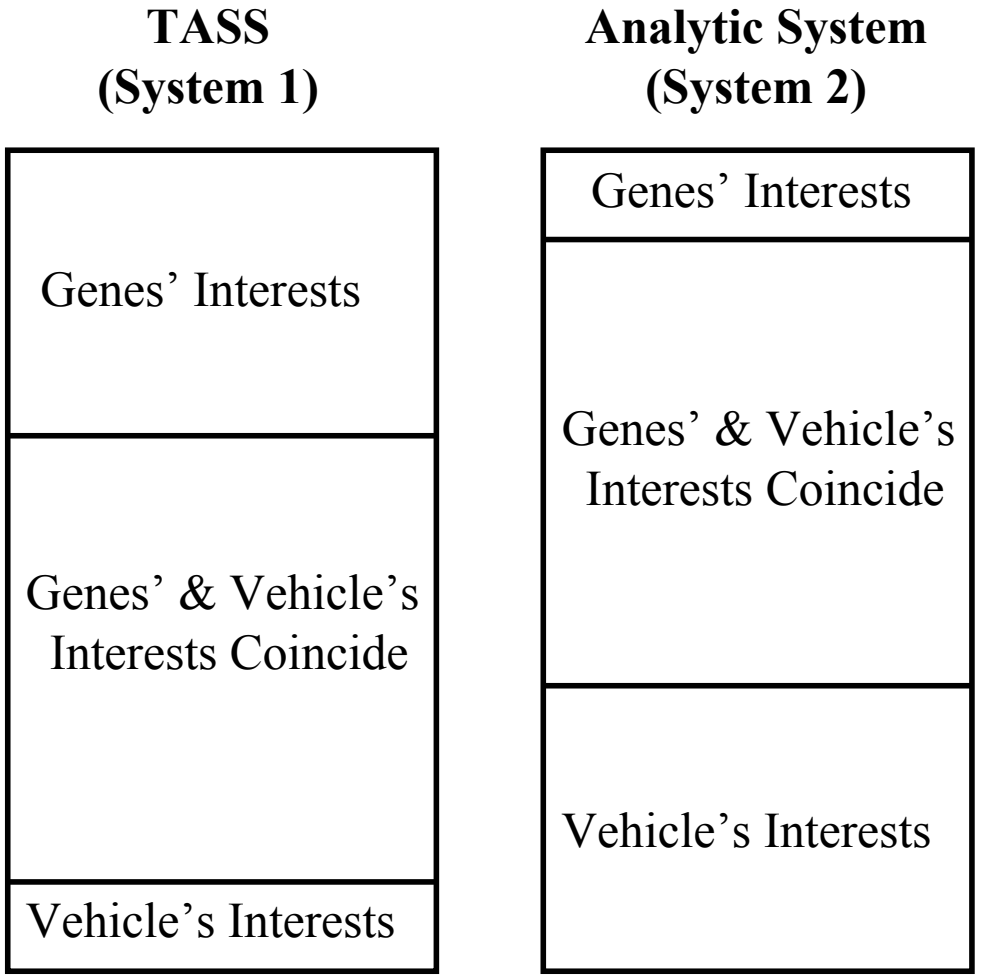
Is It Time for a Tri-Process Theory? Distinguishing the Reflective and the Algorithmic Mind

Keith E. Stanovich
University of Toronto

Missing Components of Current Dual-Process Theories and Their Implications for the Rationality Debate

Goal Structure



Goals reflecting genetic and vehicle interests
in TASS and in the analytic system

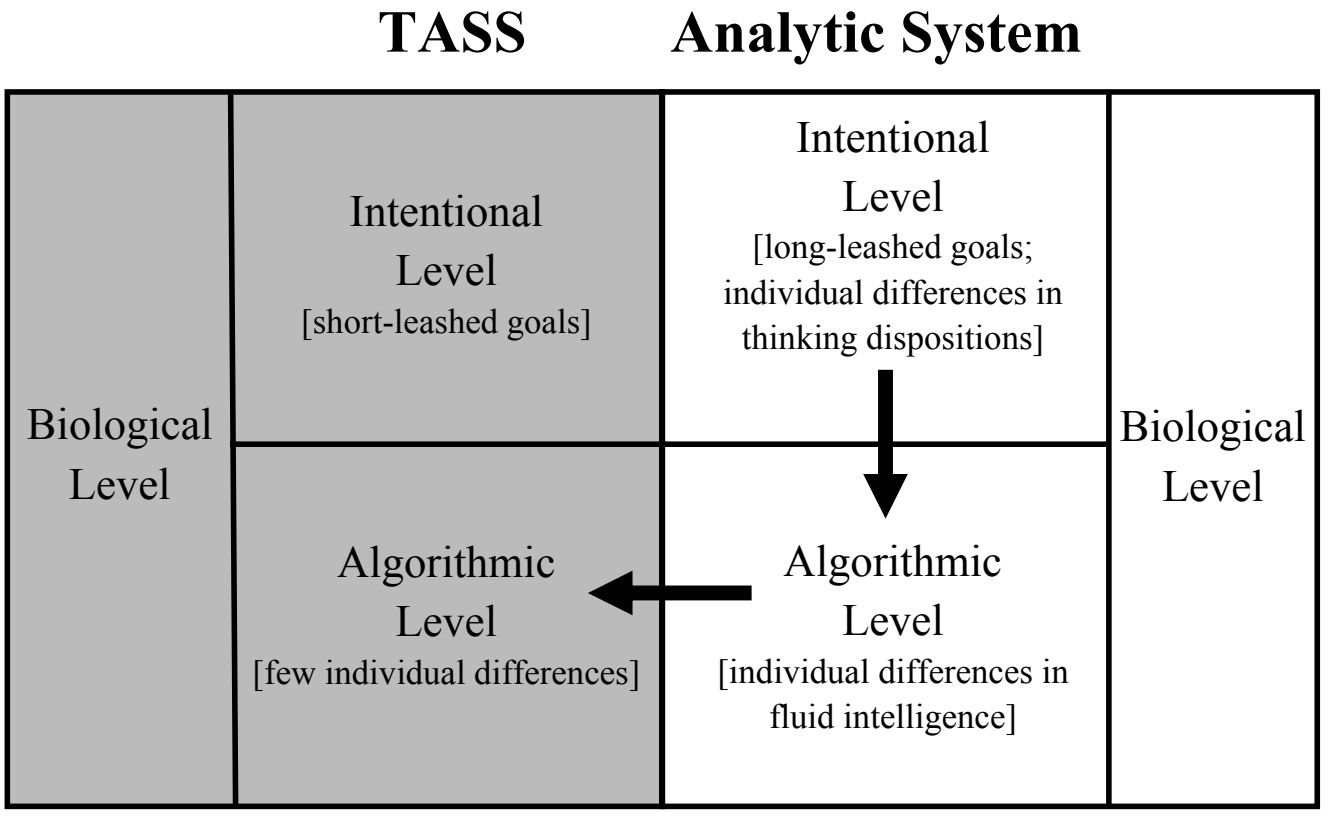
Evolutionary Reinterpretations of Heuristics and Biases Tasks

Task	Normative Response	Adaptive Response
1. Wason Selection Task	P & ~Q	P & Q
2. Linda Problem	P(A&B) < P(A)	P(A&B) > P(A)
3. Covariation Detection	w(cell A) = w(cell D)	w(cell A) > w(cell D)
4. Probability Learning	maximizing	prob matching
5. Argument Evaluation Tasks	belief bias = 0	belief bias > 0

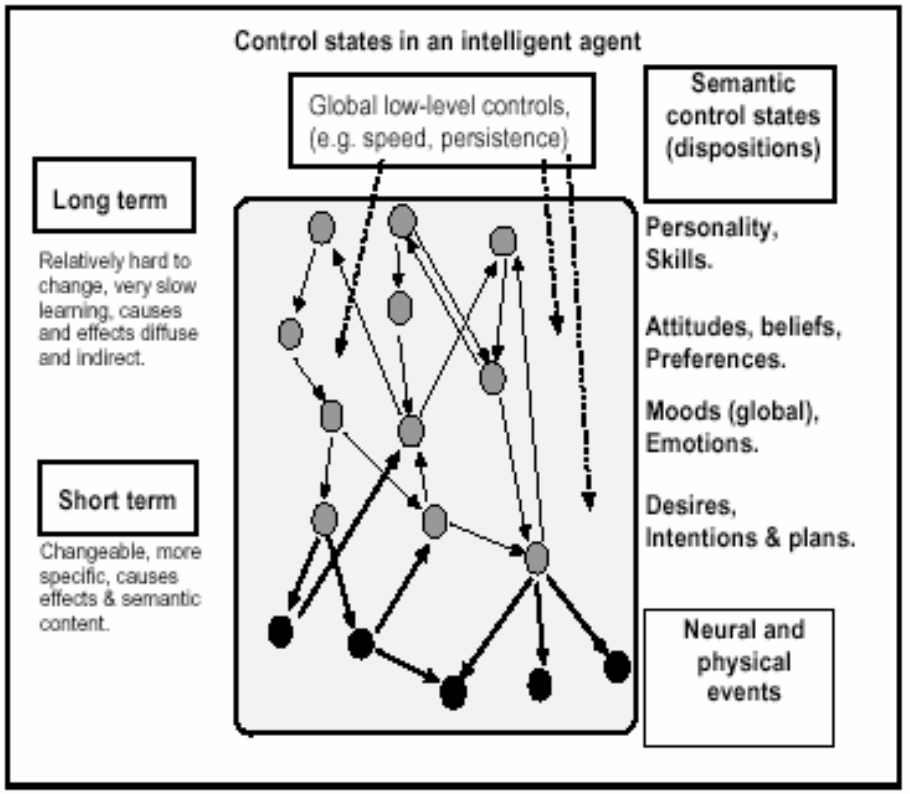
Various Processes in TASS: The Autonomous Set of Systems

- processes of implicit learning
- overlearned associations practiced to automaticity
- processes of behavioral regulation by the emotions
- processes of classical and operant conditioning
- encapsulated modules

