The CLARION Cognitive Architecture: A Tutorial

Part 4 – The Motivational and Meta-Cognitive Subsystems

Nick Wilson, Ron Sun, Michael Lynch
Cognitive Science, Rensselaer Polytechnic Institute

Outline

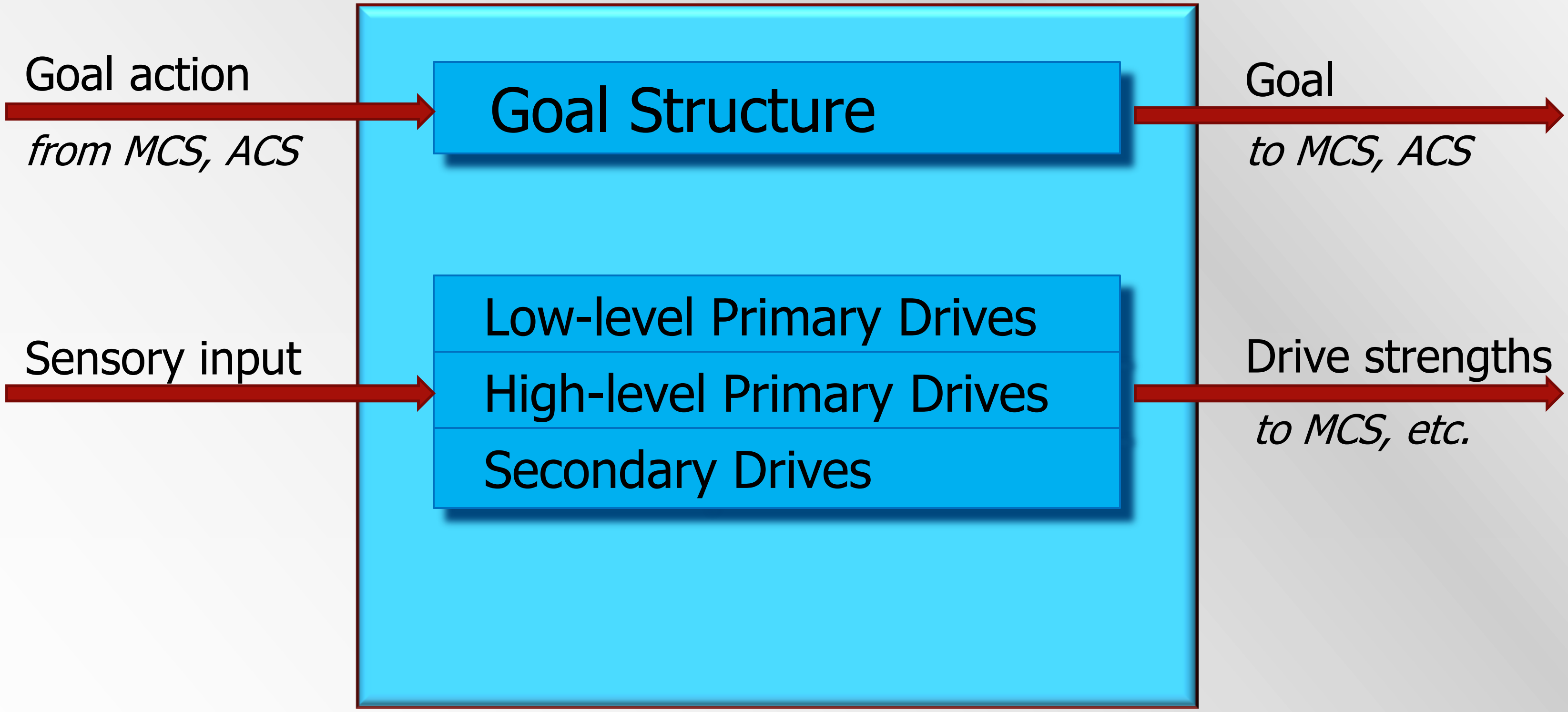
1. The Motivational Subsystem
    1. Introduction
    2. Drives
        1. Low-level Primary Drives
        2. High-level Primary Drives
        3. Drive Strengths
    3. Goal Structure
        1. Goal List
        2. Goal Stack*
2. The Meta-Cognitive Subsystem
    1. Introduction
    2. Structure & Responsibilities
3. Simulation Examples
4. Summary

Introduction

1. Motivational Subsystem
    1. **Introduction**
    2. Drives
        1. Low-level Primary Drives
        2. High-level Primary Drives
        3. Drive Strengths
    3. Goal Structure
        1. Goal List
        2. Goal Stack*
2. Meta-Cognitive Subsystem
    1. Introduction
    2. Structure & Responsibilities
3. Simulation Examples
4. Summary

The Motivational Subsystem (MS)

Motivations are, for the most part, not externally set, but internally generated. Need mechanisms and processes for capturing and explaining them.



Introduction

Implicit vs. explicit representation of motivation:

- Implicit: Drives

  The internal process of generating drives (essential motives, basic needs, or basic desires) is not readily accessible cognitively (Murray, 1938; Maslow, 1943; Hull, 1951)

- Explicit: Goals

  Explicit motivational representation consists of explicit goals (Newell, 1990; Anderson and Lebiere, 1998), which may be used for action selection

Introduction

- Motivational processes are highly complex and varied (Weiner, 1992).
  - Cannot be captured with simple explicit goal representation alone (e.g., ACT-R)
  - Motivational processes may be based on unconscious “needs” or motives, especially biologically/ecologically essential ones (cf. e.g., Maslow, 1943; Murray, 1938)
- Explicit goal representations arise to clarify and supplement implicit motivational dynamics.

Introduction

Why do we need a motivational system?

