Explaining Spiritual Experiences Using a Three-Agent Model of Cognition

Frank Heile, Ph.D. (Stanford Physics)

Stanford | Neurosciences Institute Symposium: Poster Session, Oct. 19, 2017

Introduction:

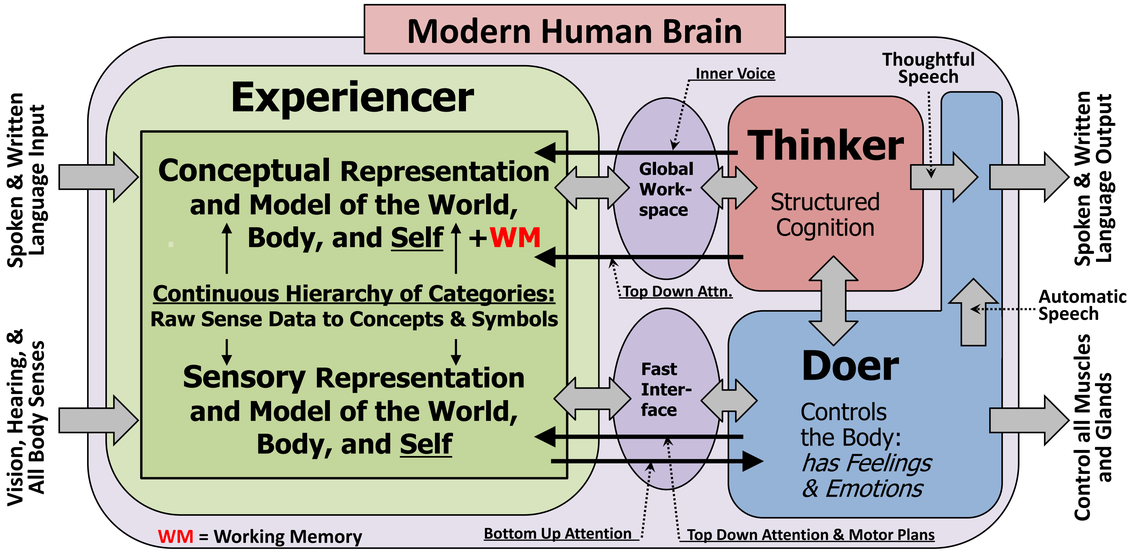
An agent, such as a human being, is an entity that can sense the world and act on the world, often in the pursuit of goals. Decomposing a complex agent into multiple sub-agents is one strategy for gaining insight into underlying mechanisms. This work proposes a decomposition of the human agent into three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

Proposed Three-Agent Model:

The Good Regulator Theorem[1] 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, 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	
Thinker	**Structured cognition** (executive function and working memory)
Doer	**Controls the body** and has emotions and feelings
Experiencer	**Constructs** and **supplies** the **sensory** and **conceptual**, **world** and **self-models**, to the Thinker and Doer.

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



The proposed Thinker and Doer agents are **consistent** with experimentally derived models of cognition in both **psychology** (Dual Process Theory [2], **DPT**) and **neuroscience** (Yin and Knowlton [3], *Nature Reviews Neuroscience 7*, **YK**):

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

Given both the Thinker **and** Doer, the Experiencer is suggested by the Good Regulator Theorem[1] since these two agents need to access a single world model.

Self-Models:

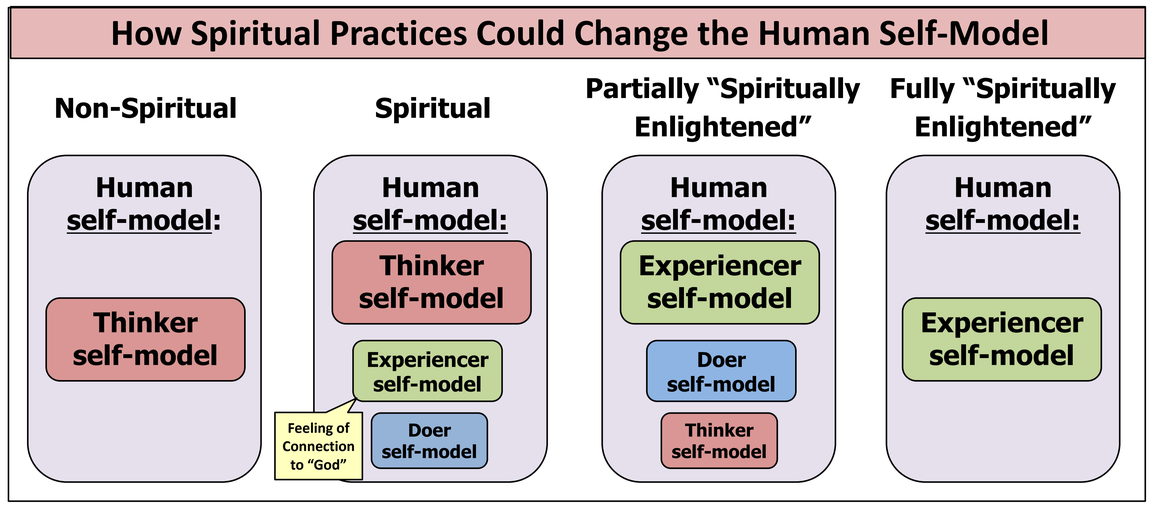
The Thinker and Doer self-models are straightforward:

Self-Models of the Agents	
Thinker	**I/Me/My** (autobiographical self) + **Simple Body Model** + goals
Doer	**Body Schema** + goals
Experiencer	**Attention Schema** + **Current Representation of the World**
Human	*Some combination of the above three sub-agent self-models*

The only problematic self-model is the Experiencer's. In fact, it might seem 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 internal representation** of the world. Therefore, the current state of top-down and bottom-up attention, which is defined as the attention schema, would need to be included in the model of the world. In other words, the **world model** would **be the attention schema plus the current representation of the world**. The way that the Experiencer is **present in** and makes **changes to** the **internal** world is through the **same** attention schema plus the current representation of the world; thus, **the Experiencer self-model is equivalent to the current world model.**

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:



"Spiritual Enlightenment:"

According to **Attention Schema Theory,**[4] the **only** conscious sub-agent is the **Experiencer**. The Thinker **appears** to be conscious (per **DPT**) but is **not truly** conscious. This appearance of consciousness is due to the Experiencer's **experience** of the Thinker's Inner Voice, and the Experiencer's **noticing** the Thinker's manipulation of Working Memory contents. That the Experiencer is the **only** conscious sub-agent, may be the basis for "spiritual enlightenment;" by using meditation, the human may achieve the realization that they are **"only" the Experiencer.**

The **Hindu Advaita Vedanta** tradition claims that "enlightenment" occurs when the human realizes that the **self-other distinction** is an "**illusion.**" If the human self-model is the Experiencer self-model, and if the Experiencer self-model is identical to the world model, then it is reasonable that the self-other distinction could be considered as an "illusion;" thus the "enlightened" human says that "**the world and I are one**." This statement explains "unitive consciousness," a synonym for "spiritual enlightenment."

---

[1] Conant & Ashby, (1970) *Every Good Regulator of a System Must Be a Model of That System*. Int. J. Systems Sci., 1, 2, 89-97.
[2] Nobel Laureate Daniel Kahneman popularized Dual Process Theory in his 2011 book, "*Thinking, Fast and Slow*."
[3] Yin & Knowlton, (2006) The role of the basal ganglia in habit formation. *Nat. rev. Neuroscience*. 7. 464-76. 10.1038/nrn1919.
[4] Graziano & Webb, (2015) *The attention schema theory: a mechanistic account of subjective awareness.* Front. Psych., *6*, 500