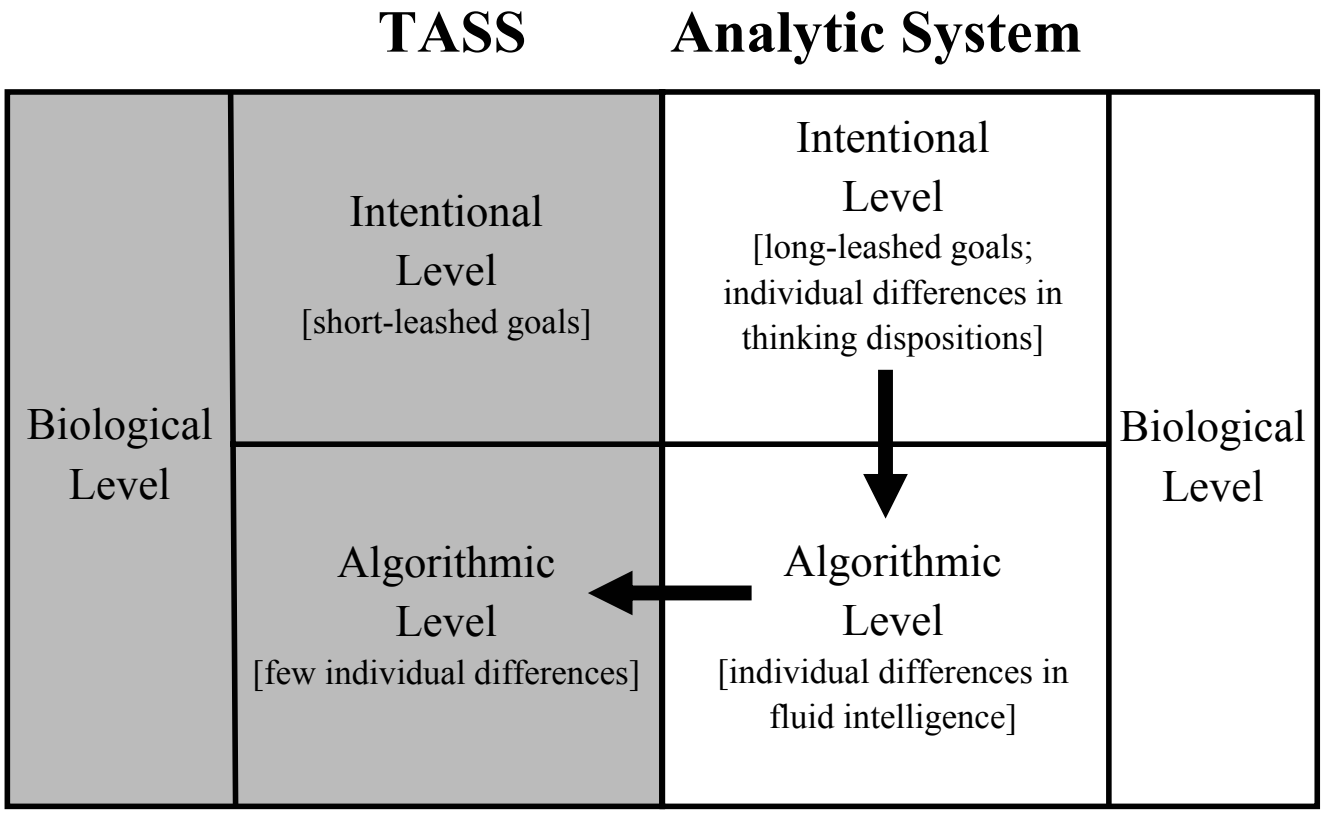
Control Structures



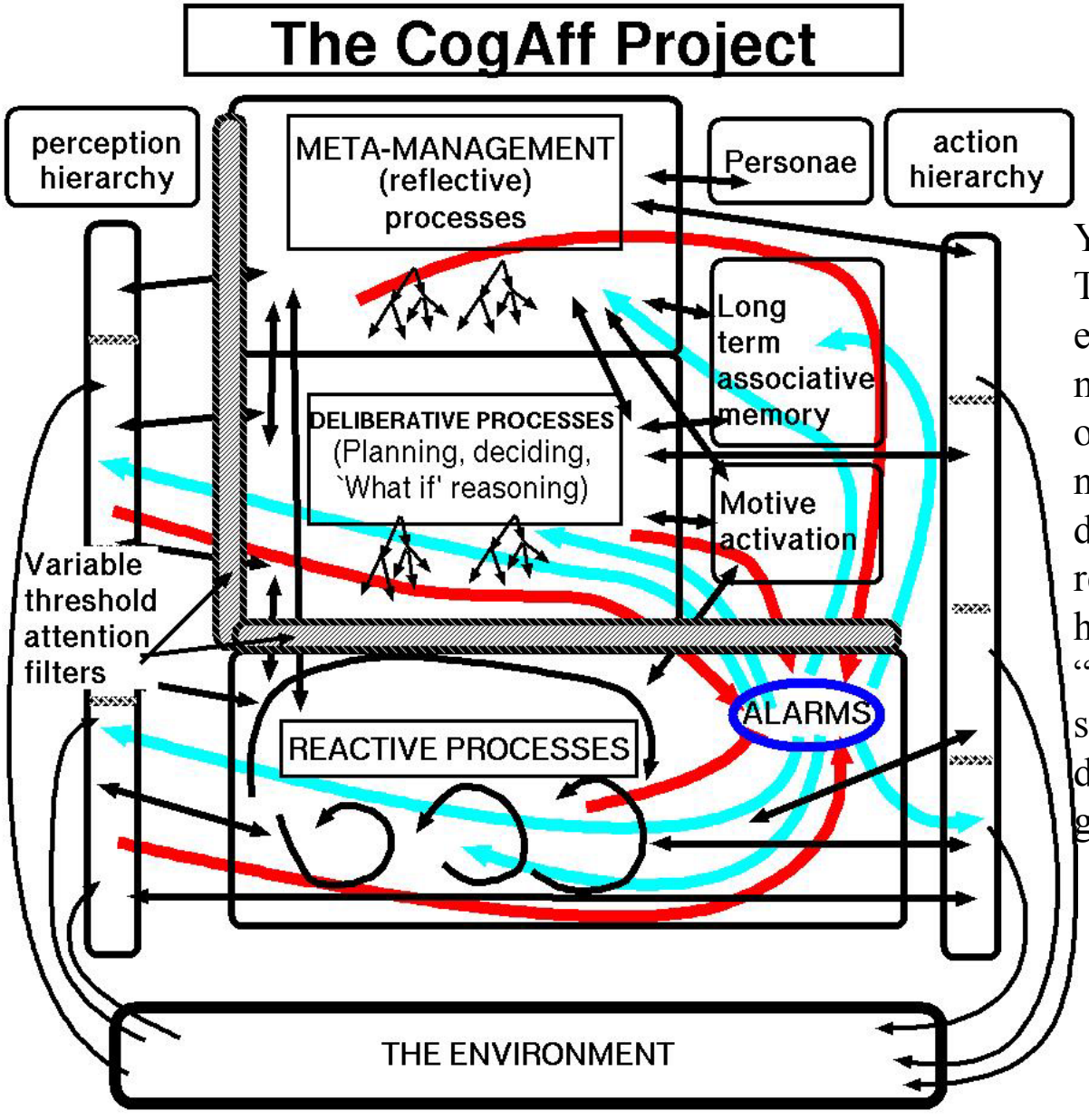
Processing Control in TASS Override by the Analytic System


Control states in an intelligent agent
Global low-level controls, (e.g. speed, persistence)
Semantic control states (dispositions)
Long term
Relatively hard to change, very slow learning, causes and effects diffuse and indirect.
Short term
Changeable, more specific, causes effects & semantic content.
Personality, Skills.
Attitudes, beliefs, Preferences.
Moods (global), Emotions.
Desires, Intentions & plans.
Neural and physical events

