Computation and Intentionality: A Recipe for Epistemic Impasse

I. SHANI
The University of Haifa, Mount Carmel 31905, Haifa, Israel; E-mail: ishani479@hotmail.com

Abstract. Searle’s celebrated Chinese room thought experiment was devised as an attempted refutation of the view that appropriately programmed digital computers literally are the possessors of genuine mental states. A standard reply to Searle, known as the “robot reply” (which, I argue, reflects the dominant approach to the problem of content in contemporary philosophy of mind), consists of the claim that the problem he raises can be solved by supplementing the computational device with some “appropriate” environmental hookups. I argue that not only does Searle himself cast doubt on the adequacy of this idea by applying to it a slightly revised version of his original argument, but that the weakness of this encoding-based approach to the problem of intentionality can also be exposed from a somewhat different angle. Capitalizing on the work of several authors and, in particular, on that of psychologist Mark Bickhard, I argue that the existence of symbol-world correspondence is not a property that the cognitive system itself can appreciate, from its own perspective, by interacting with the symbol and therefore, not a property that can constitute intrinsic content. The foundational crisis to which Searle alluded is, I conclude, very much alive.

Key words: Bickhard, computational theory of the mind, encoding, intrinsic intentionality, Searle, the Chinese room, the robot reply

1. Introduction

John Searle’s Chinese room thought experiment (1980) is, without a doubt, one of the most celebrated, and most pointed and divisive, criticisms of the computational (symbol-manipulation, information processing) theory of the mind. From the moment of its inception, Searle’s Gedankenexperiment provoked strong reactions. Many have dismissed the argument, giving it short “refutations” (see for example, Abelson, 1980; Block, 1980), yet its import continues to reverberate in the milieu of cognitive science, artificial intelligence, and the philosophy of mind.

As Searle himself made quite clear, the aim of the Chinese room argument was to show that digital computers do not, and cannot, exhibit intrinsic intentionality and hence that the intentionality of intelligent beings, humans included, does not, and cannot, consist of computer-like symbol manipulation.

In what follows, I examine Searle’s argument in some detail. I proceed to discuss one of the major rebuttals to Searle’s charge – the robot reply (RR).
I argue that RR reflects the most popular approach to the naturalization of mental content, an approach inspired by the computational theory of the mind (CTM). According to RR, what constitute mental content are symbol-world correspondence relations – in complete independence from, yet in supplementation to, the system’s cognitive architecture. I argue that not only does Searle himself cast doubt on the adequacy of this idea by applying to it a slightly revised version of his original argument, but that the weakness of this encoding-based approach to the problem of content can also be exposed from a somewhat different angle.

Capitalizing on the work of several authors, and, in particular, on that of psychologist Mark Bickhard, I argue that the existence of symbol-world correspondence is not a property that the cognitive system itself can appreciate, from its own perspective, by interacting with the symbol and therefore not a property that can constitute intrinsic content. Thus, attempts to deal with the problem of content within the dominant information-processing approach fail to live up to the expectation (shared by the majority of its own practitioners) of articulating a notion of mental content that can make epistemic sense from the first-person perspective of a psychological agent. The result, I submit, is a failure to naturalize content.

2. Part One: Intrinsic Intentionality and the Chinese Room

2.1. SEARLE’S AARGUMENT AS A QUÆ ARGUMENT

What is the Chinese room argument an argument for: The Chinese room thought experiment is devised as an attempted refutation of the view that appropriately programmed digital computers literally are the possessors of genuine mental states. As such, the argument is a direct assault on the CTM, often described as “cognitivism” (Harnad, 1990; Searle, 1992) or “the cognitive science view” (Fodor, 1975; Sayre, 1987). According to CTM, mental processes are computational processes performed on internal symbol strings; intelligent beings are, therefore, digital computers of the appropriate sort, and the challenge, from the point of view of theoretical cognitive research, is to identify, understand and (eventually) be able to construct such intelligent formal automata. The Chinese room argument aims to show that computer-like information processing is not only insufficient for the having of genuine mental states but that, in all likelihood, it plays no constitutive role in the making of intentional agency (mental life).

If appropriately programmed computers really are minds, Searle reminds us, then they “can be literally said to understand and have other cognitive states” (1980, 282). Correspondingly, Searle constructs his argument as a counterexample to the claim that (a) appropriately programmed digital