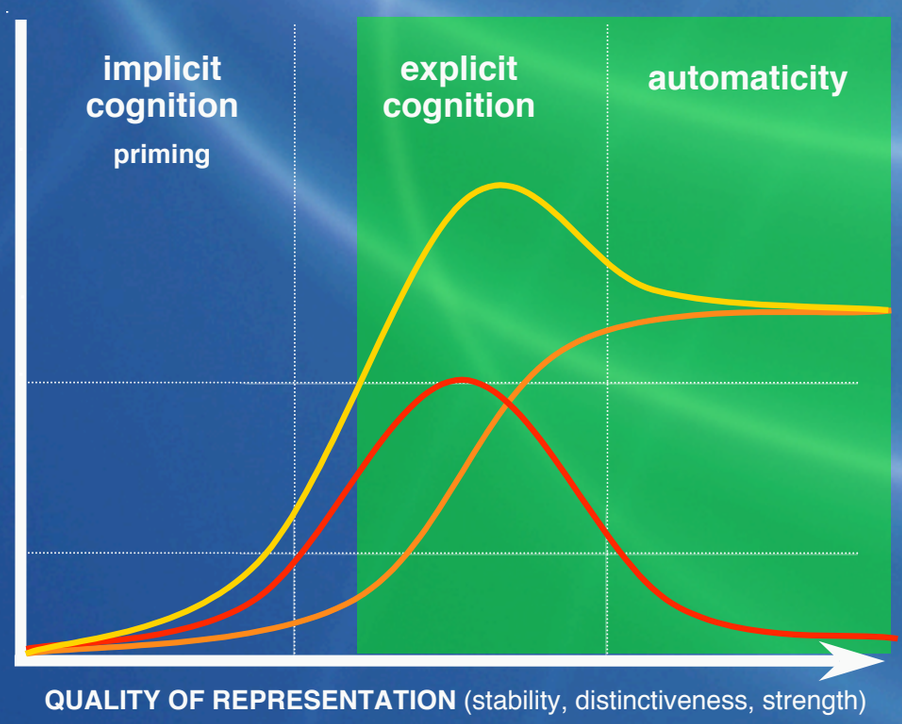
- Only the memory component of the model is used to simulate generation
- SRN predictions are taken to correspond to generation responses
- On each trial, a response is selected based on activation levels and presented as the next stimulus to the SRN

- Response selection differs in inclusion vs. exclusion:
  - In inclusion, the next input corresponds to the most activated output unit at the previous trial
  - In exclusion, that particular response is excluded and the next stimulus is randomly chosen between the other possible responses

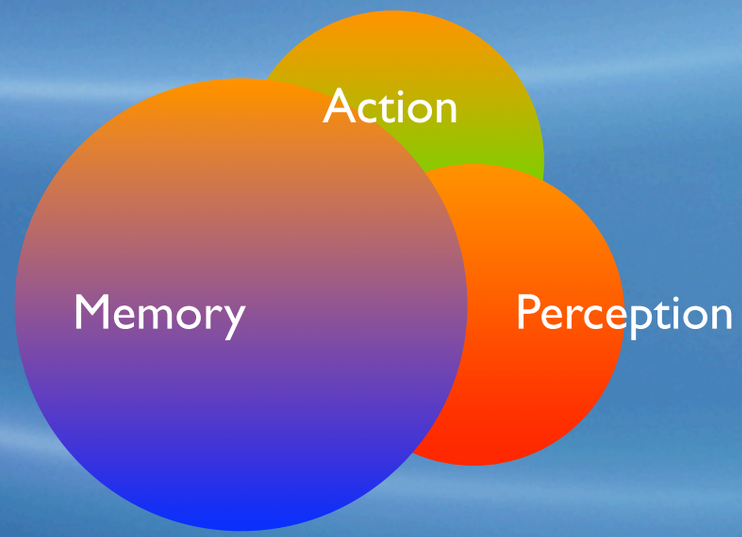


>> Testing the framework: Practice

- What are the effects of extended practice?
  - The framework predicts that extended practice will result in "automatic representations" that are simultaneously more available to both action and subjective experience, but also harder to control
  - Subjects were trained for 30 blocks of trials

