

Mirror Neurons, Mirrorhouses, and the Algebraic Structure of the Self

Ben Goertzel¹, Onar Aam, F. Tony Smith, Kent Palmer
¹Novamente LLC

November 18, 2007

Abstract. Recent psychological research suggests that the individual human mind may be effectively modeled as involving a group of interacting social actors: both various subelves representing coherent aspects of personality; and virtual actors embodying “internalizations of others.” Recent neuroscience research suggests the further hypothesis that these internal actors may in many cases be neurologically associated with collections of mirror neurons. Taking up this theme, we study the mathematical and conceptual structure of sets of inter-observing actors, noting that this structure is mathematically isomorphic to the structure of physical entities called “mirrorhouses.” Mirrorhouses are naturally modeled in terms of abstract algebras such as quaternions and octonions (which also play a central role in physics), which leads to the conclusion that the presence within a single human mind of multiple inter-observing actors naturally gives rise to a mirrorhouse-type cognitive structure and hence to a quaternionic and octonionic algebraic structure as a significant aspect of human intelligence. Similar conclusions would apply to nonhuman intelligences such as AI’s, we suggest, so long as these intelligences included empathic social modeling (and/or other cognitive dynamics leading to the creation of simultaneously active subelves or other internal autonomous actors) as a significant component.

Introduction

The thesis of this paper is that there are certain abstract algebraic structures that typify the self-structure of human beings and any other intelligent systems relying on empathy for social intelligence. These algebraic structures, called quaternions and octonions, are familiar to mathematicians, and also play a critical role in modern theoretical physics (Dixon, 1994).

The argument presented in favor of this thesis has two steps. First, it is argued that much of human psychodynamics consists of “internal dialogue” between separate internal actors – some of which may be conceived as subelves à la (Rowan, 1990), some of which may be “virtual others” intended to explicitly mirror other humans (or potentially other entities like animals or software programs). Second, it is argued that the structure of inter-observation among multiple inter-observing actors naturally leads to quaternionic and octonionic algebras. Specifically, the structure of inter-observation among three inter-observers is quaternionic; and the structure of inter-observation among four inter-observers is octonionic. This mapping between inter-observation and abstract algebra is made particularly vivid by the realization that the quaternions model the physical