A cognitive agent must address the following in its everyday activities:

- Sustainability
- Purposefulness
- Focus
- Adaptivity

Introduction

- Sustainability

  One must attend to essential needs in order to be sustainable (Toates, 1986), for example: Hunger, Thirst, etc.

- Purposefulness

  Actions must be chosen in accordance with some criteria, instead of completely randomly (Hull, 1943; Anderson, 1993)

  Criterion: enhancing sustainability (Toates, 1986)

Introduction

- Focus

  An agent must be able to direct its activities with respect to a specific purpose (Toates, 1987)

  Therefore, actions need to be: *Consistent, Persistent, Contiguous, etc.*

  But, must also be able to give up activities when necessary (Simon, 1967; Sloman, 1986)

- Adaptivity

  Must be able to change behavioral patterns (i.e. learn) for the sake of sustainability, purposefulness, and focus.

Introduction

It is reasonable to assume (see Sun, 2009):

- Dual motivational representation (more later)
- Primacy of implicit motivational processes:
  - Implicit motivational processes: more basic and more essential than explicit processes (Hull; Maslow; etc.)
  - Capture:

    Basic drives
    Basic needs
    Basic desires
    Intrinsic motives

Drives

1. Motivational Subsystem
    1. Introduction
    2. **Drives**
        1. Low-level Primary Drives
        2. High-level Primary Drives
        3. Drive Strengths
    3. Goal Structure
        1. Goal List
        2. Goal Stack*
2. Meta-Cognitive Subsystem
    1. Introduction
    2. Structure & Responsibilities
3. Simulation Examples
4. Summary

Drives

- Hull (1950):
  - developed the most detailed conception of “drives”
  - a pre-conceptual (i.e., implicit) representation of motives
- Strict notion of drives:
  - Physiological deficits that require reduction by corresponding behaviors (Hull, 1951; Weiner, 1982)
- A Generalized interpretation of drives:
  - Internally felt needs of all kinds that likely may lead to corresponding behaviors
  - Physiological or otherwise
  - For end-states or process-states
  - May or may not be reduced by corresponding behavior
- Transcends controversies surrounding the stricter notions of drive
- Accounts for behaviors and satisfies some important considerations (more later)

Drives

1. Motivational Subsystem
   1. Introduction
   2. Drives
      1. **Low-level Primary Drives**
      2. High-level Primary Drives
      3. Drive Strengths
   3. Goal Structure
      1. Goal List
      2. Goal Stack*
2. Meta-Cognitive Subsystem
   1. Introduction
   2. Structure & Responsibilities
3. Simulation Examples
4. Summary

Drives

Low-level Primary Drives:

- Mostly physiological, mostly evolutionarily formed, hard-wired
- Include:

    Food
    Water
    Sleep
    Avoiding physical danger
    Reproduction
    Etc.

Drives

1. Motivational Subsystem
    1. Introduction
    2. Drives
        1. Low-level Primary Drives
        2. **High-level Primary Drives**
        3. Drive Strength Considerations
    3. Goal Structure
        1. Goal List
        2. Goal Stack
2. Meta-Cognitive Subsystem
    1. Introduction
    2. Structure & Responsibilities
3. Simulation Examples
4. Summary

Drives

High-level Primary Drives:

- More socially oriented
- More or less hard-wired, innate, but mostly not physiological
- See, e.g., Sun (2009) for details and justifications

Drives

High-level Primary Drives:

the following drives were posited (Sun, 2009; Murray, 1938; James, 1890; Maslow, 1987; Reiss, 2004):

- *Affiliation and Belongingness*:

  The drive to associate with other people and to be part of social groups

- *Dominance and Power*:

  The drive to have power over other individuals or groups

- *Recognition and Achievement*:

  The drive to excel and be viewed as being accomplished at something

Drives

High-level Primary Drives (cont.)

- *Autonomy*
- *Deference*
- *Similance*
- *Fairness*
- *Honor*
- *Nurturance*
- *Conservation*
- *Curiosity*

(see Sun, 2009 for definitions)

Drives

Primary Drive Considerations:

- Empirical data: The primary drives are largely uncorrelated and individually significant (Reiss, 2004)
- Each drive may be weighted differently by different individuals, leading to *individual differences* (even personality differences; more later)
- One aims for a "moderate mean" (Aristotle, 1953; Reiss, 2004): Desirable levels of satisfaction is often neither the highest nor the lowest (e.g., low food, high drive; high food, low drive)

Drives

There are also

Secondary (Derived) Drives

- Secondary and more changeable
- Acquired in the process of satisfying primary drives
  - Gradually acquired drives through conditioning (Hull, 1951)
  - Externally set drives through externally given instructions
  - Etc.

Drives

Drives

Drive Strength Considerations (e.g., Tyrell, 1993; Sun, 2003)

- Proportional Activation*:*

  Drive activation should be proportional to the perceived deficit

- Opportunism*:*

  Opportunities need to be taken into consideration when calculating desirability of alternatives

- Contiguity of Actions*:*

  Tendency to continue the current action sequence

Drives

Drive Strength Considerations (cont.)

- Persistence:

  Actions to satisfy a drive should persist beyond minimum satisfaction

- Interruption when Necessary:

  When a much more urgent drive arises, actions for a lower-priority drive may be interrupted

- Combination of Preferences:

  A compromise candidate may be generated that is the best in terms of the combined preferences of the different drives

Drives

General structure of a drive (cf. Tyrell 1993; Toates, 1987; Sun, 2009):

$$ds_d = \alpha_d Stimulus_d \times Deficit_d + \beta_d$$

where α_d is the gain for drive *d* (consisting of three gain parameters) and β_d is the drive *d* baseline

Additionally, persistence factors, etc.