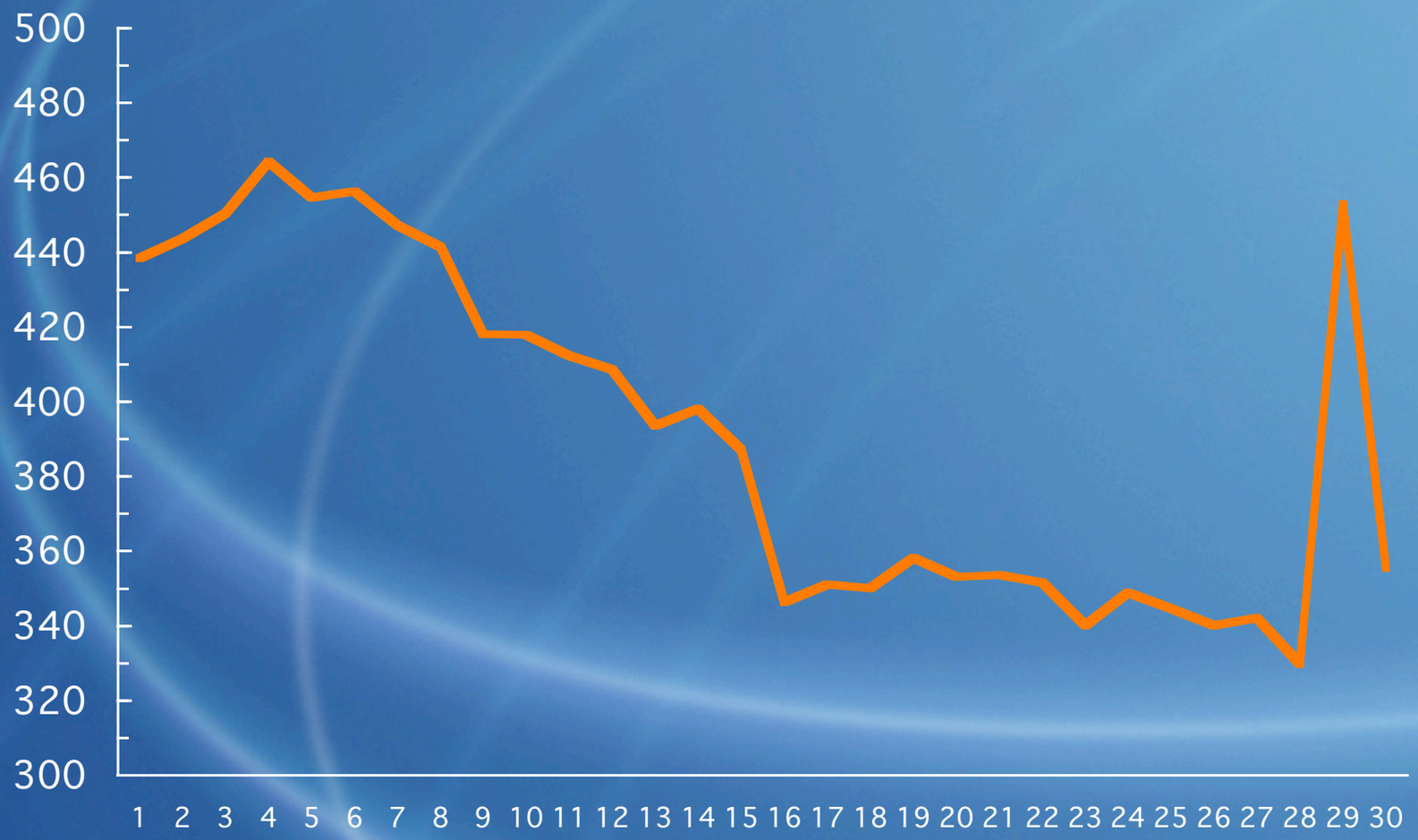


Training



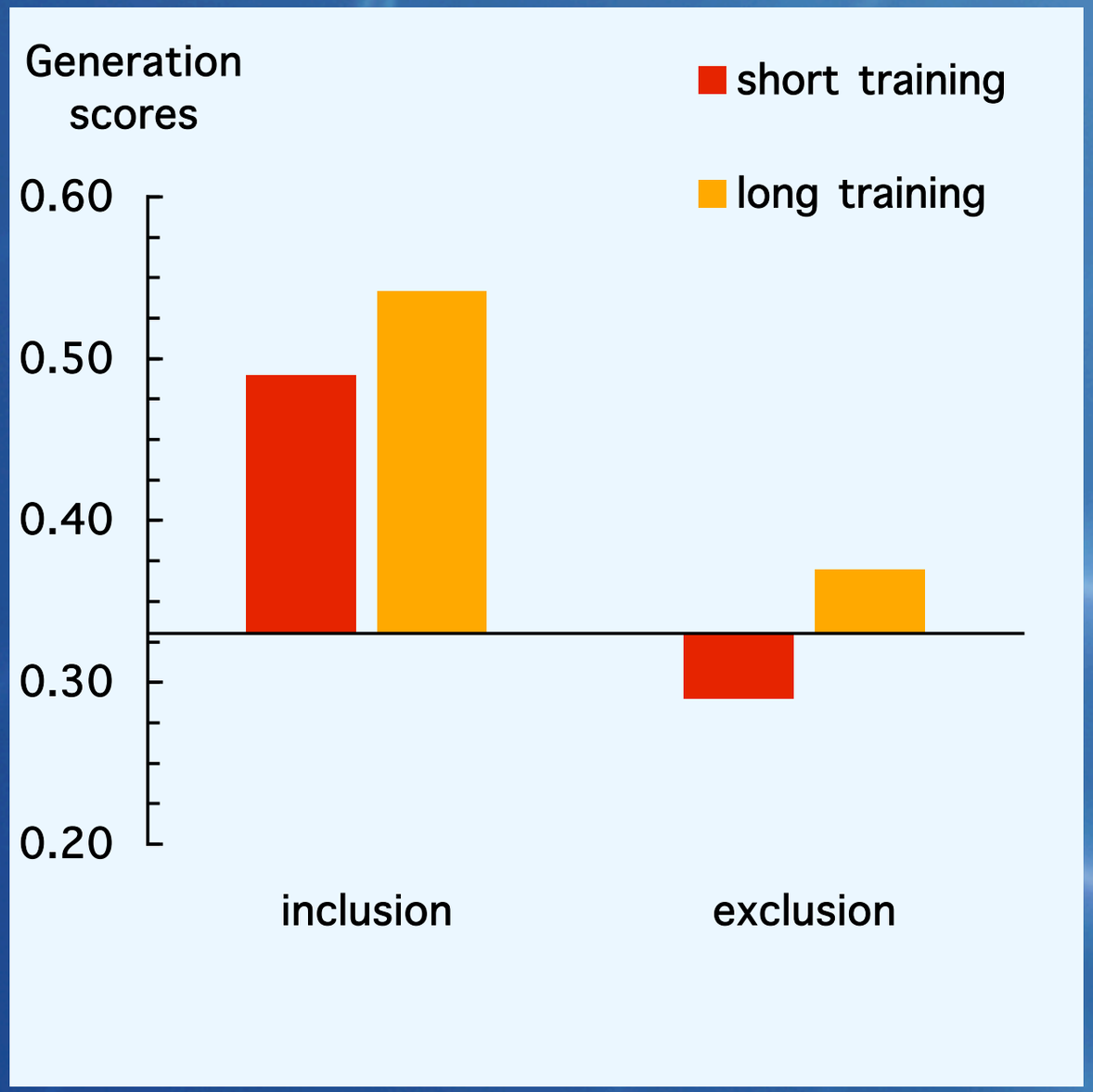
Expertise

Reaction time results



- RT cost during transfer is higher than for subjects trained during 15 blocks (119 ms vs 72 ms, $p < 0.01$)
- ➡ More practice increases sequence learning

Generation results

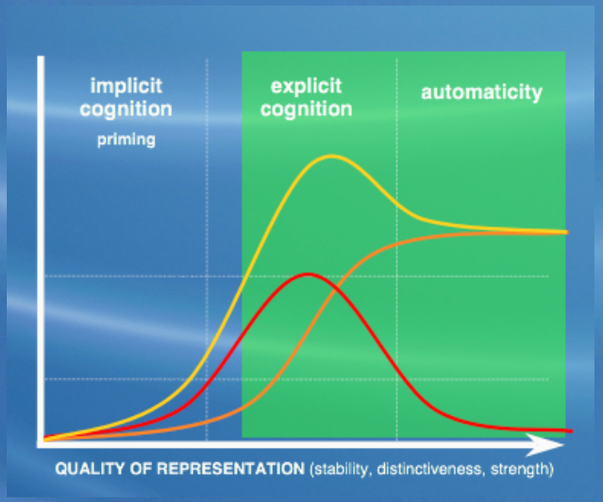


- Practice does not influence inclusion scores
- Exclusion scores are higher for participants in the long training condition

- → Strong traces are more difficult to control

Recognition results

- Participants do better discriminating old from new instances in the long training condition
  - ➡ Strong traces are more available to conscious recollection
  - ➡ Extended practice results in dissociation between control and recognition



>> (Interim) Conclusions

- Dichotomous, monolithic, distinctions are replaced by graded accounts expressed in terms of the computational objectives that different regions of the brain have evolved to solve:
  - conjunctive vs. distributed representations
  - activation-based vs. weight-based processing
  - model learning vs. task learning

- Consciousness involves a graded continuum as well as a dichotomy
  - Adaptation is a central aspect of consciousness:
    - Learning shapes conscious experience
    - Conscious experience shapes learning
  - Adaptation is a mandatory consequence of information processing
    - We learn all the time

