

# Explaining Spiritual Experiences Using a Three-Agent Model of Cognition

Frank Heile, Ph.D.

## Introduction

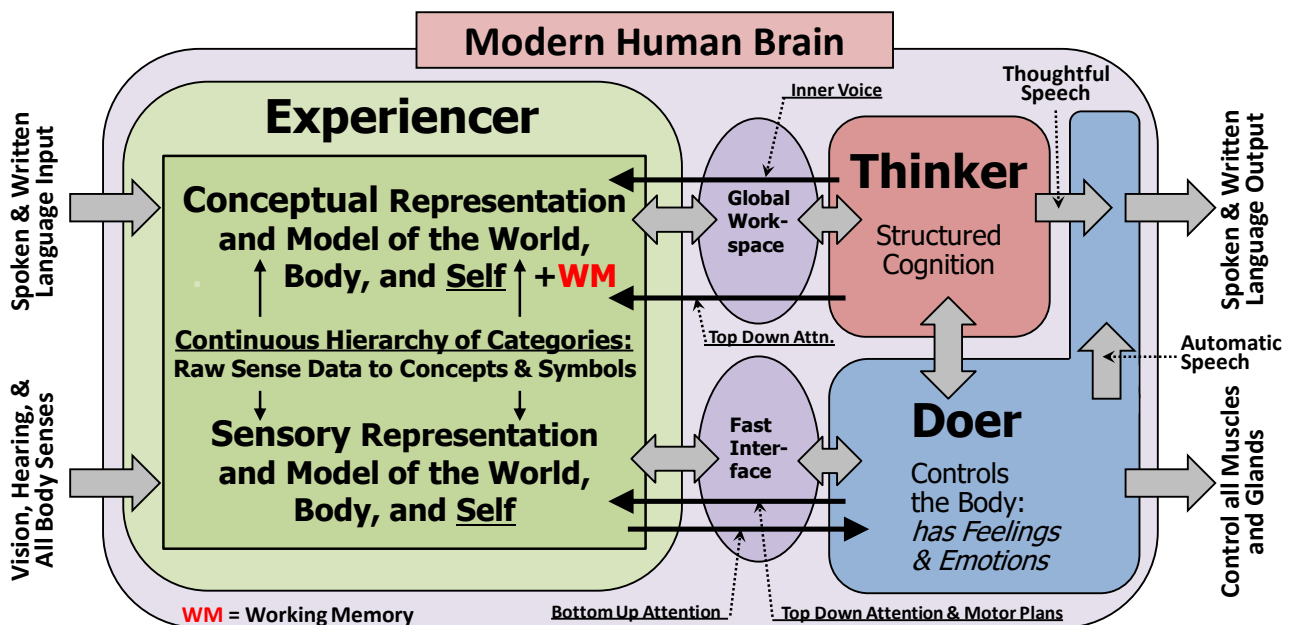
An agent, such as a human being, is an entity that can sense the world and act on the world, often in the pursuit of goals. Decomposing a complex agent into multiple sub-agents is one strategy for gaining insight into underlying mechanisms. This work proposes a decomposition of the human agent into three sub-agents in order to explain less common states of consciousness such as “spiritual experiences” and the even rarer states of “spiritual enlightenment” (also known as “unitive consciousness”).

## Proposed Three-Agent Model

The Good Regulator Theorem<sup>(1)</sup> suggests that an agent needs to contain a model of the world in order to exert control over the world. If the agent changes the world, and if the world contains the agent, then the agent’s world model must include a model of itself – a self-model. A human agent can be viewed as maintaining **two** different world models, one **sensory**, and the other **conceptual**. The **functionality** provided by the three proposed agents are:

Functionality Provided by the Agents	
<b>Thinker</b>	<b>Structured cognition</b> (executive function and working memory)
<b>Doer</b>	<b>Controls the body</b> and has emotions and feelings
<b>Experiencer</b>	<b>Constructs and supplies the sensory and conceptual, world and self-models</b> , to the Thinker and Doer.

The **interconnections** and **interfaces** between the agents are:



The proposed Thinker and Doer agents are **consistent** with experimentally derived, well established models of cognition in two different fields: in **psychology** (Dual Process Theory <sup>(2)</sup>, **DPT**) and in **neuroscience** (Yin and Knowlton <sup>(3)</sup>, *Nature Reviews Neuroscience* 7, **YK**):

Consistency with Other Models	
<b>Thinker</b>	<b>DPT:</b> System 2: slow, deliberative, explicit, & conscious <b>YK:</b> Associative Network / Action-Outcome contingency system
<b>Doer</b>	<b>DPT:</b> System 1: fast, intuitive, implicit, & subconscious <b>YK:</b> Sensorimotor Network / Stimulus-Response habit system

Given both the Thinker **and** Doer, the Experiencer is suggested by the Good Regulator Theorem<sup>(1)</sup> since these two agents need to access a single world model.

### Self-Models

The Thinker and Doer self-models are straightforward. The only problematic self-model is the Experiencer's:

Self-Models of the Agents	
<b>Thinker</b>	<b>I/Me/My</b> (autobiographical self + a simple body model)
<b>Doer</b>	<b>Body Schema</b>
<b>Experiencer</b>	<b>Attention Schema</b> (per AST <sup>(4)</sup> this is AKA <b>Awareness</b> )
<b>Human</b>	<i>Some combination of the above three sub-agent self-models</i>

In fact, it might appear that the Experiencer does **not** require a self-model since it does **not change** the external world; hence it is **not in** the external world. However, the Experiencer **directs** attention, and directing attention does **change** the **current representation** of the world. A model of the current state of top-down and bottom-up attention is defined as attention schema. Therefore, attention schema is the model of how the Experiencer changes the current representation of the world. Thus, the attention schema is the self-model of the Experiencer. In attention schema theory<sup>(4)</sup> this would also be known as awareness.