Drives

- *Deficit*$_d$ can represent an innate psychological sensitivity to a certain need (including, in some cases, a perceived physiological deficit)
- *Stimulus*$_d$ may consist of an internal evaluation of the relevance of the current state relative to a particular drive, taking into consideration:
  - Sensory information
  - Maybe working memory
  - Maybe current goal
  - A meta-cognitive subsystem filtered interpretation of the above (will be discussed later)

Drives*

A few examples of calculating drive strengths:

- Food:

$$ds_{food} = 0.95 \times Max(0.30 \times Deficit_{food}, Stimulus_{food} \times Deficit_{food})$$

- Avoiding Physical Danger:

$$ds_{danger} = 0.98 \times Stimulus_{danger} \times Sensitivity_{danger}$$

- Affiliation and Belongingness:

$$ds_{a\&b} = 0.50 \times Deficit_{a\&b} \times Stimulus_{a\&b} + 0.20$$

- Recognition and Achievement:

$$ds_{r\&a} = 0.40 \times Deficit_{r\&a} \times Stimulus_{r\&a} + 0.10$$

Drives

BIS versus BAS: avoidance vs. approach drives (reward-seeking vs. punishment avoiding; cf. Gray, 1987):

Approach Drives	Avoidance Drives	Both
Food	Sleep	Affiliation & Belongingness
Water	Avoiding Danger	Similance
Reproduction	Avoiding Unpleasant Stimuli	Deference
Nurturance	Honor	Autonomy
Curiosity	Conservation	Fairness
Dominance & Power		
Recognition & Achievement		

Table 1. Approach versus avoidance drives.

Drives

BIS versus BAS: avoidance vs. approach drives (cf. Gray, 1987; Clark & Watson, 1999; Cacioppo, Gardner, & Berntson, 1999): Justifications for the division

This determination is based on seeking positive rewards vs. avoiding punishments (i.e., negative rewards).

- This determination does not involve complex reasoning, mental simulation, etc., because the processes of drives are reflexive and immediate (Heidegger 1927; Dreyfus 1992).
- Some drives come with intrinsic positive rewards (e.g., food, reproduction, fairness/vengeance, dominance, affiliation, achievement, etc. ---- essentially all the drives in the BAS), while others do not have related intrinsic positive reward (e.g., sleeping, avoiding danger, avoiding the unpleasant; so, mostly, they are for avoiding negative rewards

Drives

Questions?

Goal Structure

1. Motivational Subsystem
    1. Introduction
    2. Drives
        1. Low-level Primary Drives
        2. High-level Primary Drives
        3. Drive Strengths
    3. **Goal Structure**
        1. Goal List
        2. Goal Stack*
2. Meta-Cognitive Subsystem
    1. Introduction
    2. Structure & Responsibilities
3. Simulation Examples
4. Summary

Goal Structure

- Drives: provide the context within which explicit goals are set
- Goals: more clear-cut, more specific, and more explicit
- Goals: different from drives:
  - Multiple drives may be activated at the same time while only one goal may be pursued at a time
  - Drives are diffused in focus whereas goals are often more specific (e.g., Anderson and Lebiere, 1998)
  - Drives are more implicit, while goals are more explicit (Murray, 1938; Maslow, 1943; Hull, 1951)
  - Drives are more hardwired, whereas goals are more flexibly created, set, and carried out (Hull, 1951; Sun, 2009)

Goal Structure

- Goals provide specific, tangible motivations for the actions chosen and performed in the ACS

  (Note: Actions are chosen in the ACS on the basis of the current input state and the current goal)
- The goal structure enables sequencing of actions (behavioral routines)
- It also facilitates communicating motivation to the other subsystems (and to other agents)
- Implemented in CLARION as:

  A goal list

  A goal stack*

Goal Structure

1. Motivational Subsystem
    1. Introduction
    2. Drives
        1. Low-level Primary Drives
        2. High-level Primary Drives
        3. Drive Strengths
    3. Goal Structure
        1. **Goal List**
        2. Goal Stack*
2. Meta-Cognitive Subsystem
    1. Introduction
    2. Structure & Responsibilities
3. Simulation Examples
4. Summary

Goal Structure

Goal List

- Randomly accessible linear structure that contains a set of goal items
  - Each slot of the list can contain a goal chunk
  - A goal chunk is made up of:
    - A goal identification dimension
    - A number of parameter dimensions
  - Location of goal items on the list is irrelevant (i.e., not a stack or queue)
- Goal items on the goal list compete with each other to become active (e.g., using Boltzmann distribution)
- A more psychologically realistic approach toward modeling of motivational control of behavior

Goal Structure

- Goal List (cont.)
  - Goal items have a recency based base-level activation (BLA) that allows goals to decay over time:

$$B_i^g = iB_i^g + c \times \sum_{l=1}^{n} t_l^{-d}$$

Where t_l is the l^{th} setting of goal i and iB_i is the initial value.

  - Goals compete using BLA through a Boltzmann distribution

Goal Structure

- Goal List (cont.)
  - Can approximate stack-like behaviors when needed (using BLA)
  - Goal list can handle more complex or more subtle situations (than a goal stack or queue)
  - Goal alternation (e.g., task switching) is easier using a goal list
  - Can be used to generate complex behavioral “routines” (sequences; e.g., navigating a route, TOH, etc.)

Goal Structure

- Goal List (cont.)

Goal actions: set a goal, remove a goal, etc.