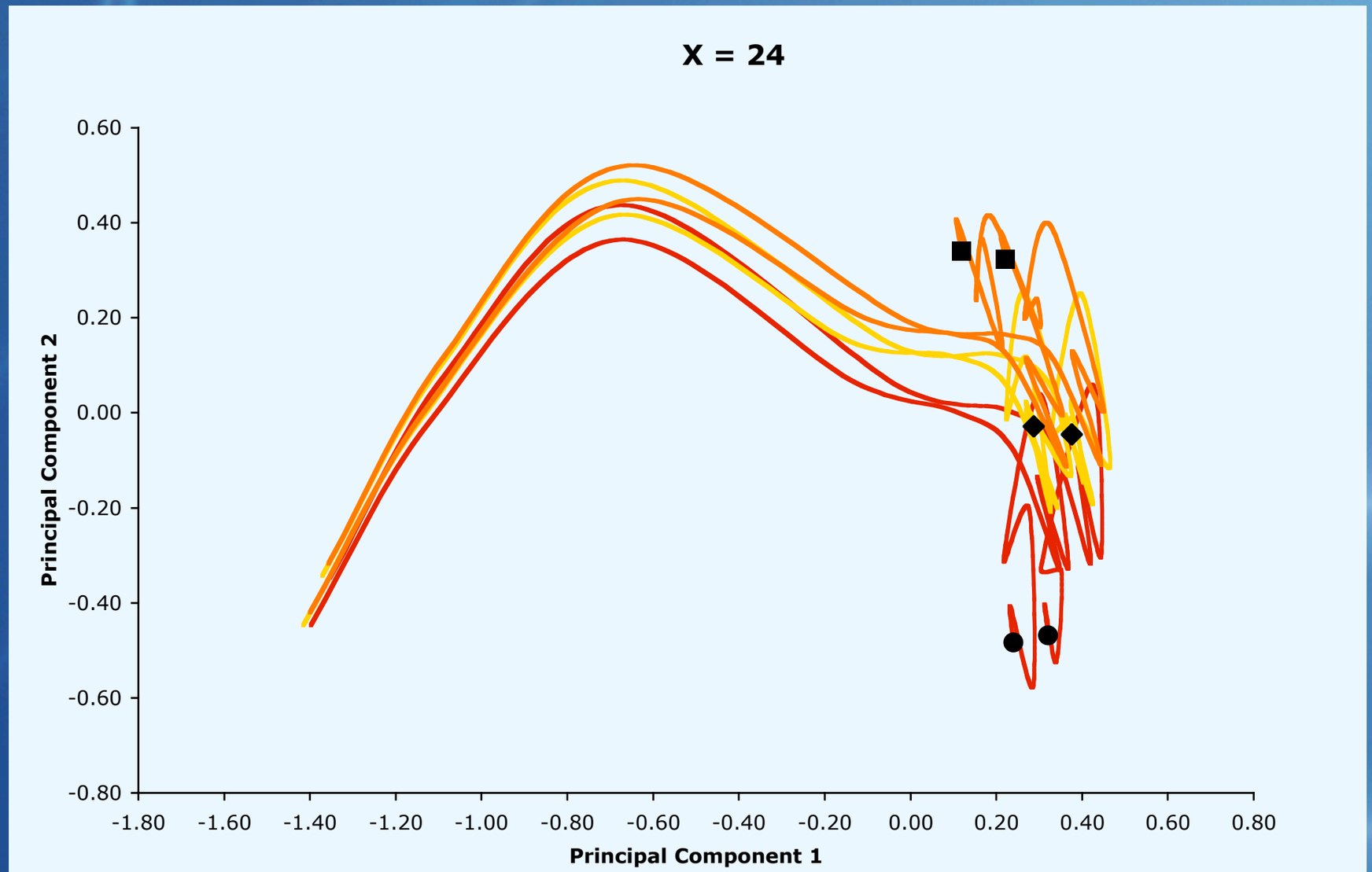
- The function of consciousness is to enable flexible, adaptive control over behavior

>> Being Virtual

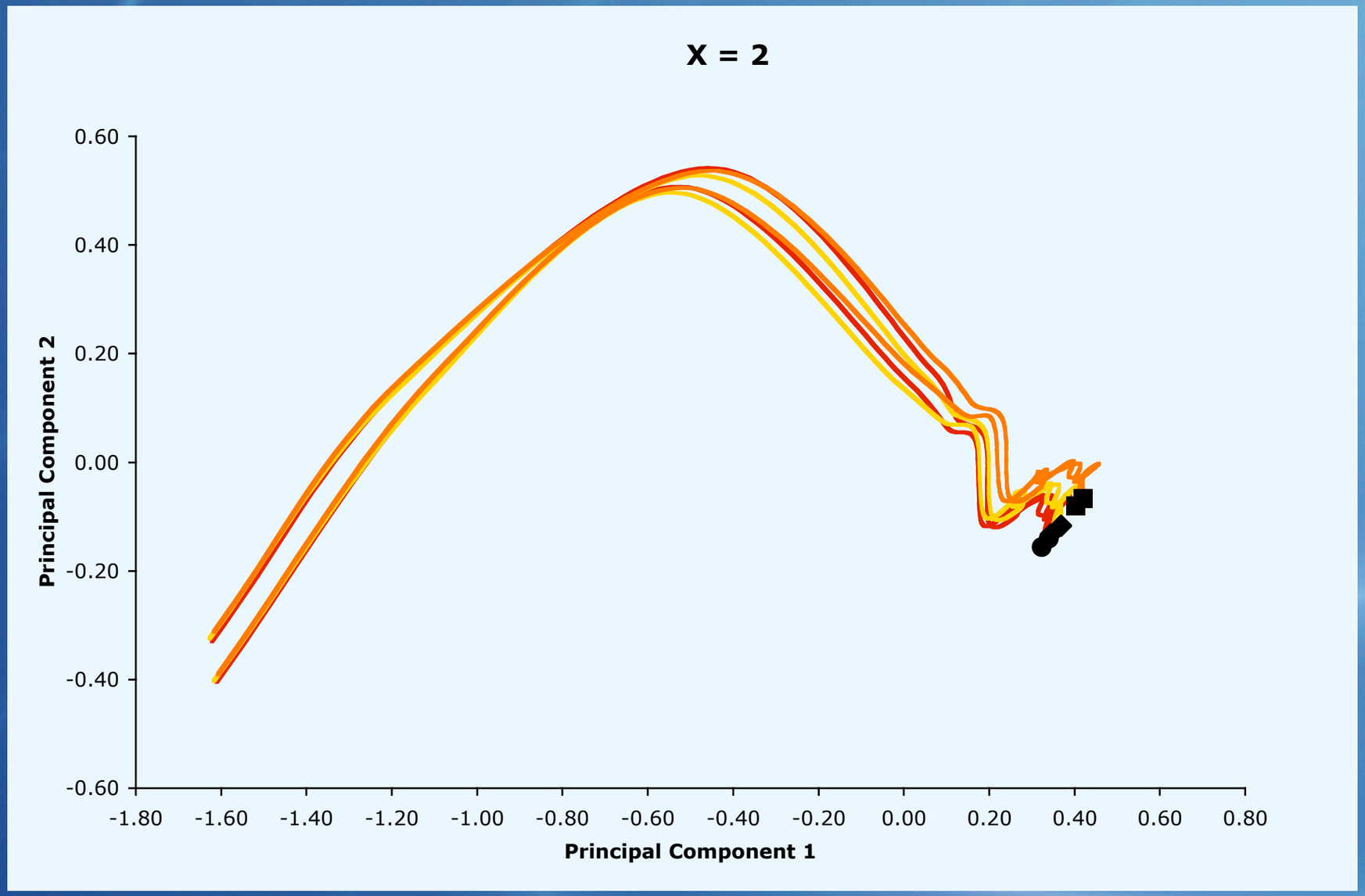
- The framework suggests the necessary conditions under which representations are most likely to be available to form the contents of conscious experience
- What is missing? What might the sufficiency conditions be?

