Crazy Ideas June 2015

Consciousness and Rationality Explained

John Rushby

Computer Science Laboratory
SRI International
Menlo Park, California, USA

Preamble

- I talked about the evolutionary function of consciousness in 2012
- I've now improved the treatment to include rationality
- It explains some hitherto puzzling features
- And is obviously correct
- But you may think it's a crazy idea

Consciousness

- "Consciousness is a fascinating but elusive phenomenon; it is impossible to specify what it is, what it does, or why it evolved" [Johnson-Laird, Mental Models]
- Most attempts to understand or explain consciousness focus on subjective experience or qualia
  - "The hard problem of consciousness is the problem of explaining how and why we have qualia or phenomenal experiences–how sensations acquire characteristics, such as colors and tastes" [Chalmers]
  - . . . materialist theories of mind omit the essential component of consciousness, namely that there is something that it is (or feels) like to be a particular conscious thing [Nagel, What Is It Like To Be A Bat?]
- They go wrong at the first step!

Rationality

- "Man is a rational animal" [Medieval, scholastic period]
- Hierarchy of life: nutritive (plants), perceptual/instinctual (animals), rational (man) [Aristotle]
- Rationality: capacity for deliberative imagination [Aristotle]
- Modern Neuroscience finds that most of what we (humans) do is driven by instinctual, automated processes
  - System 1
  - Lots of specialized modules, fast, works well enough
  - Same as in animals
- Then there is a deliberative mechanism, looks like rationality
  - System 2
  - Slow, easily tired, can work well but has puzzling features

Puzzles of Rationality

- System 2 claims it made a decision at time t but sensors and imaging says it was made by System 1 at time $t - \delta$ [Libbet]
- Split brain studies show that System 2 makes up reasons why System 1 did something
- In general, System 2 seems more a watcher than a doer
- And a creator of post-hoc rationalizations for decisions already executed by System 1

What Really Is Special About Humans?

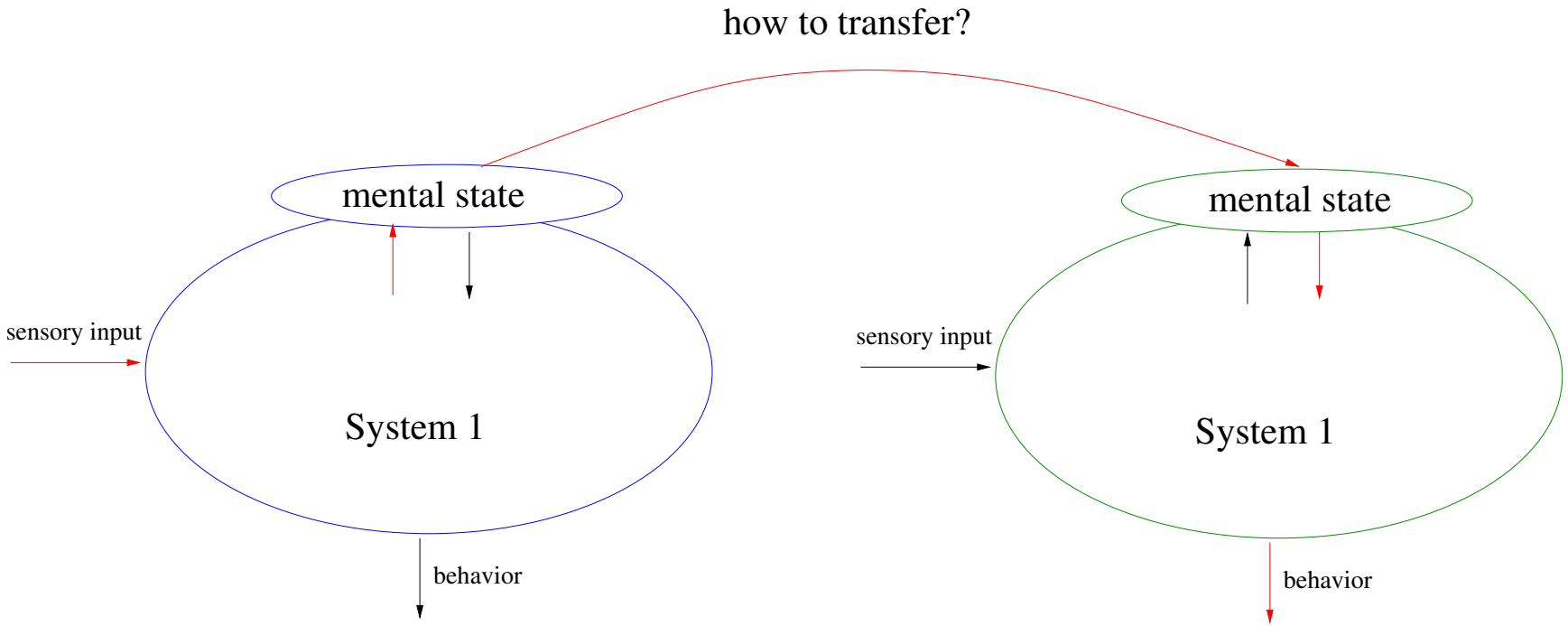
- Rationality? Seems uniquely human, but only a small part of what we do
- Consciousness? What is it like to be a bat?
- No, the uniquely human attribute is our ability to perform novel actions as a cooperative group
  - A single human is feeble thing
  - But collectively we rule the world
- Social insects and hunting pack mammals (wolves) form cooperative groups
  - But their behavior is programmed by evolution
  - Individual actions adjust parameters of existing behaviors
  - Cannot create new ones