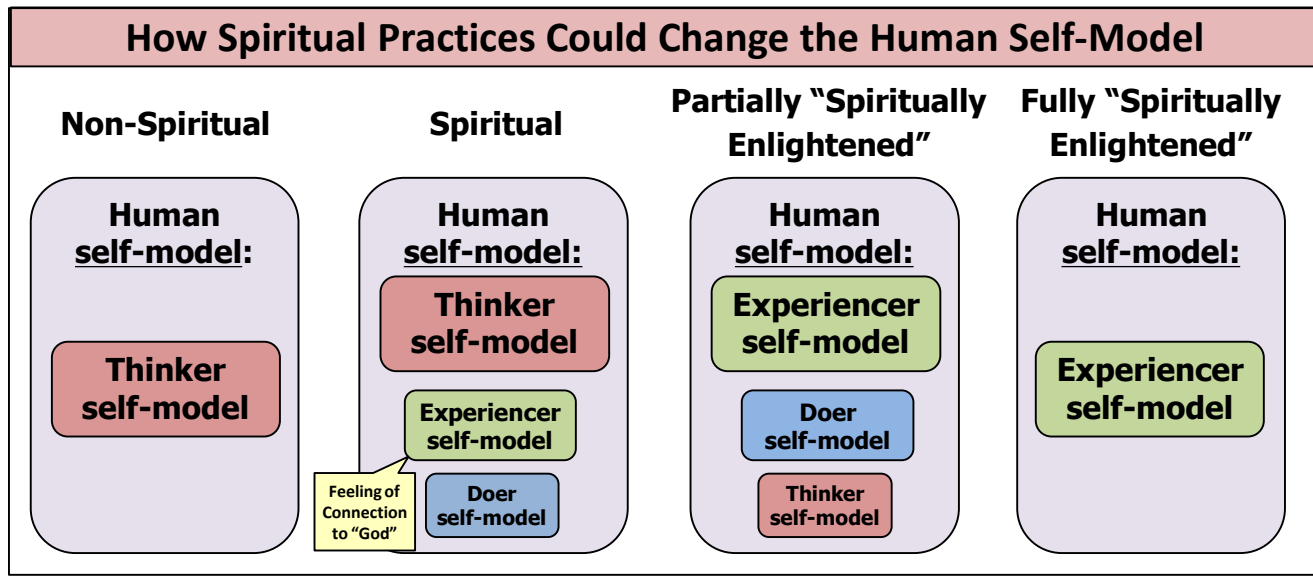
### Spirituality

Spirituality can be understood through the changes in the human agent's self-model resulting from spiritual practices. For example, a normal modern non-spiritual human would identify fully with the Thinker's self-model. Humans engaged in spiritual practices would include more of the Experiencer's self-model in the human self-model. A **fully** "spiritually enlightened" human is, therefore, **completely** identified with the Experiencer's self-model.

<sup>(1)</sup> Conant & Ashby, (1970) *Every Good Regulator of a System Must Be a Model of That System*. Int. J. Systems Sci., 1, 2, 89-97.

<sup>(2)</sup> Nobel Laureate Daniel Kahneman popularized Dual Process Theory in: (2011) *Thinking, Fast and Slow*.

<sup>(3)</sup> Yin & Knowlton, (2006) *The role of the basal ganglia in habit formation*. Nat. rev. Neuroscience. 7. 464-76. 10.1038/nrn1919.



## "Spiritual Enlightenment"

**Attention Schema Theory (AST)**,<sup>(4)</sup> proposes that conscious awareness is only present in a brain that has an attention schema. Since the Experiencer is the only sub-agent that has an attention schema, AST would imply that the Experiencer would be the only conscious sub-agent.

Dual Process Theory (DPT),<sup>(2)</sup> agrees that the Doer is subconscious; however, DPT claims that the Thinker is conscious. I claim that the Thinker only appears to be consciousness due to the Experiencer's **experience** of the Thinker's Inner Voice, and the Experiencer's **noticing** the Thinker's manipulation of the Working Memory contents. Thus, the Experiencer can be aware of the Thinker's activity, even though the Thinker, by itself, is not conscious. This fact, that the Experiencer is the **only** conscious sub-agent, may be the basis for spirituality; by using meditation, the human may achieve the insight and realization that they, the conscious entity, are **"only" the Experiencer**.

The **Hindu Advaita Vedanta** tradition claims that "enlightenment" occurs when the human realizes that the **subject-object distinction** is an "illusion" (also known as "nonduality"). The AST model of awareness is "SM+AS+CRW;" where SM is the Self-Model, AS is the Attention Schema, and CRW is the Current Representation of the World. If the human self-model is just the Experiencer self-model (which is AS), then the model of awareness becomes "AS+AS+CRW," which reduces to "\_\_\_+AS+CRW." This is a state of **self-less awareness**; thus, it is reasonable to say that the subject-object distinction would seem to disappear in this state of consciousness. In other words, the "Fully Spiritually Enlightened" human could say that **"the world and I are one."** This statement also explains "unitive consciousness," a synonym for "spiritual enlightenment."

<sup>(4)</sup> Graziano & Webb, (2015) *The attention schema theory: a mechanistic account of subjective awareness*. Front. Psych., 6, 500