➡ Importance of self-representations (meta-representations) acquired through learning about the consequences of actions directed towards other agents

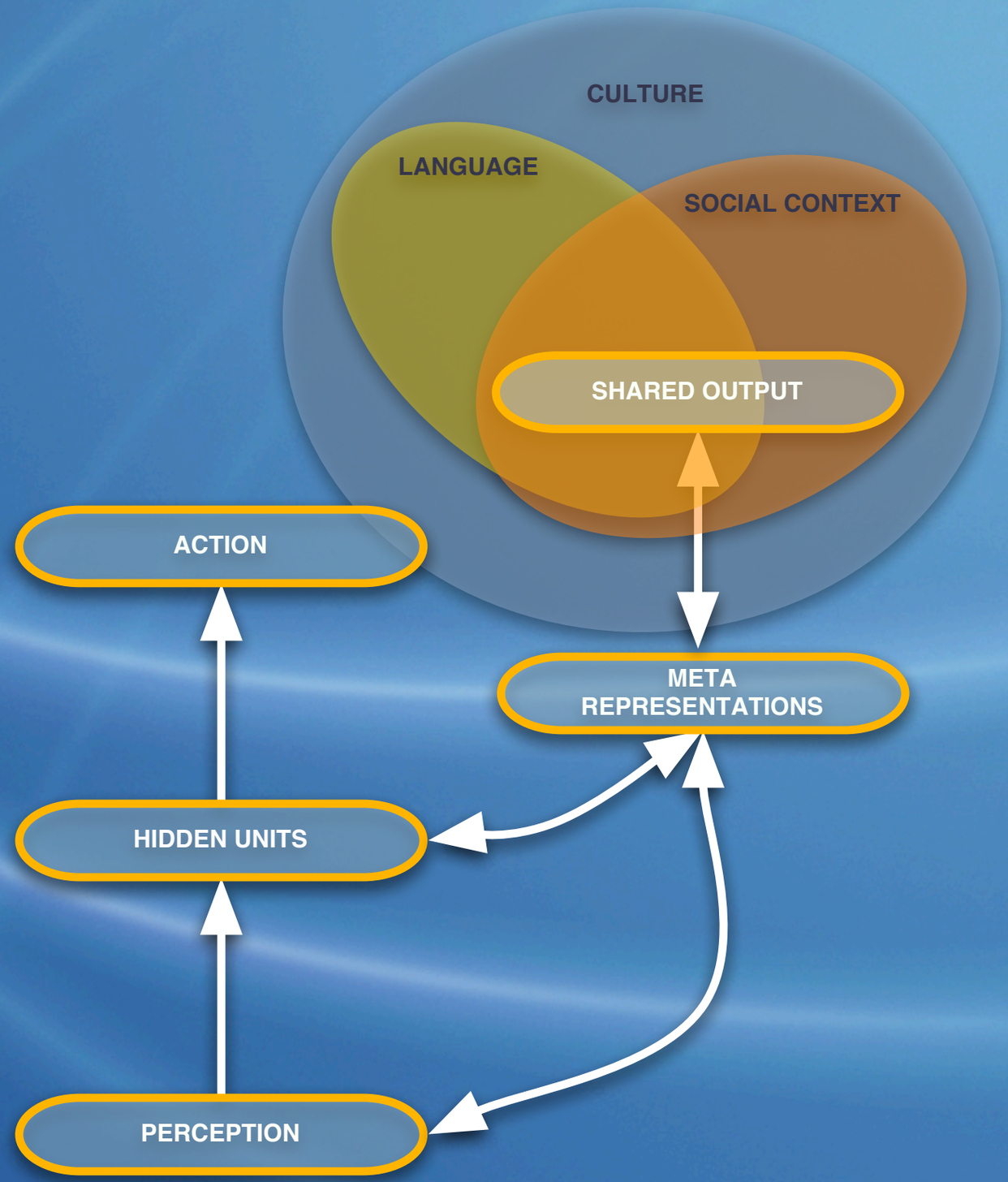
Implicit or explicit?

