

Algorithmicity and Consciousness

Why should one believe that conscious awareness is solely the result of organizational complexity? What is the connection between consciousness and combinatorics: transformation of quantity into quality? The claim that the former is reducible to the other seems unconvincing—as unlike as chalk and cheese! In his book¹ Penrose is at least attempting to compare like with like: the enigma of consciousness with the progress of physics.

And even if one were to become convinced on “scientific” grounds that there is no option but to regard consciousness as an ultimately combinatorial matter, how would one go on to formulate a cogent account of the evident *subjectivity* of consciousness: the fact that a conscious being has access only to its own—and no other—consciousness?

I doubt whether it is possible to construct a theory of consciousness relating, but not necessarily reducing the phenomenon to other, presumably less problematic phenomena, and yet at the same time doing full justice to its uniquely subjective character. In the words of Hermann Weyl, such a theory

may be objectively adequate, but it is shattered by the desperate cry of Judas: Why did I have to be Judas! The impossibility of an objective formulation to this question strikes home, and no answer in the form of an objective insight can be given. Knowledge cannot bring the light that is I into coincidence with the murky, erring human being that is cast out into an individual fate.

Now both Penrose and the AI people appear to share the belief that such a theory *is* possible—and presumably necessary! Indeed the latter, at least in Penrose’s portrayal, claim already to be in possession of one. Penrose himself seems to believe that a satisfactory “objective” theory of consciousness may emerge from future developments in physics. In any event, the fact that both parties believe such a theory to be formulable in principle already distinguishes them from most philosophers who take consciousness seriously, e.g. Descartes, Husserl who regard consciousness as an irreducible “given”.

It seems to me that an “objective” theory of consciousness would, as a consequence of its “external” or “extensional” character”, fail to do justice to the subjective or “intensional” phenomenon of actually *being conscious*, just as mathematical theories of time necessarily fail to

¹ *The Emperor’s New Mind.*

capture the phenomenon of *being in time*. If it is pointed out that scientific theories are not intended to capture the essence of the subjective, then I can only reply that this amounts to admitting—as I believe one finally must—that scientific knowledge is limited.

Whether thought and objective reality are, in the last analysis, algorithmic in character is a question of some subtlety. The Gödel incompleteness theorem establishes the “nonalgorithmic” character of the truth of a particular arithmetical statement **A** *at the level of a specified formalism*. On the other hand, the truth of **A** *can* be established within a *strengthened* formalism, thereby reinstating “algorithmicity” of the truth of **A**. But now, inevitably, new “non-algorithmically verifiable” statements make their appearance at the level of the strengthened formalism, and the whole affair starts all over again. The point would seem to be that, while “algorithmicity” is manifested locally at higher and higher levels of description and is likely to be an essential constituent of such descriptions, nevertheless diagonal arguments show that algorithmicity *cannot* be imposed *globally*, that is, uniformly and simultaneously, on *all* levels of description.

The apparent ubiquity of algorithmicity, or, more generally, symbolizability, should not mislead us into believing that the world is itself an algorithm or a symbol.