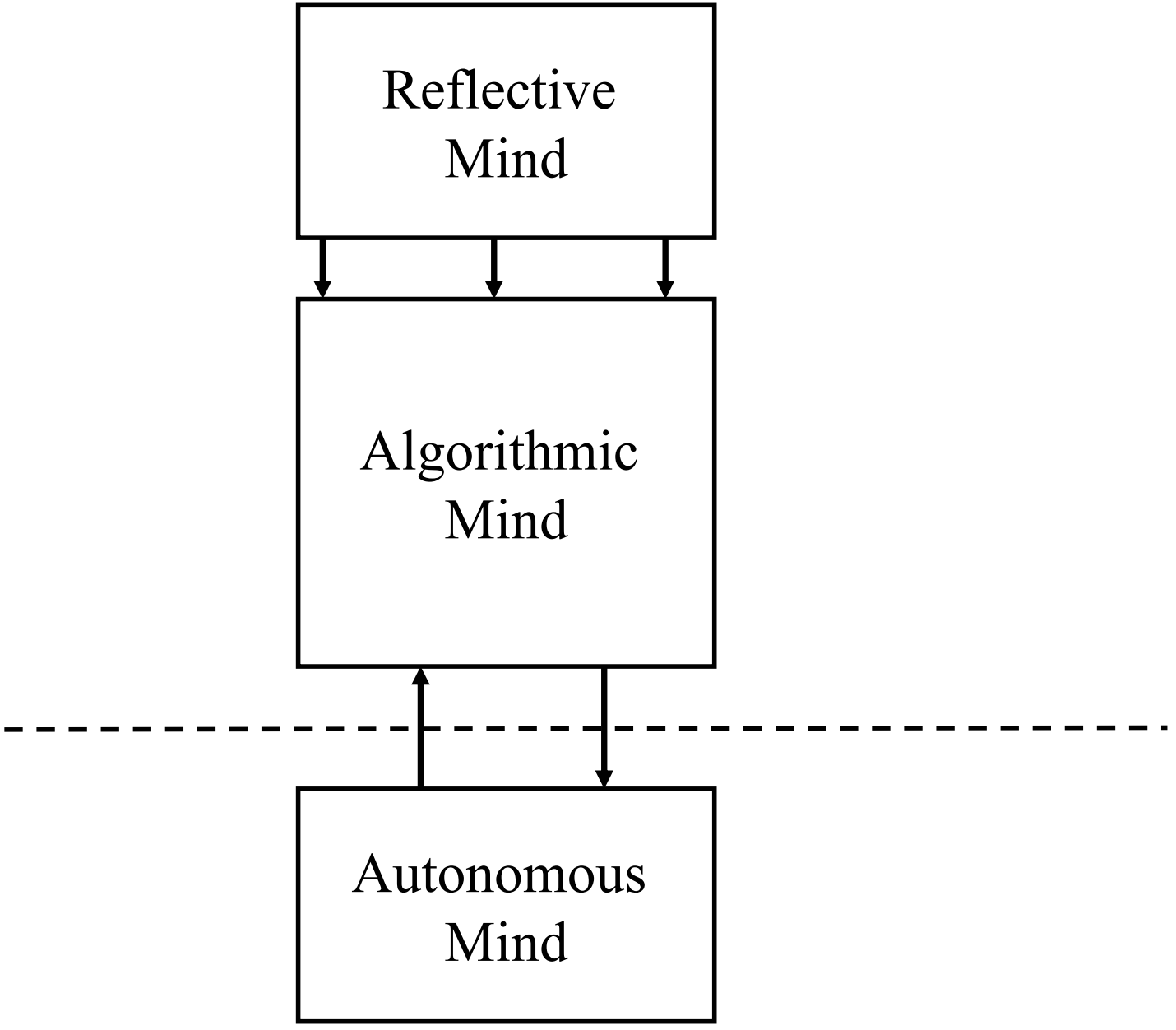
Control Structures

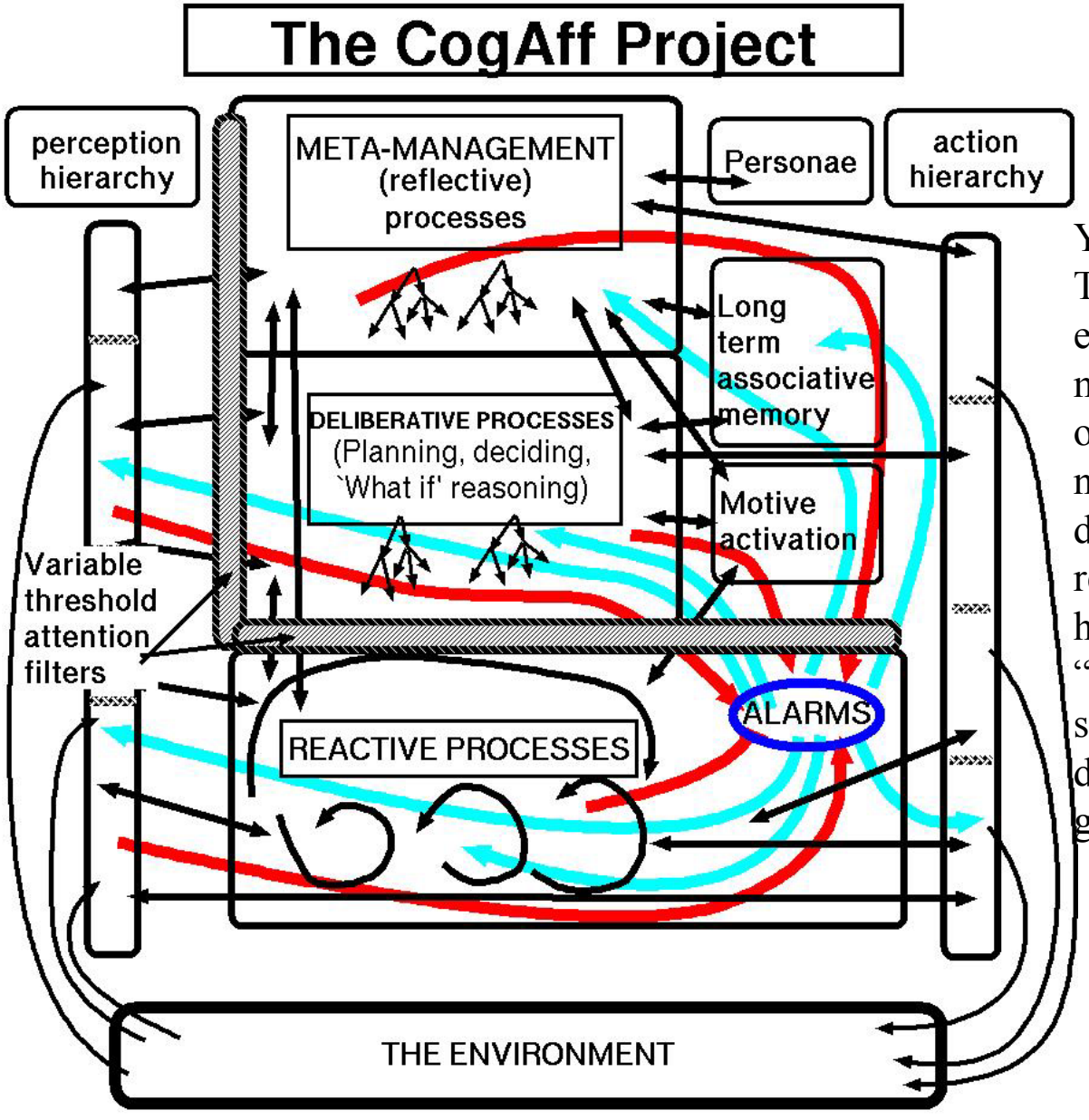


Processing Control in TASS Override by the Analytic System

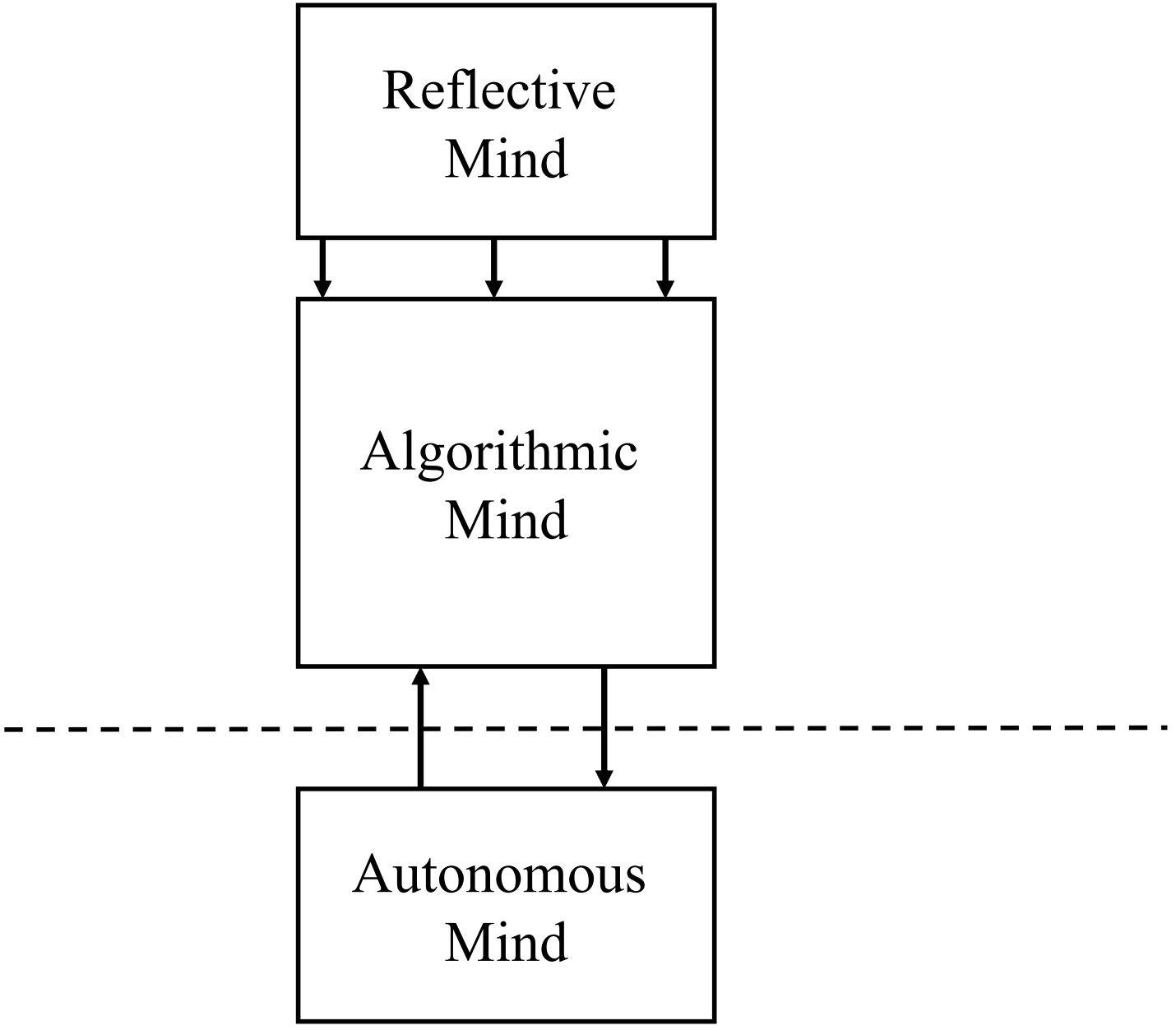


Your Mind (A. Sloman, 2003). The framework incorporates evolutionarily ancient mechanisms co-existing and co-operating or competing with new mechanisms capable of doing different tasks (e.g., reasoning about what might happen). The figure gives an "impressionistic" overview of some of the complexity (e,g., different sorts of emotions are generated at different levels).


Reflective Mind
Algorithmic Mind
Autonomous Mind



Your Mind (A. Sloman, 2003). The framework incorporates evolutionarily ancient mechanisms co-existing and co-operating or competing with new mechanisms capable of doing different tasks (e.g., reasoning about what might happen). The figure gives an "impressionistic" overview of some of the complexity (e,g., different sorts of emotions are generated at different levels).