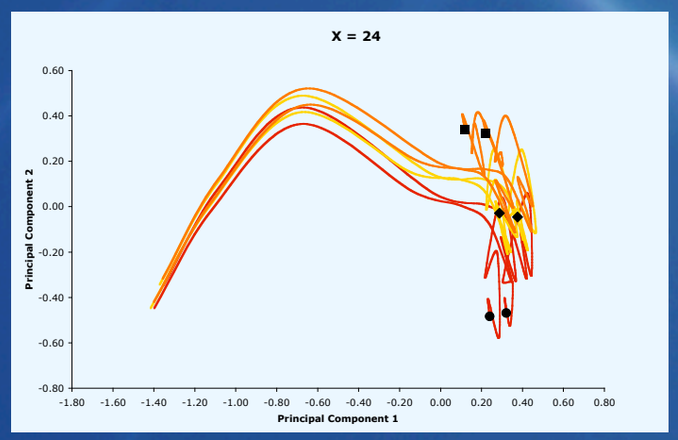


Failure to discriminate



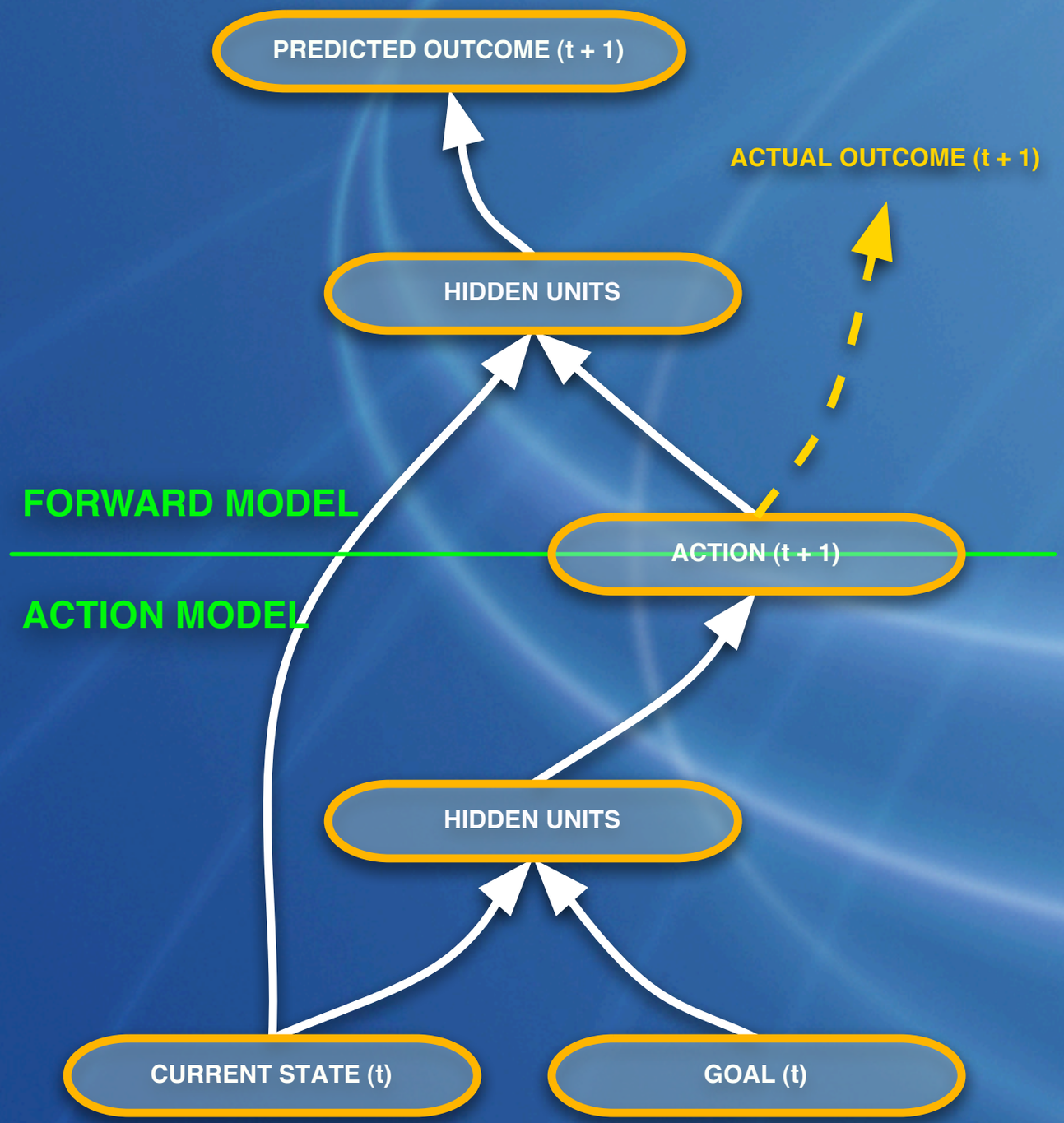
Meta Representation





Knowledge in the network vs. knowledge for the network

"Forward" modeling



- Actions are never reinforced directly
- The forward component becomes a model of the world
- A way of linking action and consciousness in such a way that virtual representations of yourself can develop as a result of your interactions with others
- Consciousness is knowledge of the consequences of your actions, turned inwards (the "enactive view")

Assumptions about Self

S1: A crucial adaptive advantage for any organism is its ability to predict future states of its environment

S2: Successful anticipation of future states in an environment that changes constantly requires organisms equipped with learning mechanisms

S3: Successful anticipation of future states based on current states requires a model of the environment to be built

- Such a model can be extremely simple, consisting of elementary associative links between current states and future states, or very complex, consisting for instance of a simulation of relevant aspects of the environment such that the future consequences of current actions can be explored in a flexible way

Assumptions about Self