- *Insert i {dim, value}*

  i is the value of the goal symbol for the goal dimension, *{dim, value}* are optional parameters to be set along with the goal

- *Delete i {dim, value}*
- *Do-nothing*

Goal Structure*

1. Motivational Subsystem
    1. Introduction
    2. Drives
        1. Low-level Primary Drives
        2. High-level Primary Drives
        3. Drive Strengths
    3. Goal Structure
        1. Goal List
        2. **Goal Stack***
2. Meta-Cognitive Subsystem
    1. Introduction
    2. Structure & Responsibilities
3. Simulation Examples
4. Summary

Goal Structure*

Goal Stack

- Linear structure of multiple items in which only the top item may be accessed
  - → Only one goal item may be active at a time
- Items may be added to or removed from the top of the stack
- A currently active goal becomes inactive when a new goal is added on top of it
- Becomes reactivated when all goals on top of it are removed

Goal Structure*

Goal stack actions:

- *Push i {dim, value}*

  i is the value of the goal symbol for the goal dimension, *{dim, value}* are optional parameters to be set along with the goal

- *Pop*
- *Do-nothing*

Goal Structure*

- The Goal Stack allows for the emergence and application of relatively fixed patterns of behavior (“routines”)
- Problems:
  - Too idealistic for modeling cognitive processes realistically
  - Much of the subtlety and complexity involving goal coordination is lost
  - Rough approximation and abstraction of a complex motivational and meta-cognitive process

Goal Structure

Questions?

Meta-Cognitive Subsystem

1. Motivational Subsystem
    1. Introduction
    2. Drives
        1. Low-level Primary Drives
        2. High-level Primary Drives
        3. Drive Strengths
    3. Goal Structure
        1. Goal List
        2. Goal Stack*
2. **Meta-Cognitive Subsystem**
    1. Introduction
    2. Structure & Responsibilities
3. Simulation Examples
4. Summary

Meta-Cognitive Subsystem

1. Motivational Subsystem
    1. Introduction
    2. Drives
        1. Low-level Primary Drives
        2. High-level Primary Drives
        3. Drive Strengths
    3. Goal Structure
        1. Goal List
        2. Goal Stack*
2. Meta-Cognitive Subsystem
    1. **Introduction**
    2. Structure & Responsibilities
3. Simulation Examples
4. Summary

Meta-Cognitive Subsystem

- Meta-cognition refers to (Flavell, 1976):
  - … knowledge concerning one's own cognitive processes and products
  - … the active monitoring and consequent regulation and orchestration of processes in relation to the cognitive objects or data on which they bear
  - … usually in the service of some concrete goal or objective
- Drives and goal structures (in the MS) lead to the need for meta-cognitive control, regulating
  - Goal structure
  - Other cognitive processes for achieving goals

Meta-Cognitive Subsystem

- Like the ACS, the MCS is:
  - Action-oriented
  - Comprised of two levels:
    - The bottom level consists of implicit decision networks
    - The top level consists of (groups of) rules
- *Mostly the bottom level takes effective control
  - Meta-cognition is often fast, effortless, and implicit (e.g., Reder and Schunn, 1996)
  - Under some circumstances, the top level can also exert influence

Structure & Responsibilities

Structure & Responsibilities

Several types of meta-cognitive processes:

- Behavioral aiming
  - Goal setting
  - Reinforcement setting
- Information filtering
  - Focusing of inputs to the various other subsystems
- Information acquisition
  - Selection of learning methods
- Monitoring buffer
  - Monitoring performance of the various subsystems

Structure & Responsibilities

Meta-cognitive process types (cont.):

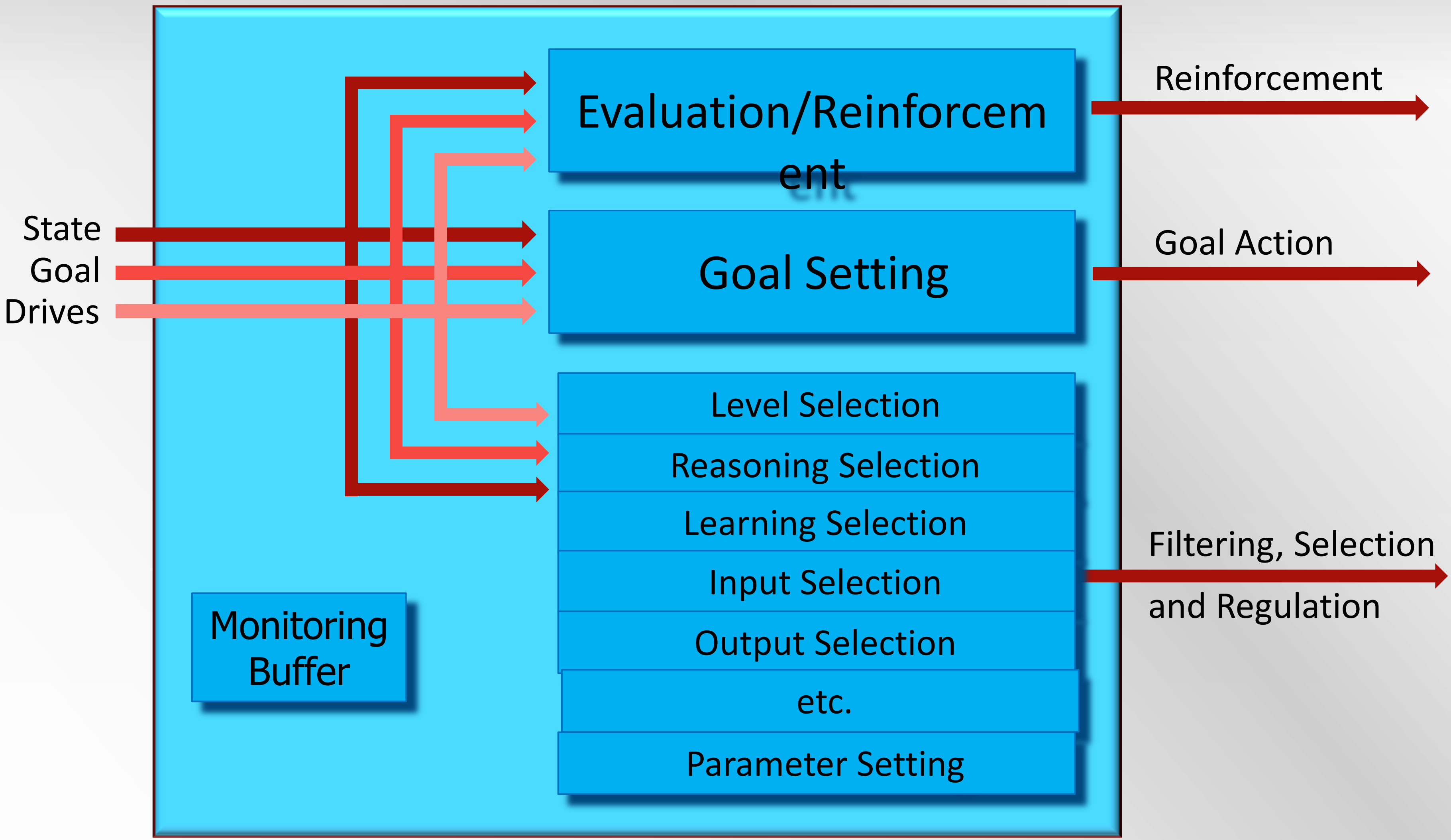
- Information utilization
  - Selection of reasoning methods
- Outcome selection
  - Selection of outputs
- Cognitive modes
  - Level selection/integration
- Parameter setting