Reflective Mind
Algorithmic Mind
Autonomous Mind

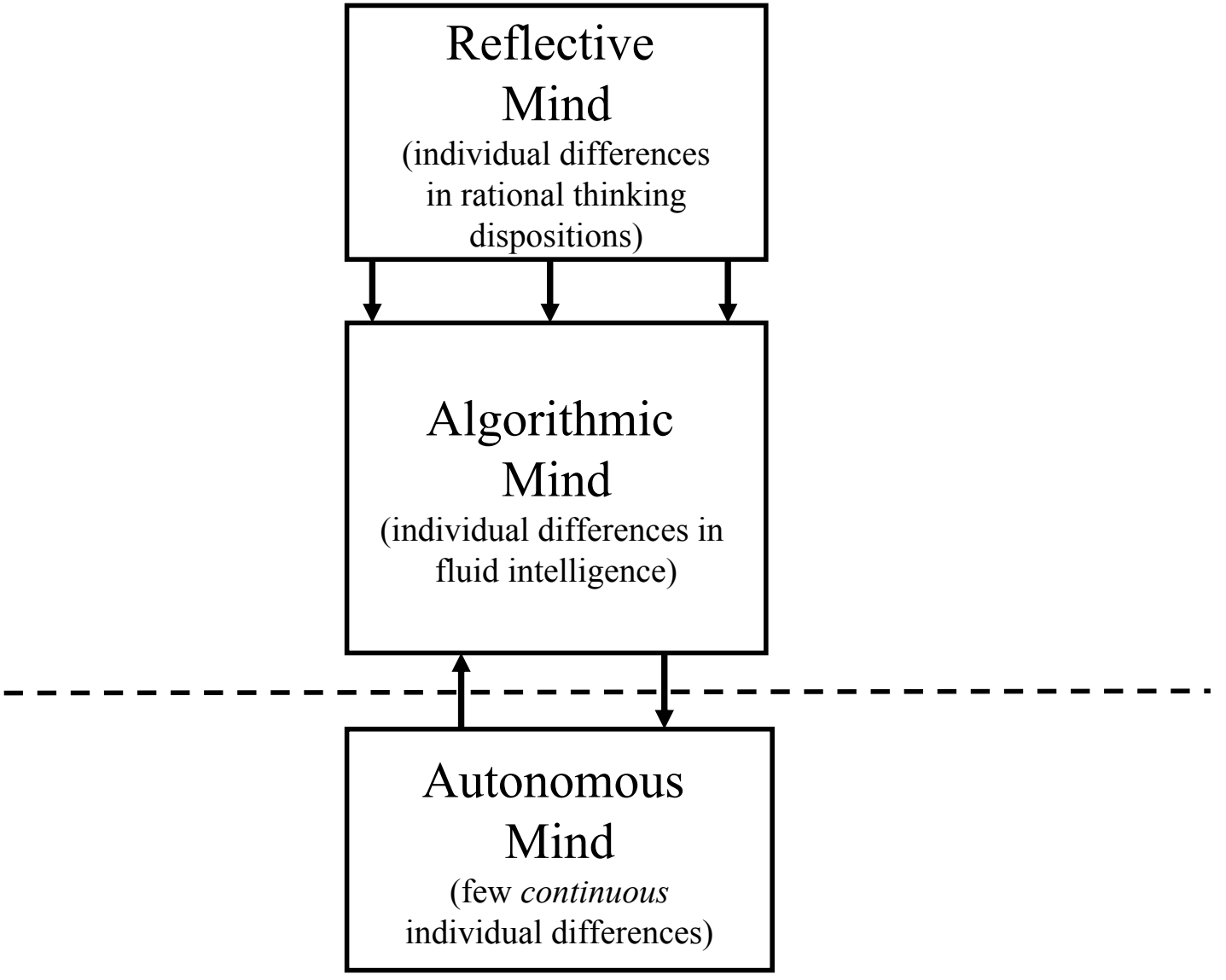
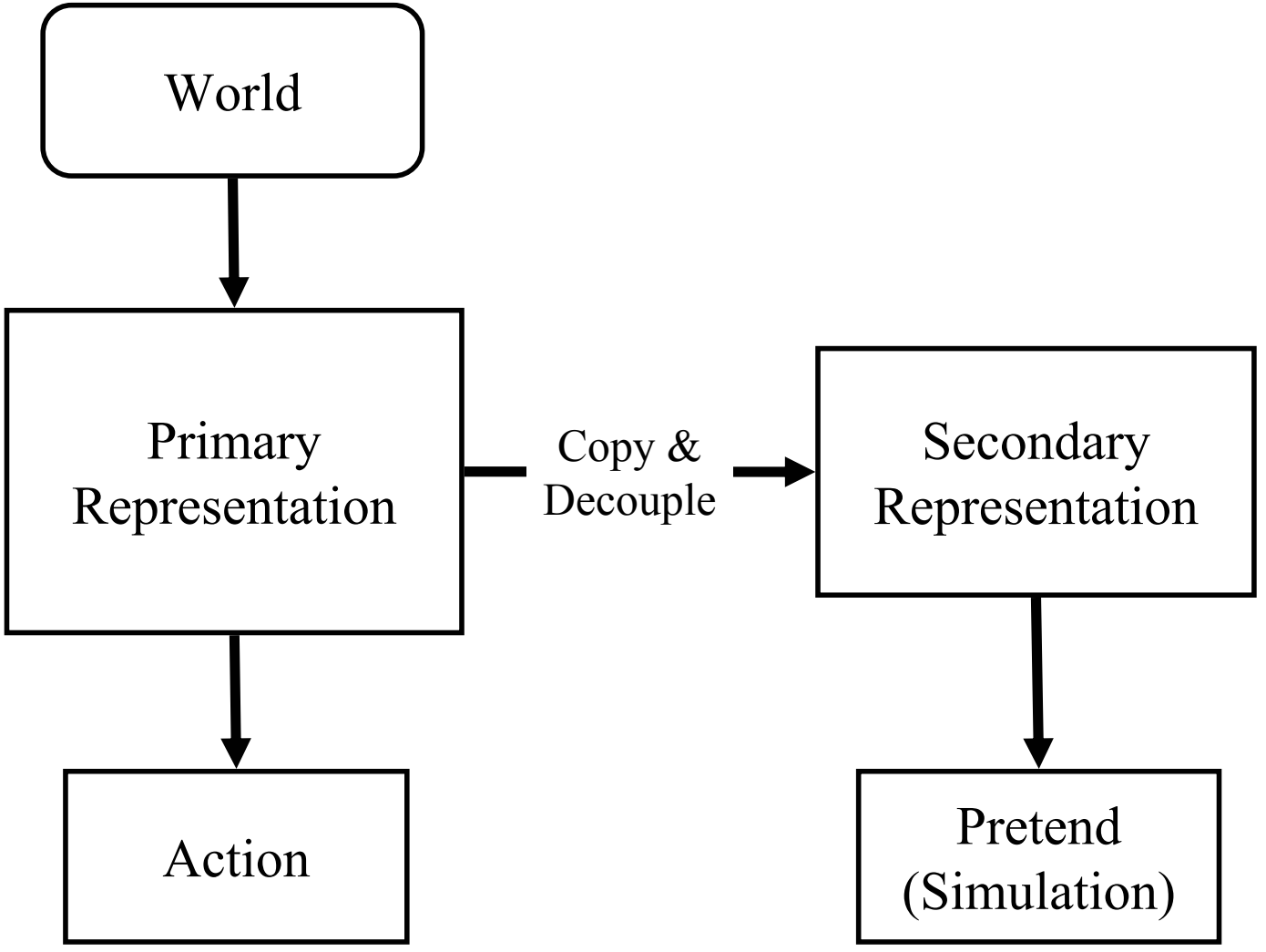


Figure 5. Individual Differences in the Tripartite Structure



Adapted from Leslie (1987)

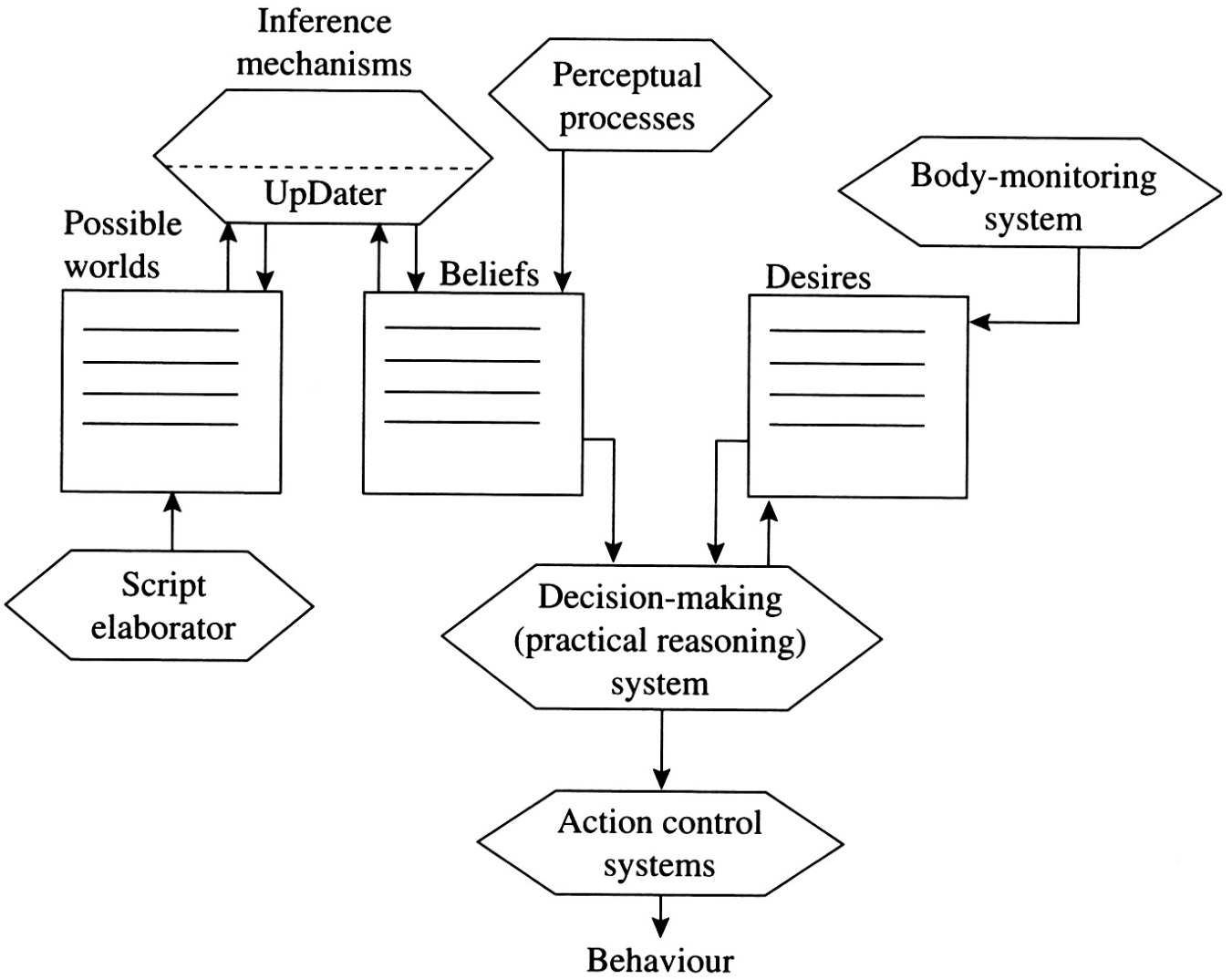
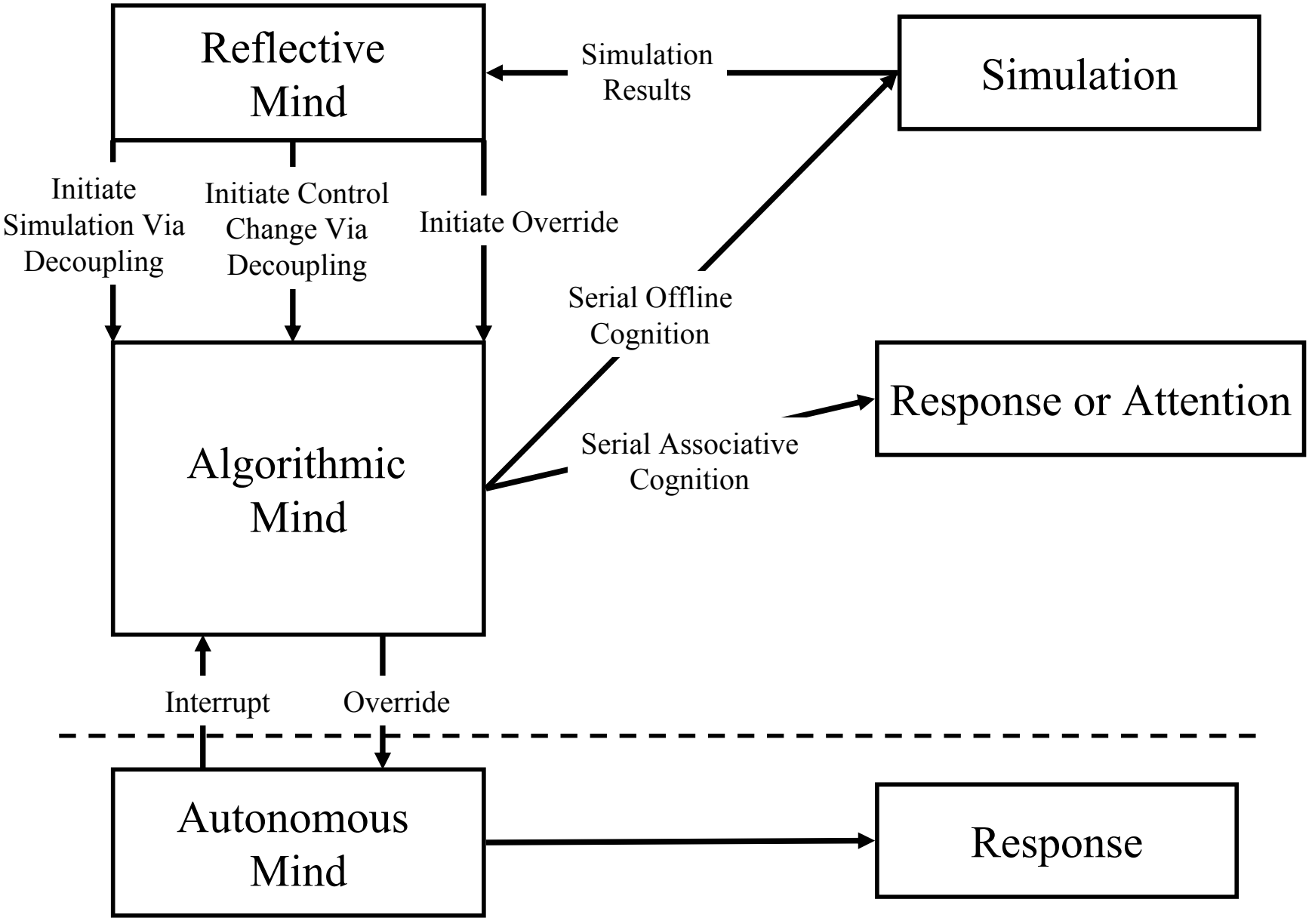