S4: When the environment includes other agents, and particularly potentially hostile agents, a crucial adaptive advantage for any organism is its ability to successfully predict the behavior of these agents

S5: Successful anticipation of the behavior of other agents requires a model of how the behavior of these agents is influenced by the environment and by their own internal states

- Again, such models can be very simple or very complex. More complex models, because they are more flexible and more detailed, provide adaptive advantages to the organism that possesses them

Assumptions about Self

S6: From 1—5, it follows that organisms equipped with sufficiently powerful learning mechanisms and with sufficiently developed neural resources will develop detailed models of the internal states of agents it encounters in its environment

S7: An organism that has developed such a detailed model of other agents is conscious in the fullest sense because in so doing it has developed the ability to entertain a third-person perspective on itself. This third-person perspective of a system upon itself arises when this system has developed a simulation of itself based on its emulation of other agents: You are an emulation of what it takes to be an agent. Your "self" is virtual.

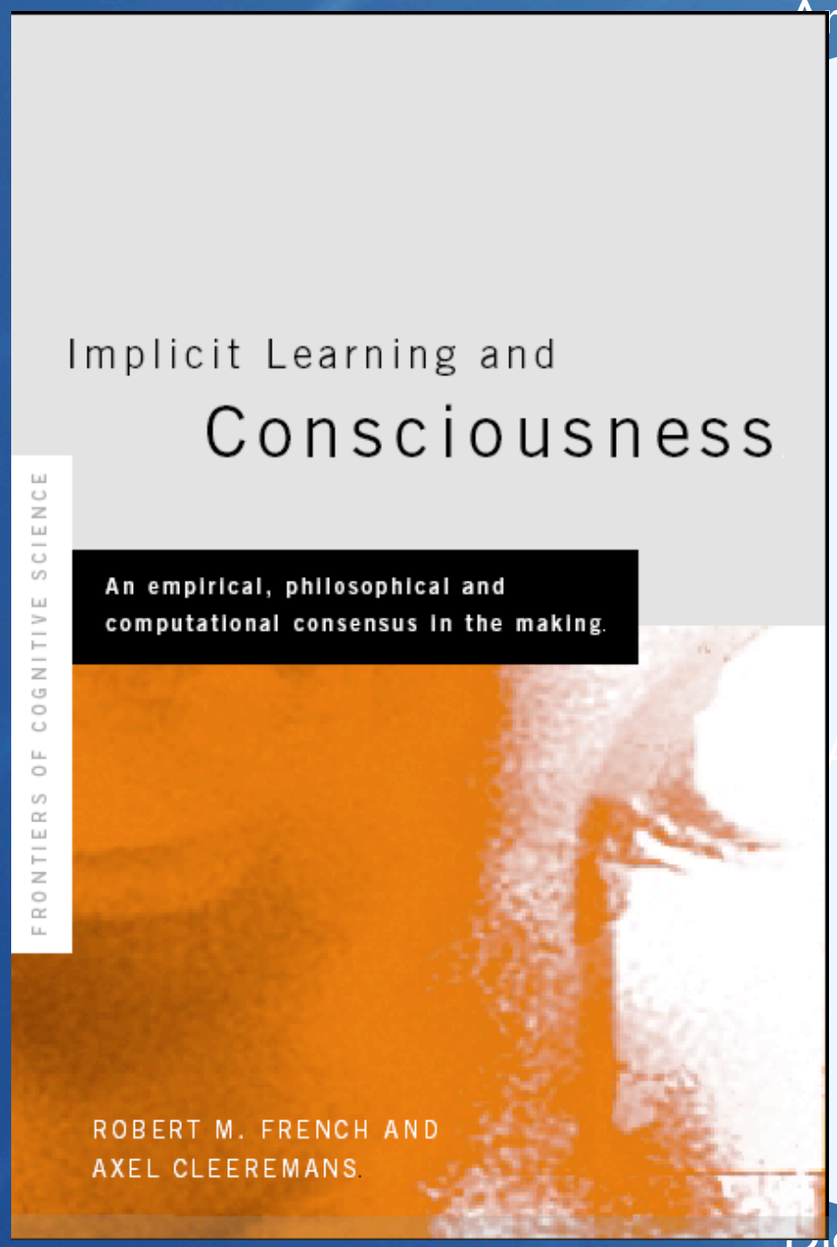
Building models of yourself through action