Consciousness and Rationality as Enablers Of Novel Group Behavior

- Traditional models of consciousness and rationality focus on what they do for the individual . . . for me
- Instead, let's look at how they enable group behavior
- Imagine a pre-human ancestor facing a ravine
- System 1 suggests using a fallen tree as bridge
- But the tree is too big to move, needs help
- Another individual watches the struggles, will he help?
- No. Would your dog help?
  - Second individual no idea what is going on.
  - Neither does the first individual. . . just follows System 1 instructions without introspective insight into its actions

Here's The Problem

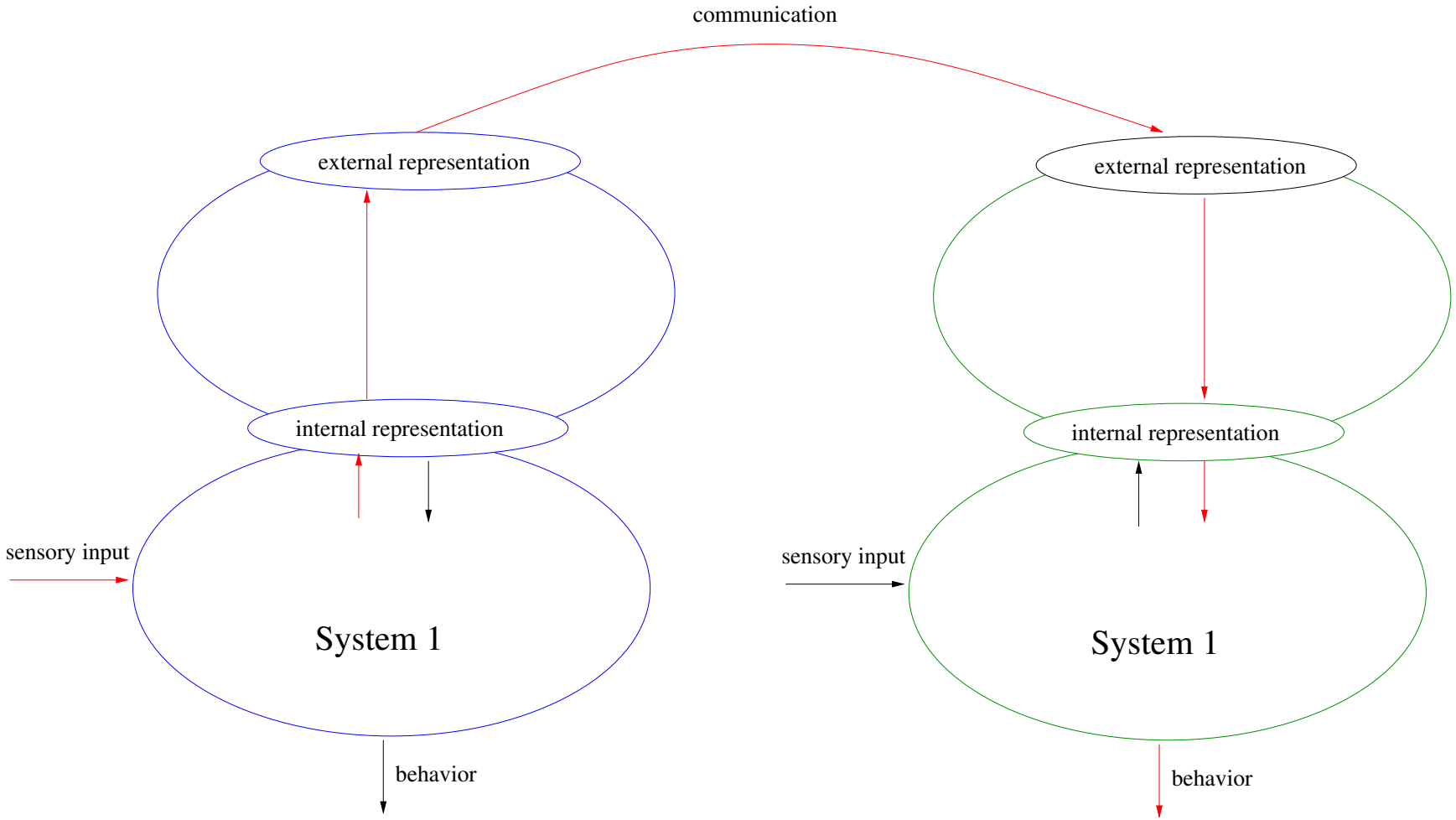
- To get cooperation, we have to transfer some of the mental state from the first individual to the second
- Can't just transfer raw neural state: may have different configurations (imagine two robots: one Java and one C++)



Here's The Solution

- Have to abstract the mental state of the first individual up to some succinct and shared representation
- Communicate that
  - Doesn't have to be language
  - Could be demonstration, mime
- The second individual then compiles upper representation down to System 1 state and lets that go to work
- With luck, its System 1 will then suggest similar/cooperative behavior since it has a similar mental state
- Abstraction/concretion will be the task of a system separate from System 1