Fig. 2.2. The mental mechanisms posited by Nichols and Stich (2003)


Reflective Mind
Simulation Results
Simulation
Initiate Simulation Via Decoupling
Initiate Control Change Via Decoupling
Initiate Override
Serial Offline Cognition
Algorithmic Mind
Response or Attention
Serial Associative Cognition
Interrupt
Override
Autonomous Mind
Response

Humans as Cognitive Misers
(Robyn Dawes)

"Following Dawes (1976), some favored the metaphor of the "cognitive miser" by emphasizing limited mental resources, reliance on irrelevant cues, and the difficulties of effortful correction" (Krueger & Funder, 2004, pp. 316-317)

The rule that human beings seem to follow is to engage the brain only when all else fails--and usually not even then

-- David Hull, *Science and Selection: Essays on Biological Evolution and the Philosophy of Science*, 2001, p. 37

Humans as Cognitive Misers

Stage 1: Default to TASS

Stage 2: Display a focal bias and rely on serial associative cognition

Focal Bias

relates to:

singularity principle (Evans, Over, & Handley, 2003)

principle of truth (Johnson-Laird)

focussing effects (Legrenzi, Girotto, & Johnson-Laird, 1993)

effect and effort in determining relevance (Sperber, Cara, & Girotto, 1995)

automatic belief acceptance (Gilbert, 1991)

focalism in social psychological theory (Wilson et al., 2000)

When the information processor is strongly disposed to deal only with the most easily constructed cognitive model, then a focal bias is being demonstrated.

The most easily constructed model tends to represent:

only one state of affairs (Evans, Over, & Handley, 2003)

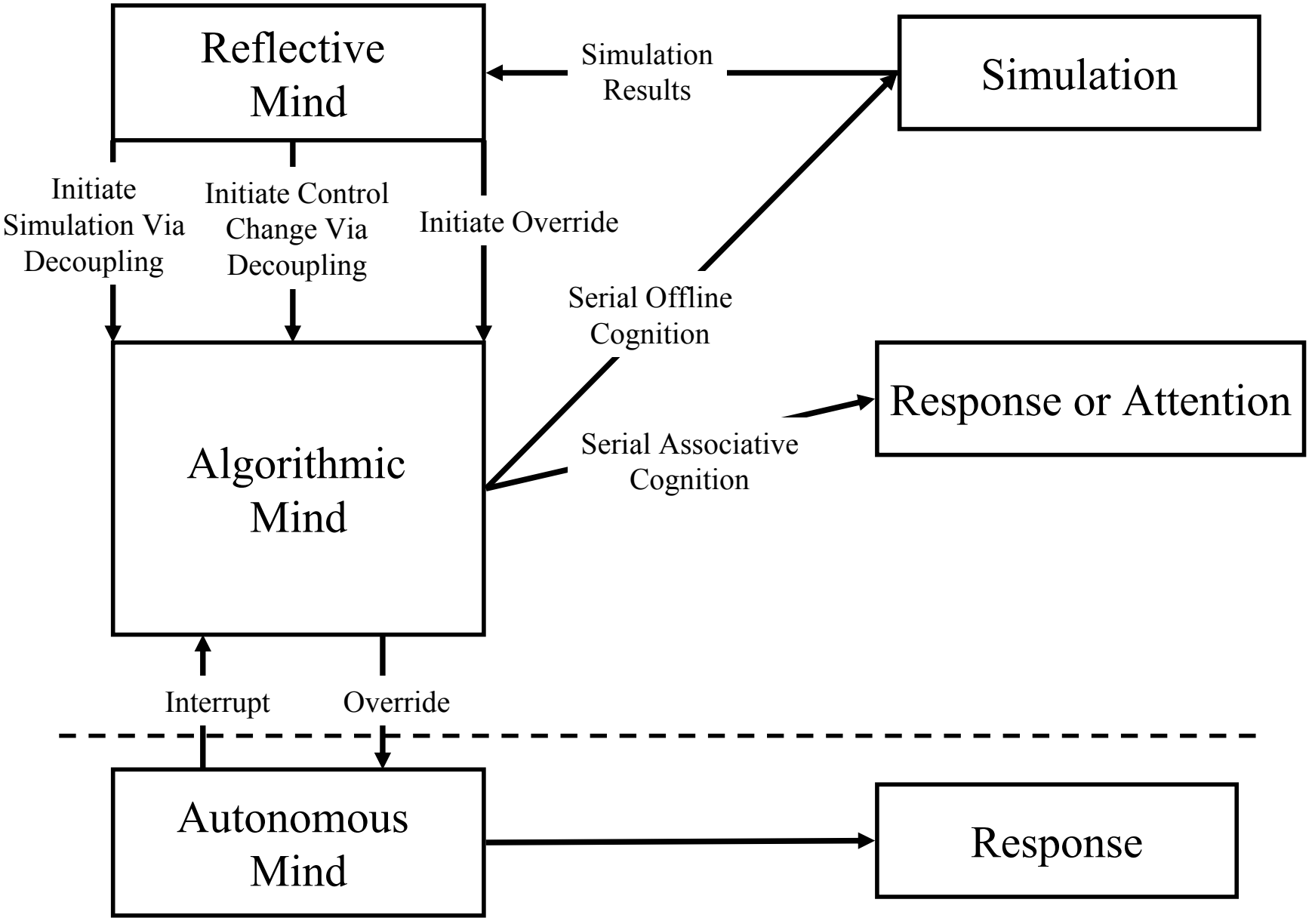
that is modelled as true (Johnson-Laird)

and/or is accepted as given (Gilbert, 1991; matching bias)

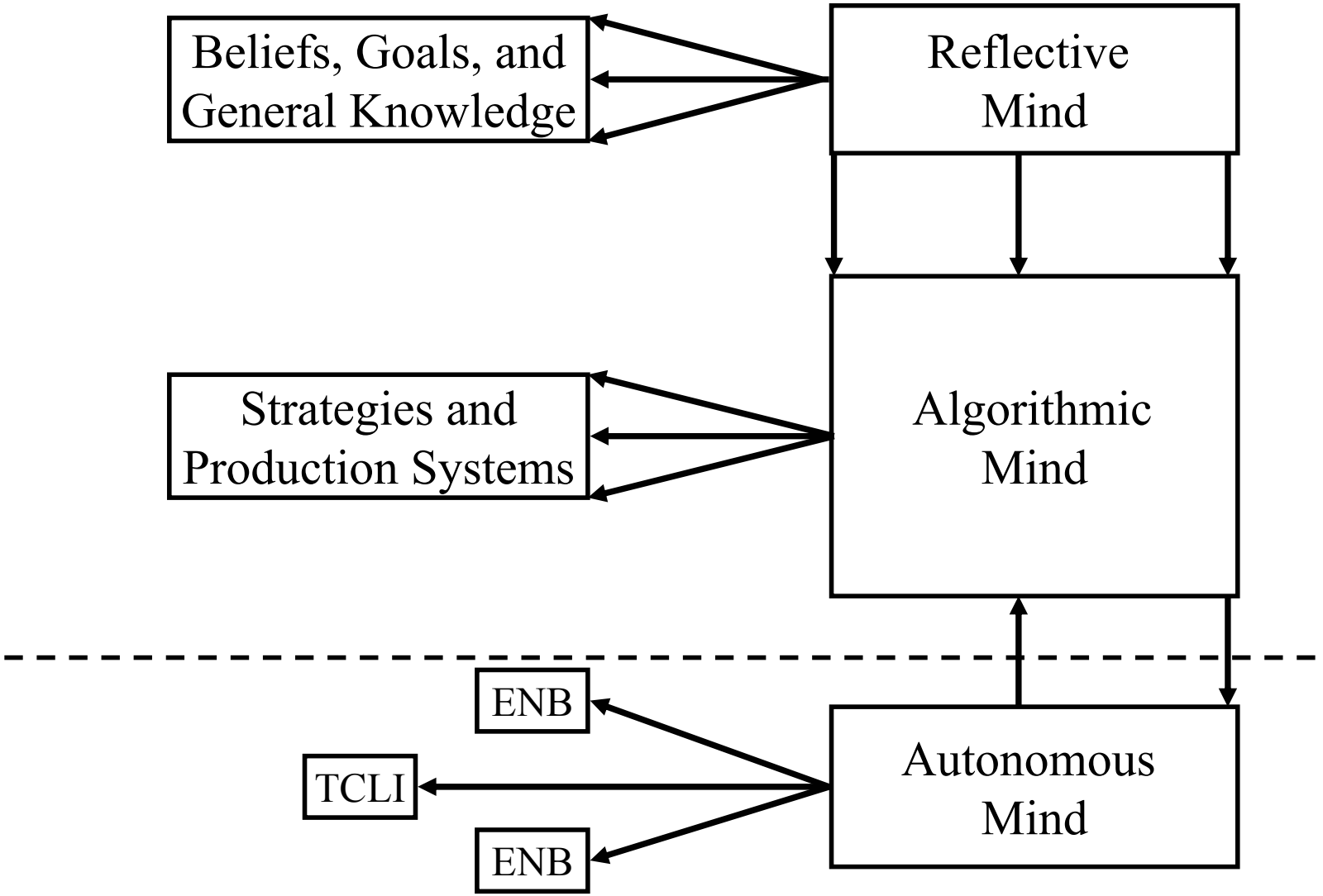
that is closest to what we already believe in (myside bias, belief bias)

that minimizes effort (Sperber)

that ignores moderating factors (which might necessitate the construction of alternative models; Wilson et al., 2000)


Reflective Mind
Simulation Results
Simulation
Initiate Simulation Via Decoupling
Initiate Control Change Via Decoupling
Initiate Override
Serial Offline Cognition
Algorithmic Mind
Response or Attention
Serial Associative Cognition
Interrupt
Override
Autonomous Mind
Response