Structure & Responsibilities

The MCS is divided into modules correspondingly:

- Goal-setting
- Reinforcement
- Filtering
- Learning and reasoning
- Level selection
- Parameter setting
- Monitoring buffer
  - ACS performance & learning
  - NACS performance & learning
- And others

The Meta-Cognitive Subsystem (MCS)



Structure & Responsibilities

Goal-setting

- The process of goal setting may be performed by the MCS (or by the ACS)
- Maps the strengths (activations) of the drives and the current input state (and possibly other factors) to a new goal
- *Two goal setting possibilities:
  - Balance-of-interests (preferred):
    - Each drive votes for multiple goals
    - Goal with the highest combined score becomes the new goal
      - ("Bonus" is given to the current goal, to prevent thrashing and to promote persistence)
  - Winner-take-all:
    - Drive with the highest strength wins (deterministically or stochastically)
    - New goal is the goal that best attends to the winning drive

Structure & Responsibilities*

Goal-setting (cont.)

- for setting goals, calculate goal strengths:

$$gs_g = \sum_{d=1}^{n} Relevance_{s,d \rightarrow g} \times ds_d$$

where $Relevance_{s,d \rightarrow g}$ is a measure of how well the goal addresses the drive (given the current input state)

- Stochastic selection of a goal based on gs_g (e.g. using Boltzmann distribution)

Structure & Responsibilities

Reinforcement/Evaluation

- MCS enables reinforcement learning --- addresses the key issue underlying reinforcement learning:
  - How to come up with an appropriate reinforcement signal
- The external world does not provide a simple, scalar reinforcement signal
  - The world simply changes (into a "new state") after an action is performed
  - An appropriate reinforcement signal has to be determined "internally" by synthesizing various external/internal information (including drives, goals and sensory information)

Structure & Responsibilities

Reinforcement (cont.)

- Reinforcement/evaluation module (e.g., Sun and Fleischer, 2011):
  - Takes external and internal sensory information, drives, and the goal (and possibly other factors) as input
  - Evaluates the current input state in terms of the current goal, in the context of the currently active drives,
  - Determines if the goal and/or the drives are satisfied or not (binary reinforcement) or the degree to which the current state satisfies the goal and/or the drives (graded reinforcement)
- *Minimally necessary: an evaluation for states that directly satisfy a goal in some way (not every step)

  Sequential decision learning (such as Q-learning) can automatically propagate reinforcement to temporally adjacent states and actions (as well as similar states/actions if BP nets are used)

Structure & Responsibilities*

Reinforcement (cont.)

- Can be accomplished either implicitly, explicitly, or both

  Implicitly (at the bottom level):

  - A neural network generates reinforcement signals from the current drives, goals and sensory information (and other factors)

  Explicitly (at the top level):

  - A reinforcement function is specified explicitly via a set of rules, which map certain inputs to reinforcement signals

Structure & Responsibilities

Filtering

- Attention focusing of input and output may be based on:
  - Current goal
  - Current drive activations
  - Current sensory input

  As well as, e.g.,
  - Working memory
  - Performance of the subsystems (as determined by the monitoring buffer)
- Allows dimensions to be either suppressed or maintained before being delivered to a subsystem
  - Different subsystems can receive different information, to allow different foci

Structure & Responsibilities

Learning and Reasoning Methods Selection

- Learning and reasoning methods in the ACS and the NACS may be selected by the MCS
- (However, learning and reasoning methods in the NACS may also be selected by the ACS.)
- These learning and reasoning methods can be set separately within each subsystem and for different modules within each subsystem

Structure & Responsibilities

Level Selection/integration

- At which level of the ACS action recommendations are made: Use current input state, goal, and drive activation information to determine.
- One possibility: level selection may be determined by some drive strengths using an inverted U-curve (Yerkes and Dodson, 1908; Hardy and Parfitt, 1991)

The following inverted U-curve follows a parabolic trajectory:

$$P_{BL} = ax^2 + bx + c$$

where x is some combination of some drive strengths (possibly *max* of BIS drives), $a < 0$, and $0 < c < 1$

Structure & Responsibilities

Parameter Setting

- Includes various parameters:
  - Bottom-level learning rate in the ACS
  - Q-learning discount factor for bottom-level learning in the ACS
  - Temperature in stochastic selection in the ACS
  - Rule learning thresholds in the ACS (extraction, specialization, and generalization)
  - Various parameters in the NACS . . .
  - Etc. etc.