Solution in Pictures



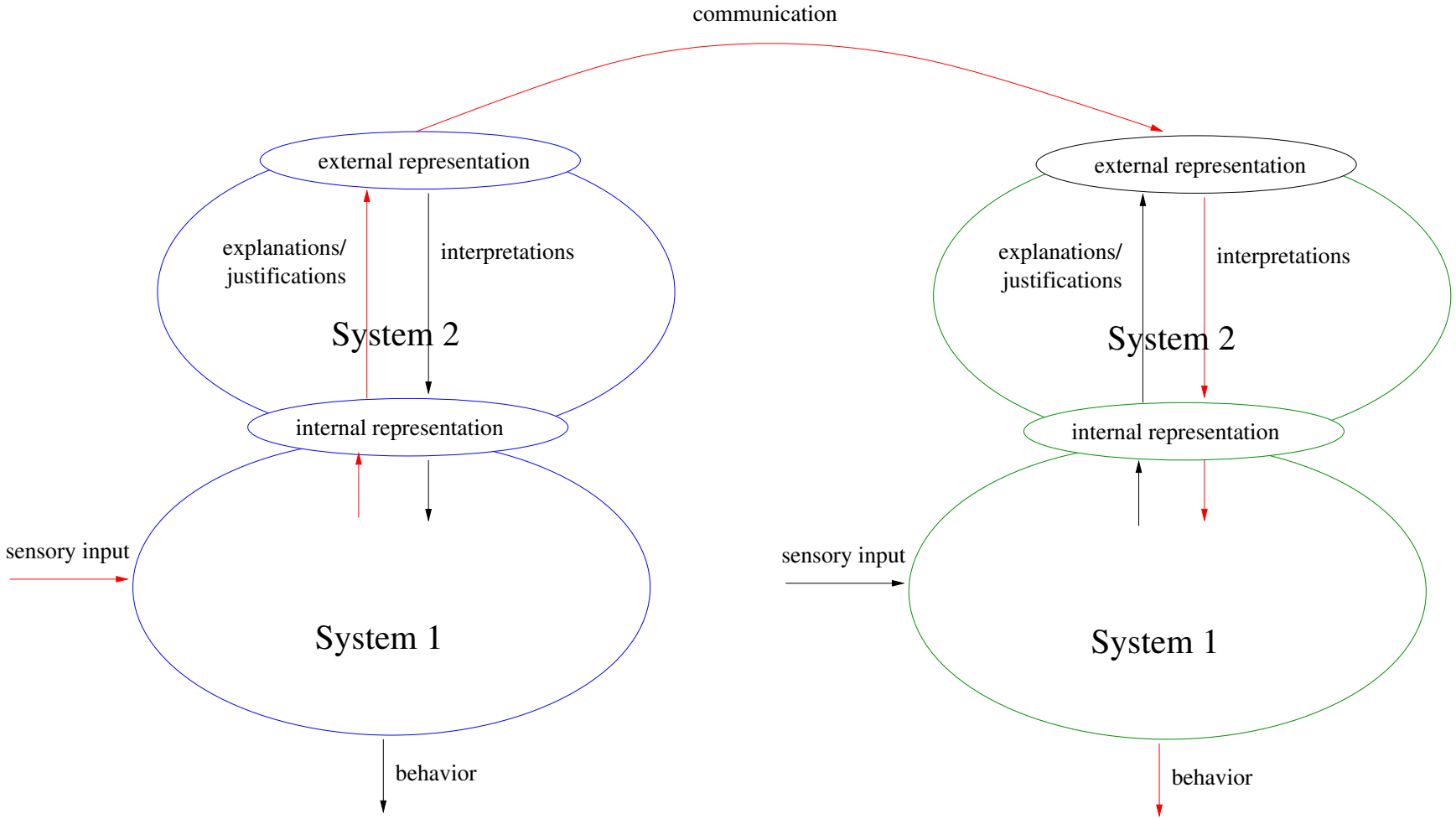
Implementation of Solution

- Second system must be able to "look" at state of the first
- The neo-cortex does that
- Will be made of similar mechanisms to System 1 (evolution)
  - Cause-and-effect reasoning
  - Elementary logical deduction
  - Mental models for some kinds of phenomena (i.e., mental simulations built on logical and cause-effect reasoning)
- That's consciousness!
- A part of the brain that looks at the brain
- Reflection in computer science terminology

More About the Implementation

- Abstraction is like concretion working in reverse
- Likely use the same mechanism in both directions
  - Unlikely to evolve a matched pair of separate mechanisms
- That's System 2
- Primarily there to explain/justify what System 1 has done
  - So it can construct a communicable abstraction
- And to interpret these back down to System 1
  - To create similar mental states in other individuals
- But could also work on its own within a single individual
  - Hey! That looks like human rationality

The Full Picture



Evaluation, Related Work

- Explains purpose of consciousness—cf. Johnson-Laird
- And why rationality has the form it does
- Based on truly unique human capacity: novel group behavior
- Reveals qualia as an epiphenomenon
- Sperber and Mercier:
  - Purpose of human reasoning is evaluation of possibly false information supplied by others

  I say we need reasoning to communicate anything at all
- Baumeister, Masicampo, and DeWall:
  - "The purpose of human conscious thought is participation in social and cultural groups"
  - Makes groups more effective

  I say it is needed to make groups work at all

Conclusion

- I don't know how to develop this to a theory that can be subject to test and refutation
- But Sperber and Mercier, and Baumeister, Masicampo, and DeWall have experimental evidence that supports my theory as much as their own
- A crazy idea?
- Or obviously true?