Knowledge Structures

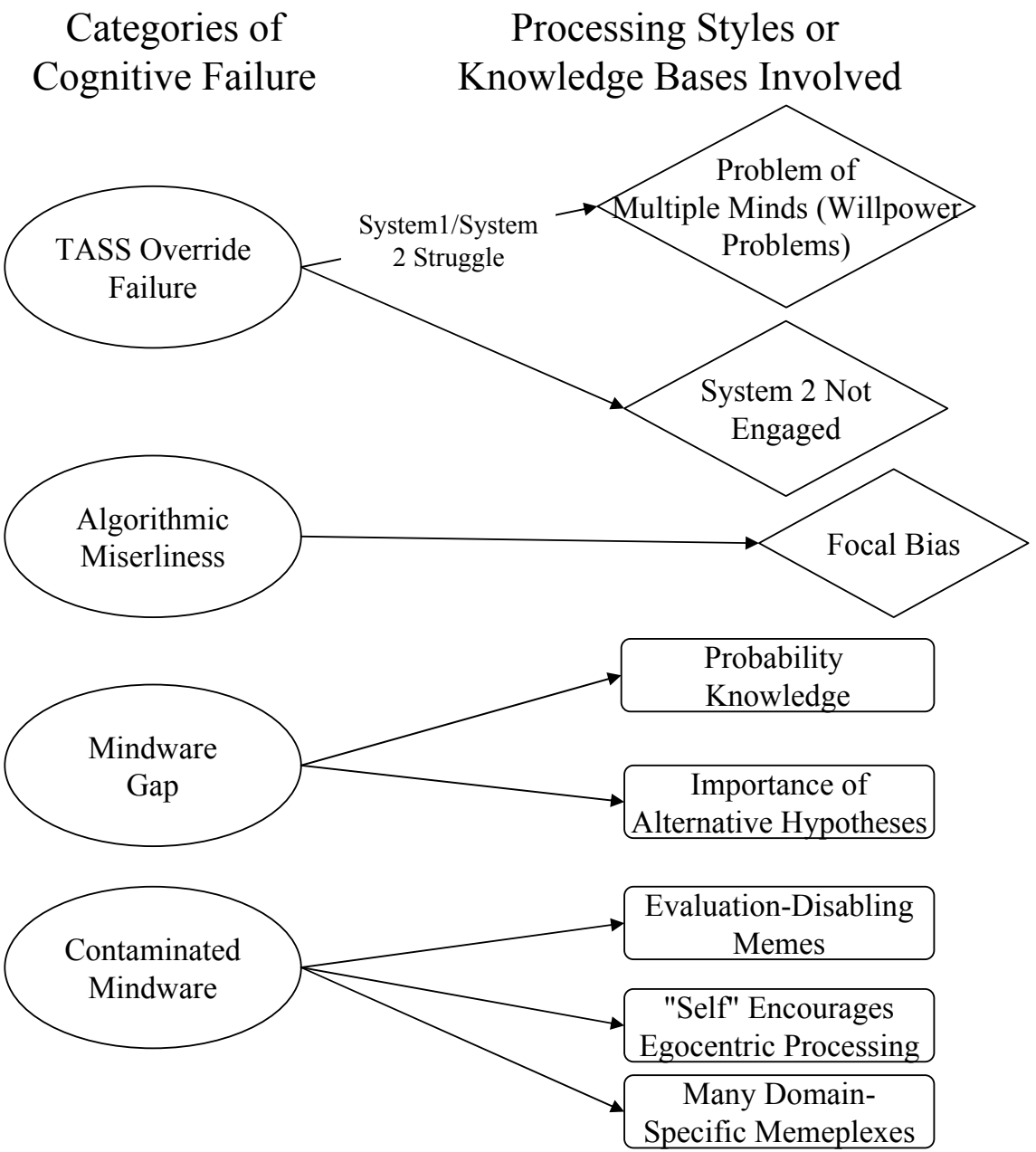


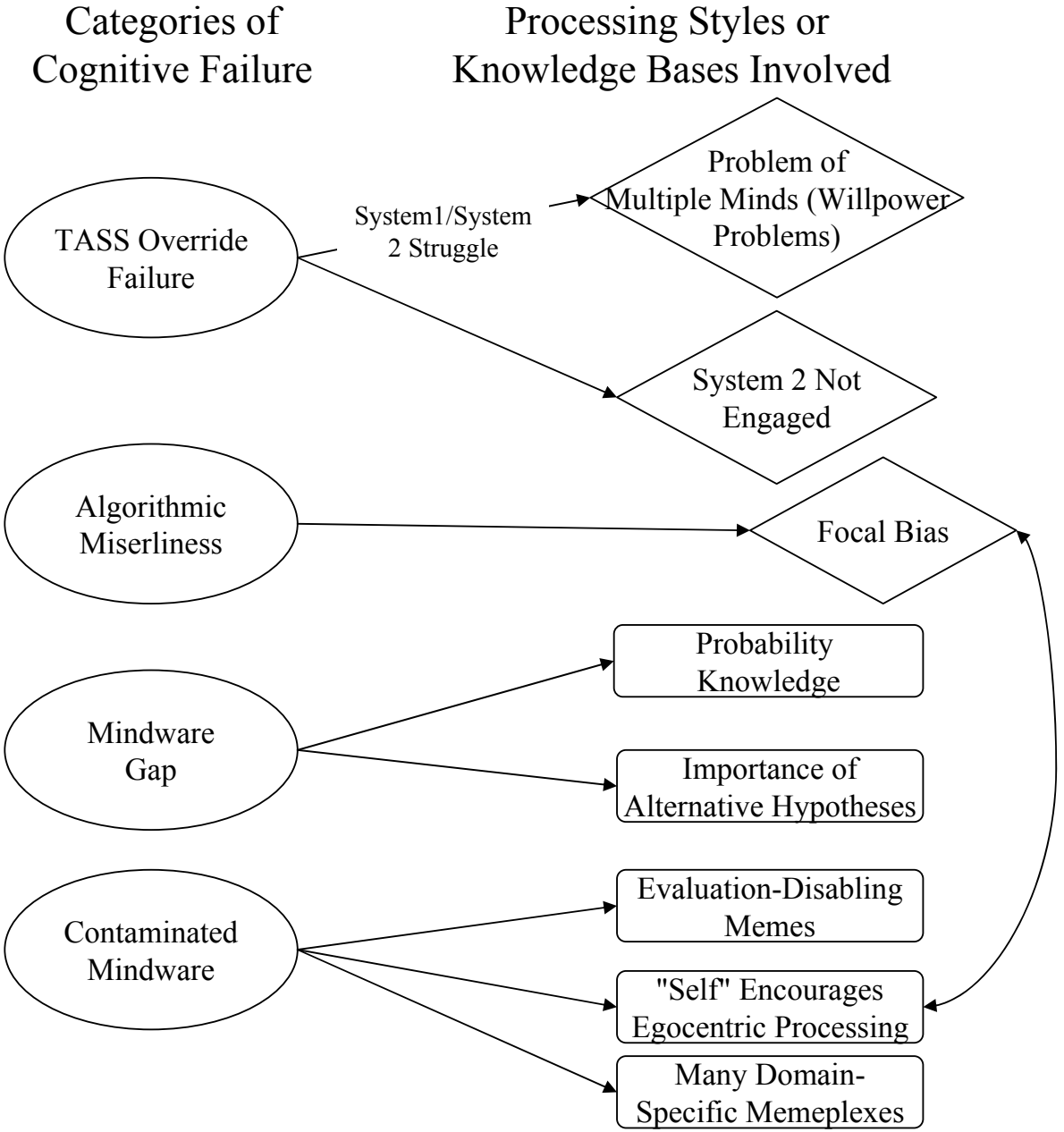
ENB = Encapsulated Knowledge Base

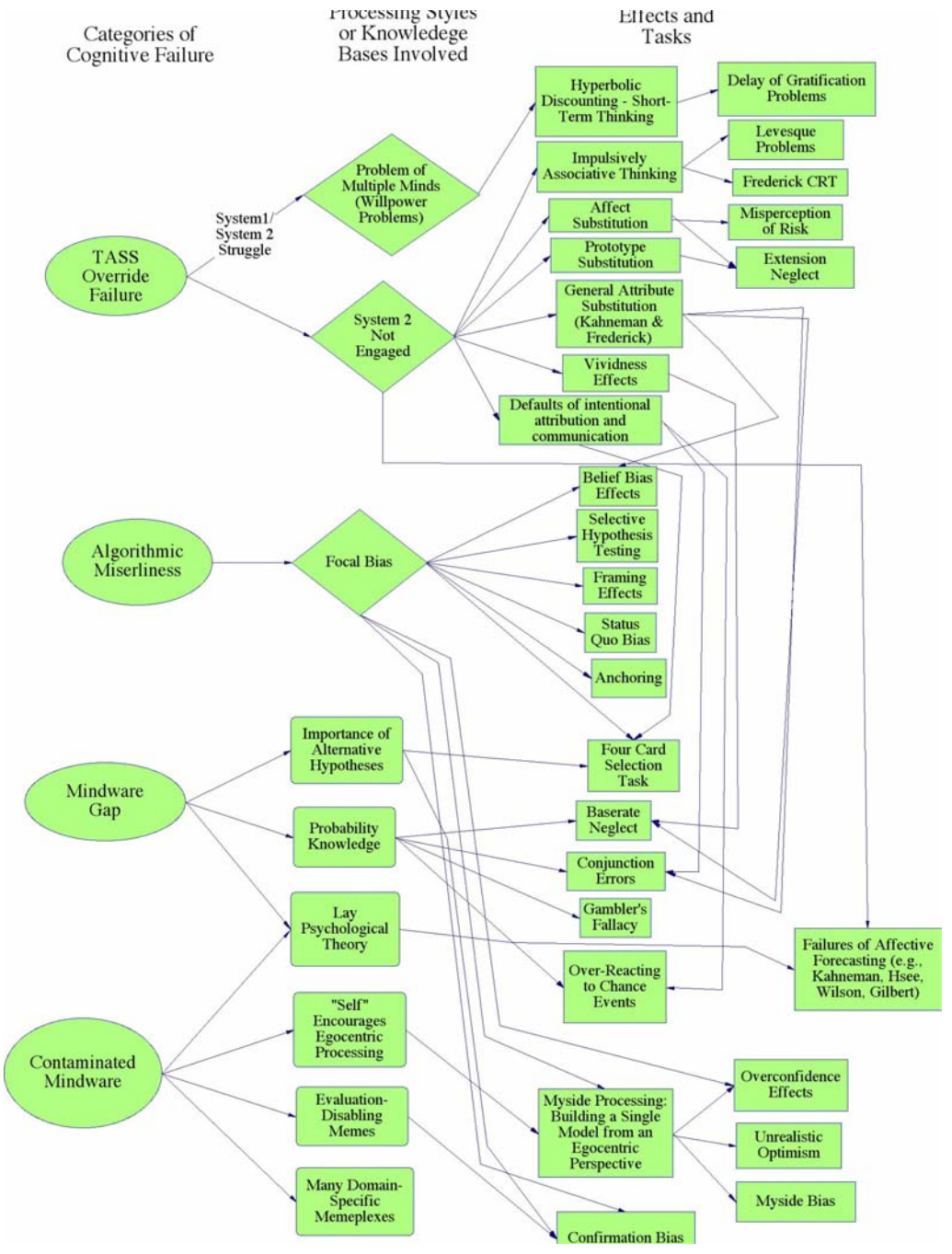
TCLI = Tightly Compiled Learned Information

Types of Cognitive Failure

1. TASS override failure
2. mindware gaps
3. contaminated mindware
4. miserliness in the algorithmic mind


Categories of Cognitive Failure
Processing Styles or Knowledge Bases Involved
TASS Override Failure
System1/System 2 Struggle
Problem of Multiple Minds (Willpower Problems)
System 2 Not Engaged
Algorithmic Miserliness
Focal Bias
Mindware Gap
Probability Knowledge
Importance of Alternative Hypotheses
Contaminated Mindware
Evaluation-Disabling Memes
"Self" Encourages Egocentric Processing
Many Domain-Specific Memeplexes


Categories of Cognitive Failure
Processing Styles or Knowledge Bases Involved
TASS Override Failure
System1/System 2 Struggle
Problem of Multiple Minds (Willpower Problems)
System 2 Not Engaged
Algorithmic Miserliness
Focal Bias
Mindware Gap
Probability Knowledge
Importance of Alternative Hypotheses
Contaminated Mindware
Evaluation-Disabling Memes
"Self" Encourages Egocentric Processing
Many Domain-Specific Memeplexes


Categories of Cognitive Failure
Processing Styles or Knowledege Bases Involved
Effects and Tasks
TASS Override Failure
System1/ System 2 Struggle
Problem of Multiple Minds (Willpower Problems)
System 2 Not Engaged
Algorithmic Miserliness
Focal Bias
Mindware Gap
Importance of Alternative Hypotheses
Probability Knowledge
Lay Psychological Theory
Contaminated Mindware
"Self" Encourages Egocentric Processing
Evaluation-Disabling Memes
Many Domain-Specific Memeplexes
Hyperbolic Discounting - Short-Term Thinking
Impulsively Associative Thinking
Affect Substitution
Prototype Substitution
General Attribute Substitution (Kahneman & Frederick)
Vividness Effects
Defaults of intentional attribution and communication
Belief Bias Effects
Selective Hypothesis Testing
Framing Effects
Status Quo Bias
Anchoring
Four Card Selection Task
Baserate Neglect
Conjunction Errors
Gambler's Fallacy
Over-Reacting to Chance Events
Myside Processing: Building a Single Model from an Egocentric Perspective
Confirmation Bias
Delay of Gratification Problems
Levesque Problems
Frederick CRT
Misperception of Risk
Extension Neglect
Failures of Affective Forecasting (e.g., Kahneman, Hsee, Wilson, Gilbert)
Overconfidence Effects
Unrealistic Optimism
Myside Bias

Frankfurt, H. (1971). Freedom of the will and the concept of a person. Journal of Philosophy, 68, 5-20.