It's not the child who imitates the mother, but the mother who imitates the child!

We learn to be conscious!

- Consciousness and control are graded dimensions that depend on "quality of representation"
- "Radical plasticity thesis": We learn to be conscious (through action - the enactive view)
- Minimal conditions for C?
  - Massive information-processing resources that are sufficiently powerful to simulate certain aspects of their own inner workings
  - A rich learning system that continuously attempts to predict future states (the consequences of its actions)
  - Immersion in a suitably rich environment from which models of yourself can be built

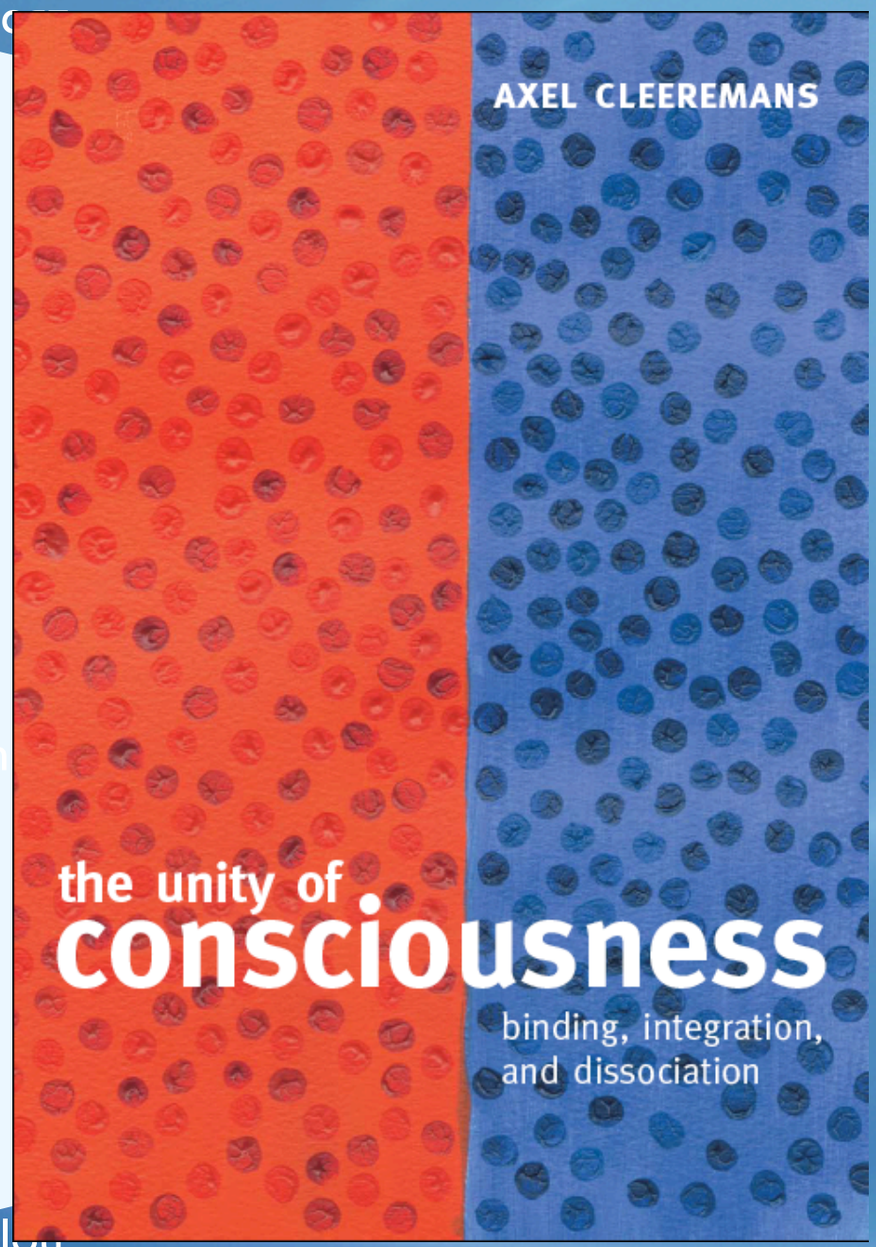
>> Discussion



Arnaud Destrebec

Axel Cleeremans

Dionyssios Theofilou



http://srsc.ulb.ac.be