Structure & Responsibilities

Monitoring Buffer

- Stores various information about the internal workings of the system:
  - Performance statistics about the ACS and NACS modules
  - Learning statistics about the ACS and NACS modules
  - Settings of parameters that can be manipulated by the MCS
  - ……
  - Etc. etc.

Structure & Responsibilities

Questions?

Simulation Examples

1. Motivational Subsystem
   1. Introduction
   2. Drives
      1. Low-level Primary Drives
      2. High-level Primary Drives
      3. Drive Strengths
   3. Goal Structure
      1. Goal List
      2. Goal Stack*
2. Meta-Cognitive Subsystem
   1. Introduction
   2. Structure & Responsibilities
3. **Simulation Examples**
4. Summary

Simulation Examples

Lack of knowledge inference task

(meta-cognitive intervention; Sun, Zhang, and Matthews 2006)

Protocol Examples:

Q: Have you ever shaken hands with Richard Nixon?

A: No…How do I know? It's not something that one would forget…(Gentner & Collins 1991)

Q: Is the Nile longer than the Mekong river?

A: I think so… Because in junior high, I read a book on rivers,… the Amazon was in there and the Nile was in there and they were big and long and important… (Collins 1978)

Q: Is Kissinger 6'6" tall?

A: If Kissinger were 6'6" tall, I would know he is very tall. I don't, so he must not be that tall. (Collins 1978)

Simulation Examples

Lack of Knowledge inference task (cont.)

- Inferences can be made based on:
  - The lack of knowledge about something, AND
  - The importance/significance of that knowledge
- To make these inferences, the MCS must:
  - Monitor the reasoning process
  - Intervene and redirect the reasoning process when necessary

Simulation Examples

Simulation setup

- Captured meta-cognitive monitoring and intervention
- Subsystems used:
  - The ACS
    - Directed reasoning of the NACS
  - The NACS
    - Performed inferences
  - The MCS
    - Selected information to be used and reasoning methods to be applied
    - Monitored the progress of inference in the NACS
    - Performed intervention by starting the “lack of knowledge” inference in the NACS

Simulation Examples

Simulation setup (cont.)

Some details:

- NACS top level contained associative rules:

$$River_a \rightarrow long\_river$$
$$River_b \rightarrow long\_river$$
$$River_c \rightarrow long\_river$$

- NACS bottom level embodied (was trained with) similar knowledge

Simulation Examples

Simulation setup (cont.)

- MCS
  - Goal-setting

    Set the initial goal as "*regular inference*"

    Choose "*LOK inference*" as the goal when the "lack-of-knowledge" condition is detected (uniformly low activation; detected in the NACS performance section of the MCS monitoring buffer)
  - Filtering

    Select relevant input dimensions to be used by the NACS
  - Reasoning

    Select the "*forward chaining with SBR*" reasoning method in the NACS

Simulation Examples

Simulation setup (cont.)

- ACS: Mainly top-level rules for directing NACS reasoning

If $goal = regular\_inference$, then perform one - step of inference in the NACS

If $goal = regular\_inference$ & chunk i is a conclusion chunk with $S_i^c > threshold_s$ and " $j : S_i^c > S_j^c$, then retrieve chunk i

If $goal = LOK\_inference$ & no conclusion chunk has $S_i^c > threshold_s$ but there are many associative rules pointing to the conclusion chunk, then the conclusion is negative

If $goal = LOK\_inference$ & no conclusion chunk has $S_i^c > threshold_s$ & there are no associative rules pointing to the conclusion chunk, then the conclusion is indeterminate

where $threshold_s$ is set to *.1*

Simulation Examples

Simulation results

- Simulation captured “lack-of-knowledge” inference exhibited by the human subjects in the protocols described earlier
- As predicted
  - When a simulated subject had a (relatively) large amount of knowledge about a conclusion but could not reach that conclusion in a particular instance, then the lack-of-knowledge inference was initiated and a negative answer was produced
  - When a simulated subject had little knowledge about a conclusion, then the no conclusion was given

Simulation Examples

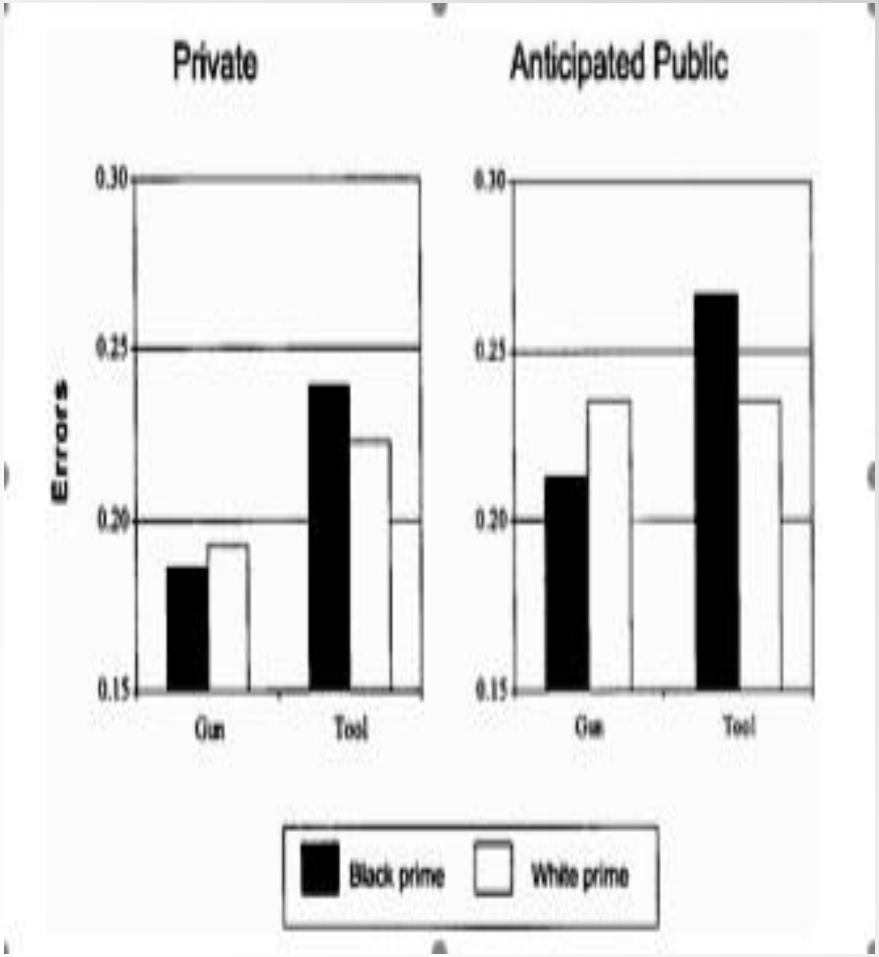
Questions about the above simulation?