Jeffrey, R. (1974). Preferences among preferences. Journal of Philosophy, 71, 377-391.

Imagine John, who has the choice between smoking (S) and not smoking (~S).
John chooses to smoke.
So we have:

John prefers to smoke

S pref ~S

John prefers to prefer not to smoke:

(~S pref S) pref (S pref ~S)

He prefers his preference to prefer not to smoke over his preference for smoking:

[(~S pref S) pref (S pref ~S)]
pref
[S pref ~S]

We might in this case say that John's third-order judgment has ratified his second-order strong evaluation

On the other hand, a third-order judgment might undermine the second-order preference by failing to ratify it:

John might prefer to smoke more than
he prefers his preference to prefer not to smoke

[S pref ~S]
pref
[(~S pref S) pref (S pref ~S)]

Velleman, J. D. (1992). What happens when somebody acts? Mind, 101, 461-481.

“your desire to act in accordance with reasons, a desire that produces behavior, in your name, by adding its motivational force to that of whichever motives appear to provide the strongest reasons for acting” (Velleman, 1992, p. 479).

The Master Rationality Motive

The desire to act in accordance with reasons.

I like to gather many different types of evidence before I decide what to do.

I like to have reasons for what I do.

I always consider the consequences before I take action.

I am only confident of decisions that are made after careful analysis of all available information.

I don't like to have to justify my actions. (R)

If a belief suits me then I am comfortable, it really doesn't matter if the belief is true. (R)

After I make a decision, it is often difficult for me to give logical reasons for it. (R)

I don't feel I have to have reasons for what I do. (R)

"As soon as language acted as a vehicle for delivering information into the mind (whether one's own or that of another person), carrying with it snippets of non-social information, a transformation in the nature of the mind began….language switched from a social to a general-purpose function, consciousness from a means to predict other individuals' behavior to managing a mental database of information relating to all domains of behavior" (p. 209)

Mithen, S. (1996). *The prehistory of mind: The cognitive origins of art and science*. London: Thames and Hudson.

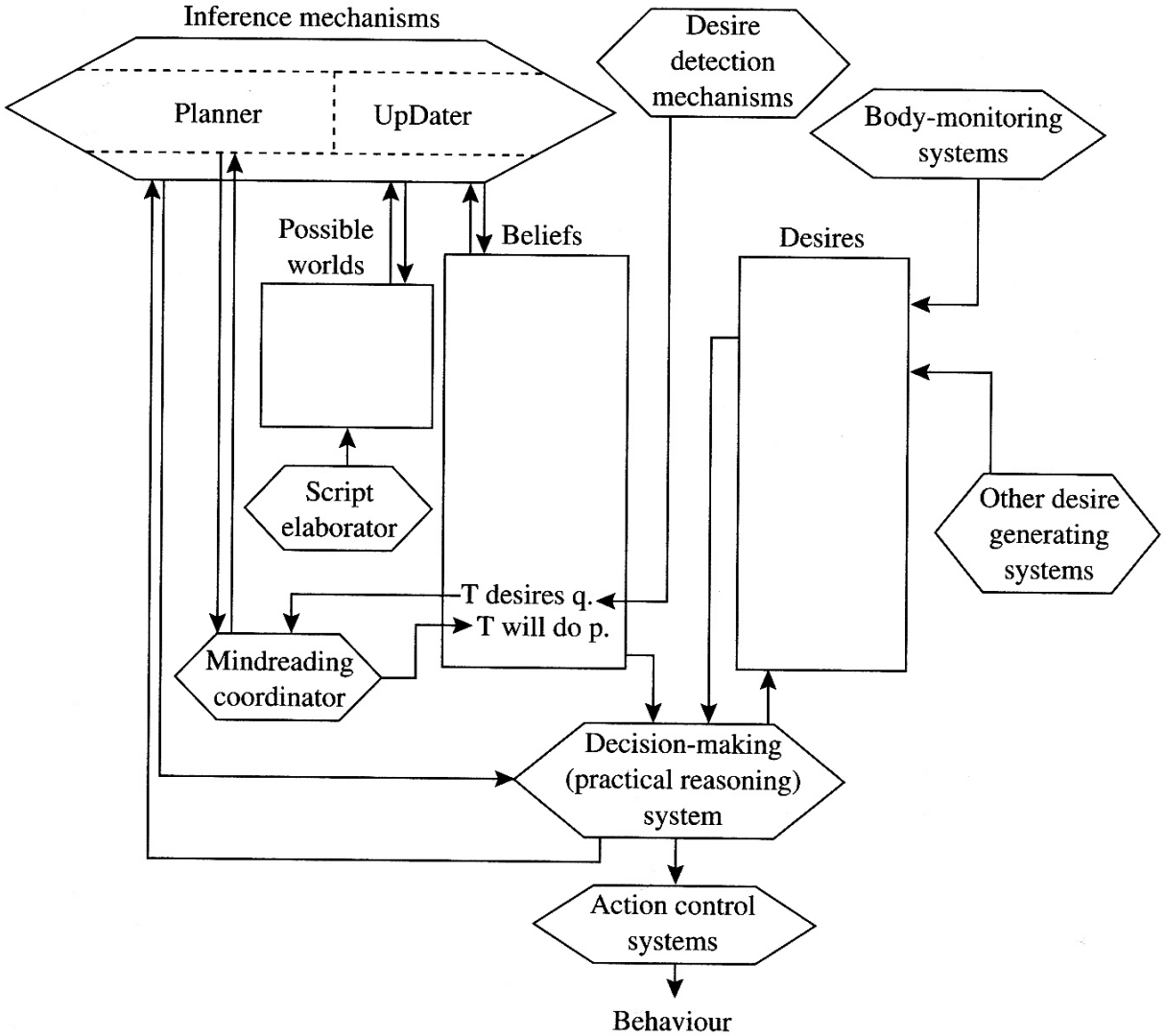


Fig. 3.4 The later mindreading system: The PWB and Desire system (Nichols & Stich, 2003)

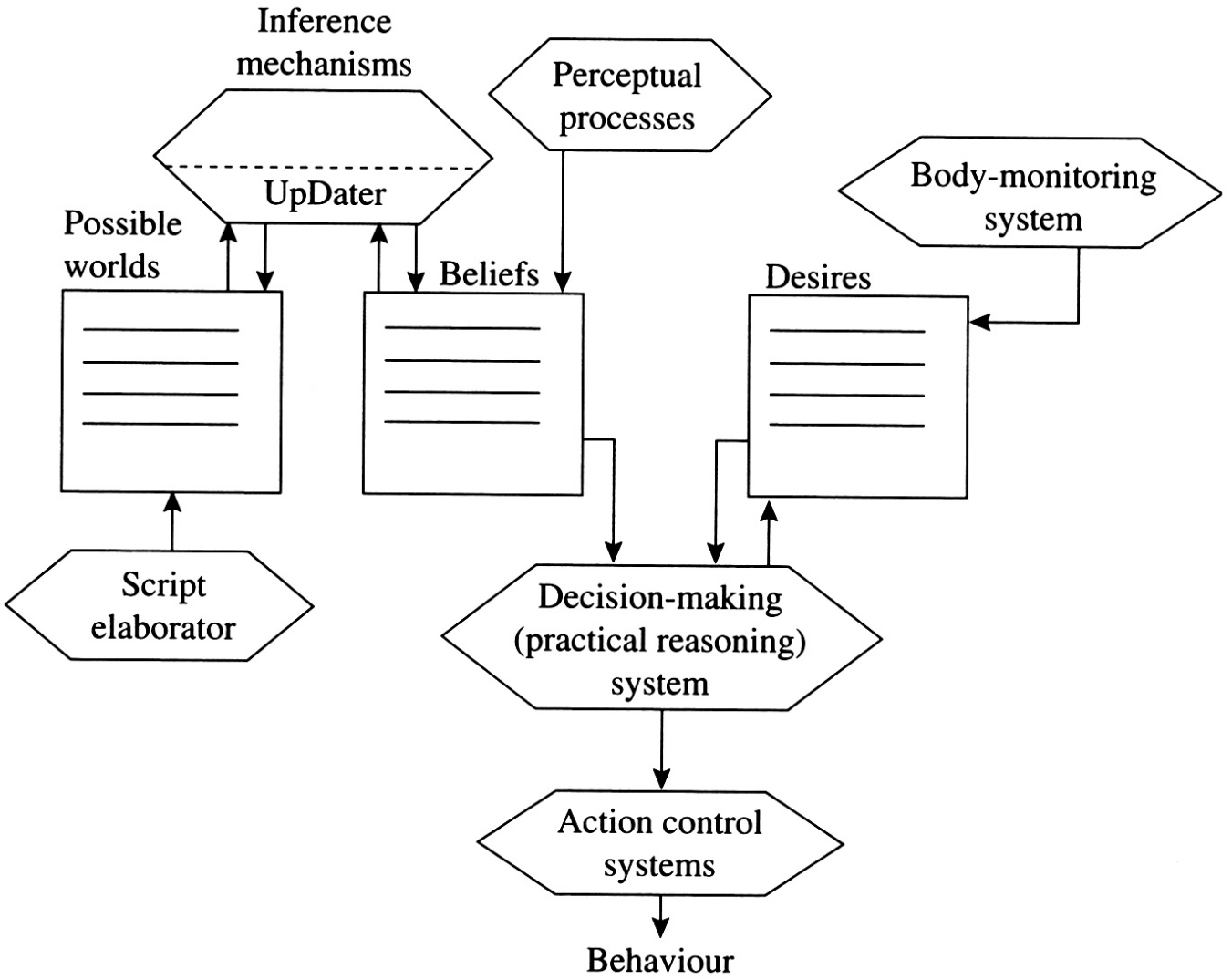
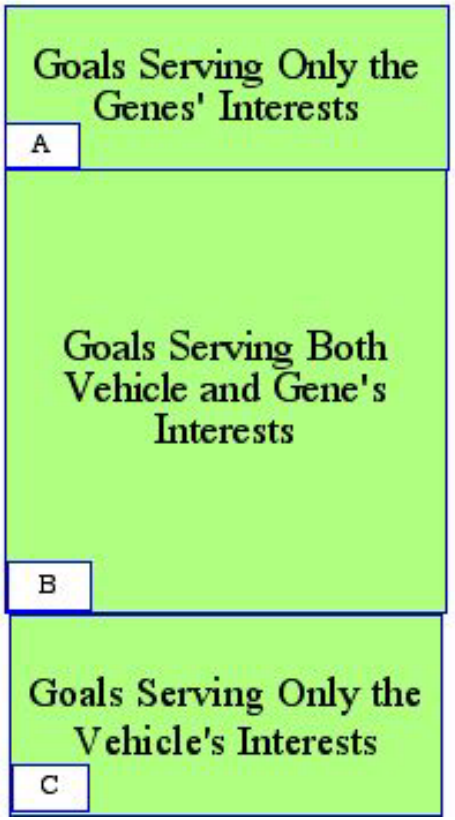
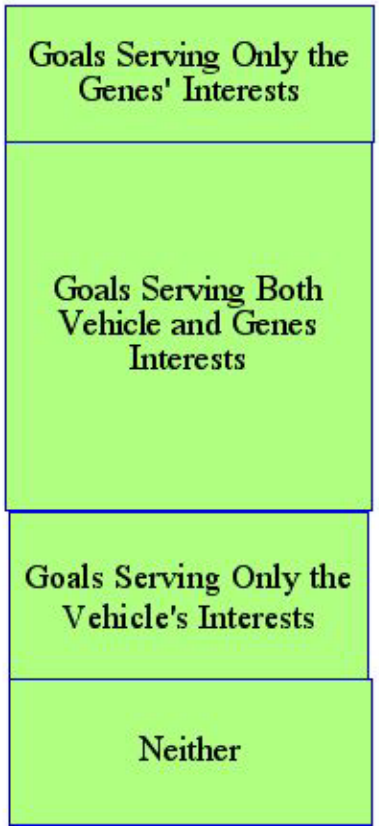


Fig. 2.2. The mental mechanisms posited by Nichols and Stich (2003)

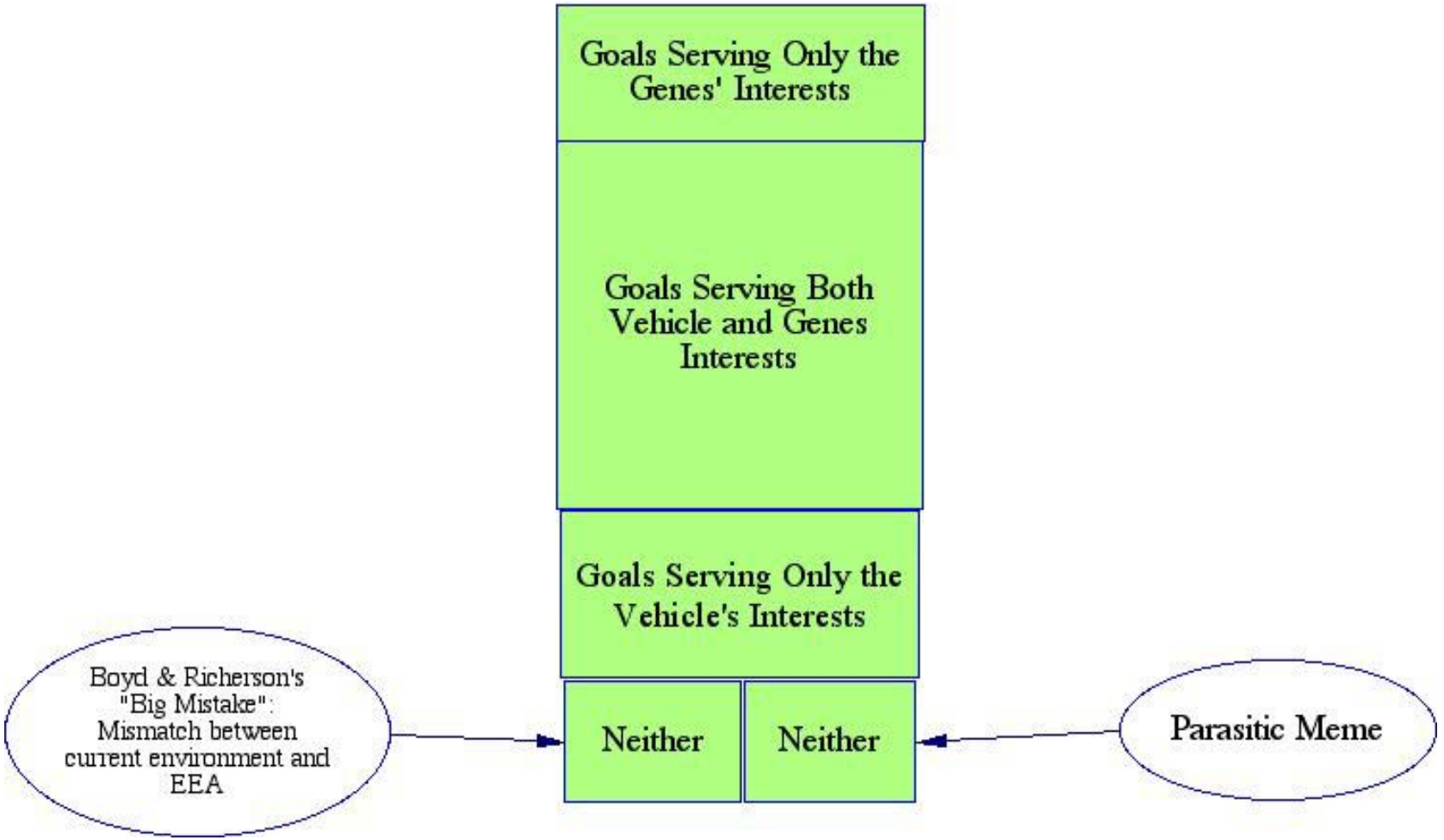
Goal Structure in Humans



Goal Structure in Humans



Goal Structure in Humans



Sampling of Effects from the Heuristics and Biases Literature

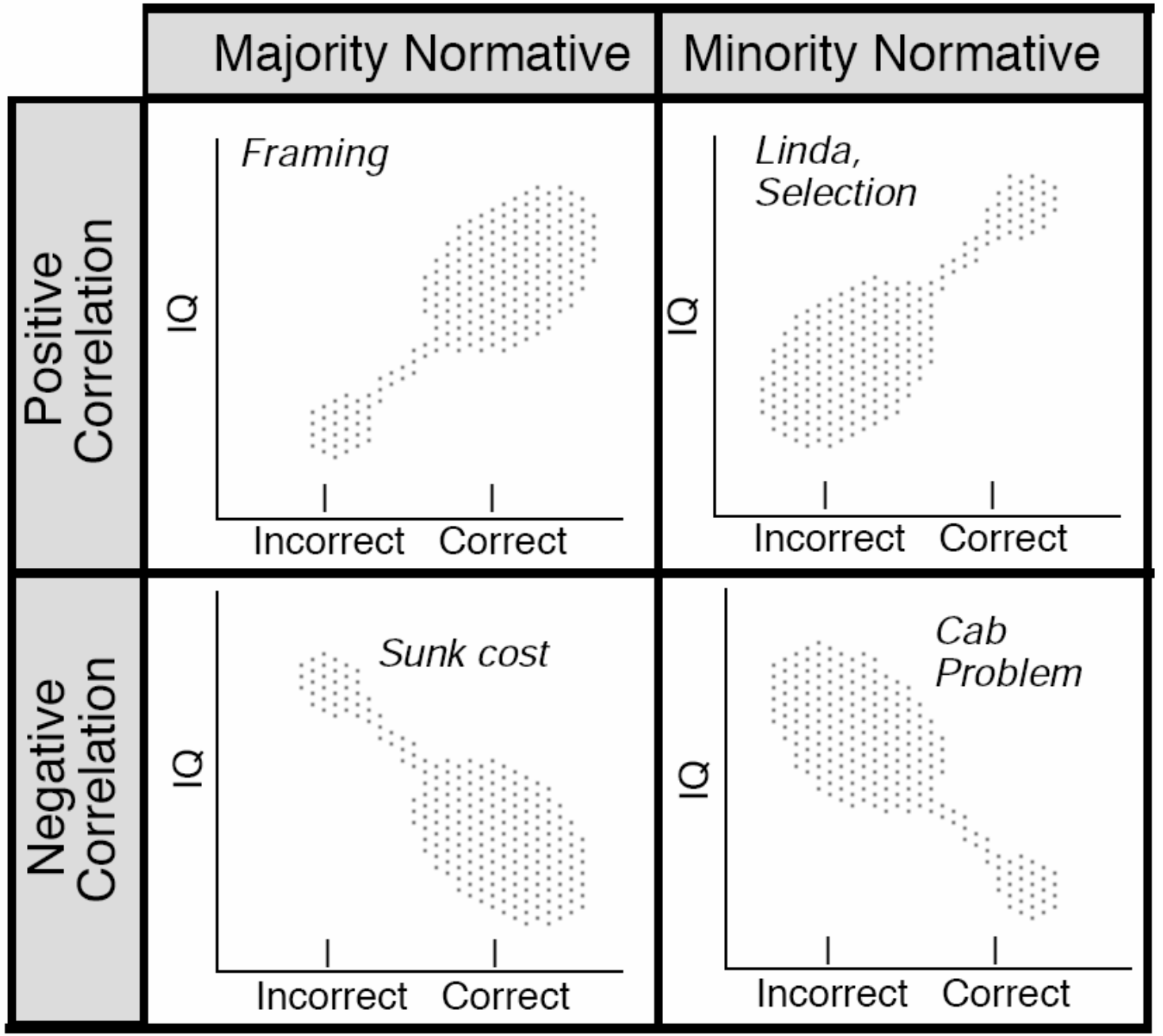
Hindsight Bias
Overconfidence Effect
Baserate Neglect
The Conjunction Fallacy
Nonregressive Predictions
Myside Bias
Inaccurate Covariation Estimation
Pseudodiagnosticity
Belief Bias
Inappropriate Anchoring
Illusory Correlation
Belief Perseverance
Preference Reversals
Outcome Bias
Commission Bias
Failure of Inconsistency Detection
Violation of SEU Axioms
Ignoring Denominator of the Likelihood Ratio
Failure to Generate Alternative Explanations
Unrealistic Optimism

People are nearly-incorrigible “cognitive optimists”. They take for granted that their spontaneous cognitive processes are highly reliable, and that the output of these processes does not need re-checking” (p. 90)

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. Cognition, 57, 31-95.

People are nearly-incorrigible “cognitive optimists”. They take for granted that their spontaneous cognitive processes are highly reliable, and that the output of these processes does not need re-checking” (p. 90)

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. Cognition, 57, 31-95.


Majority Normative
Minority Normative
Positive Correlation
Negative Correlation
Framing
Linda, Selection
Sunk cost
Cab Problem
IQ
Incorrect
Correct