Simulation Examples

Stereotyping task (Lambert et al., 2003)

- Two groups
  - Private
  - Anticipated Public
- Primes
  - 4 black faces (2 male, 2 female)
  - 4 white faces (2 male, 2 female)
- Targets
  - 4 guns
  - 4 tools (2 wrenches, 1 drill, 1 ratchet)

Simulation Examples



Findings (human data)

- The anticipated public group made more mistakes than the private group
- The error rates were significantly higher for tool targets when paired with a black face than a white face
- Error rates for gun targets did not vary based on face prime (combined over group)
- Error rates for either target when paired with a white face did not vary significantly (combined over group)

Simulation Examples

- Process Dissociation (cf. Jacoby et al.)
  - Cognitive Control Estimate

    Calculates approximate frequency of controlled (i.e., explicit) responses (P_{TL})

$$CCE = P(correct \mid congruent) - P(stereotypic\_error \mid incongruent)$$

  - Accessibility Bias Estimate

    Calculates likelihood of making stereotyped response when control failed (implicit stereotyped response)

$$ABE = \frac{P(\mathrm{stereotypic\_error} \mid \mathrm{incongruent})}{(1-c)}$$

Simulation Examples

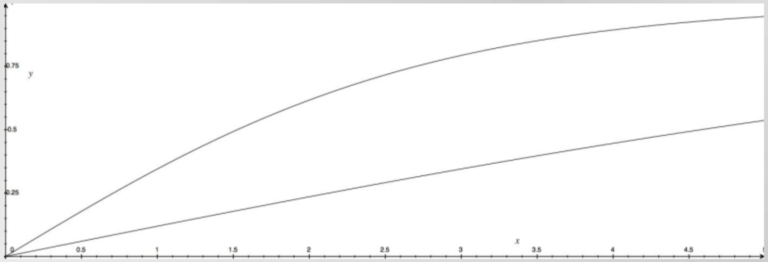
Simulation setup

- The MS – the "honor" drive
  - Deficit: reflecting individual differences --- propensity toward anxiety (trait anxiety)
  - Situational stimulus: capturing group condition difference (private vs. public)
- The MCS – level integration with an inverted u-curve (Yerkes and Dodson, 1908)
  - Input: drive strength
  - Output: P_{TL} (i.e., cognitive control)
- The ACS
  - Top Level:
    - 8 Fixed Rules: map target features to object type (reflecting prior knowledge, e.g., obtained during practice trials)
  - Bottom Level:
    - Neural Network (25 input, 5 hidden, 2 output): maps characteristics of target and prime to object type (e.g., pre-trained to map race to object type based on accessibility bias estimate)

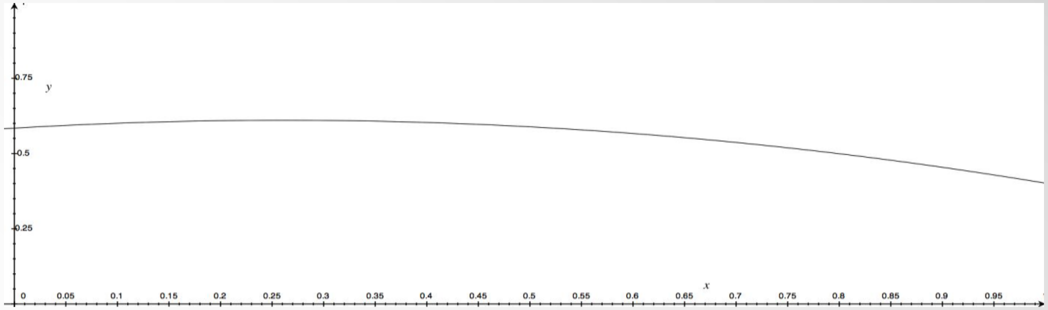
Simulation Examples

Simulation setup (cont.)

- The MS:

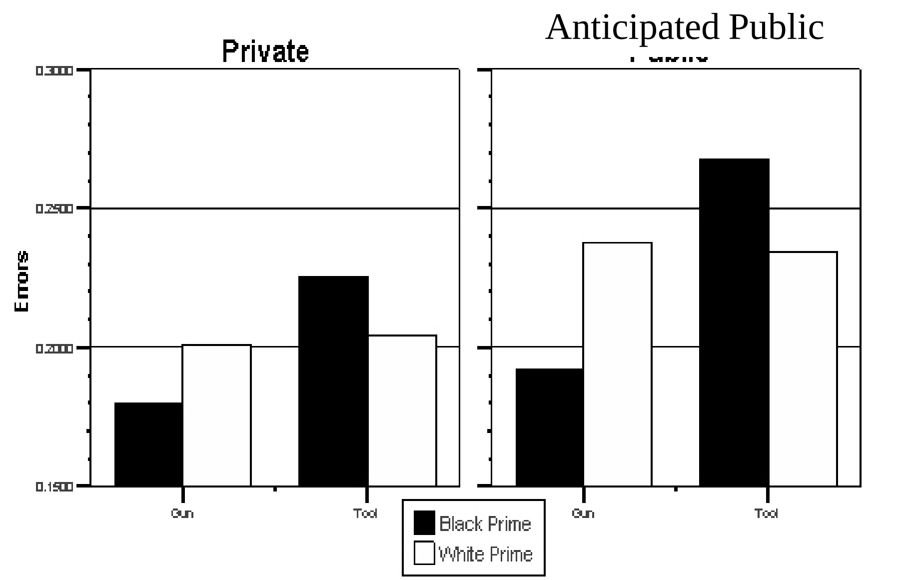


- The MCS: U-curve (Yerkes & Dodson, 1908)

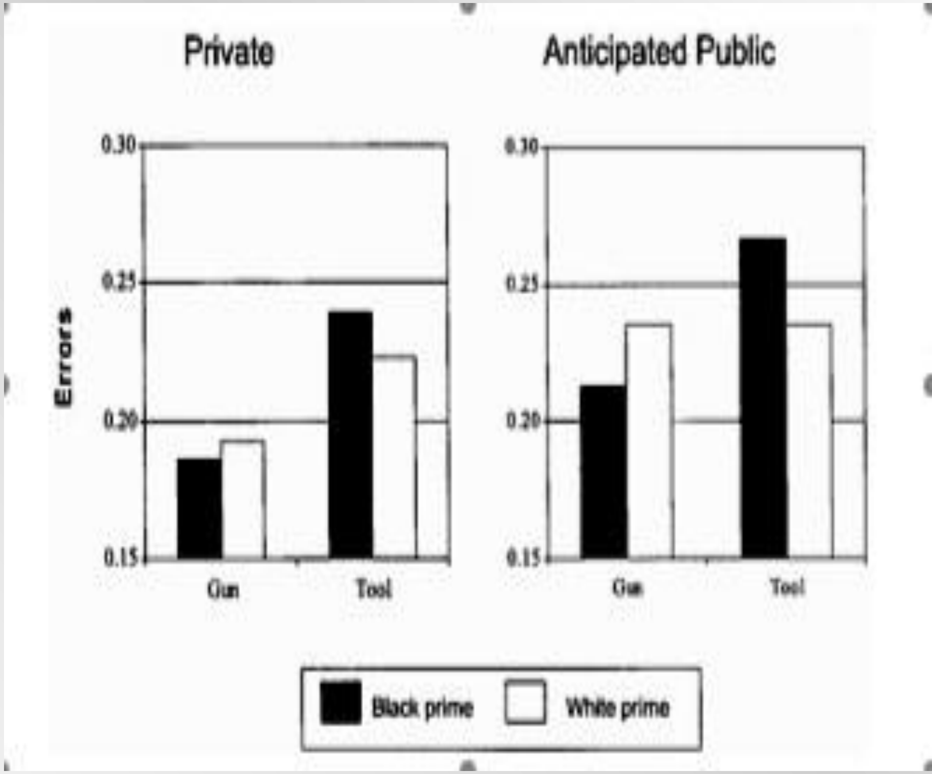


Simulation Examples

Simulation Results



Lambert Results



Simulation Examples

Simulation results (cont.)

Lambert Cognitive Control Estimate

GROUP	Black Prime	White Prime
Private	.61	.60
Public	.53	.53

Simulation Cognitive Control Estimate

GROUP	MCS	Black Prime	White Prime
Private	.60	.60	.60
Public	.53	.52	.52

Lambert et al
Accessibility Bias Estimate

GROUP	Black Prime	White Prime
Private	.56	.53
Public	.56	.49

Simulation Accessibility Bias Estimate

GROUP	ACS Black Prime	ACS White Prime	Black Prime	White Prime
Private	.57	.51	.57	.50
Public	.56	.51	.56	.51

Simulation Examples

Simulation results (cont.)

- Significant prime x object interaction in both groups
- Stronger prime x object interaction in public group than in private group
- Main effect of object
- Main effect of context over cognitive control estimates
- No effect of prime over cognitive control estimates
- Significant prime x context interaction over cognitive control estimates
- Main effect of prime over accessibility bias estimates
- No effect of context over accessibility bias estimates

Simulation Examples

Questions about the above simulation?

Summary

1. Motivational Subsystem
    1. Introduction
    2. Drives
        1. Low-level Primary Drives
        2. High-level Primary Drives
        3. Drive Strength Considerations
    3. Goal Structure
        1. Goal List
        2. Goal Stack
2. Meta-Cognitive Subsystem
    1. Introduction
    2. Structure & Responsibilities
3. Simulation Examples
4. **Summary**

Summary

The MS:

- A cognitive agent's actions must be sustainable, purposeful, focused, and adaptive
- In CLARION, these requirements are captured by drive activations at the bottom-level and a goal structure at the top level
- Low-level primary drives are mostly physiological, hard-wired, and evolutionarily formed
- High-level primary drives are more socially oriented, and likely hardwired to some extent also

Summary

The MS (cont.)

- Drive strengths are determined based on considerations of proportional activation, opportunism, contiguity of actions, persistence, interruption when necessary, and combination of preferences
- Goals provide specific, tangible motivations for actions (to be used by the ACS)
- The goal structure communicates motivation to the other subsystems

Summary

The MCS:

- Includes modules for:
  - Goal-setting
  - Reinforcement
  - Filtering
  - Learning and reasoning
  - Level selection/integration
  - Parameter setting
  - Monitoring buffer
  - Etc.

Summary

Thank You

Questions?