TOWARDS A THEORY OF REFLEXIVE INTENTIONAL SYSTEMS

ABSTRACT. (1) Intentional system: a system whose behaviour we may reliably predict via the intentional strategy, i.e., by interpreting its behaviour as a (more or less) rational consequence of its beliefs and desires. (2) Reflexive intentional system: a system that is able to interpret itself via the intentional strategy, and whose behaviour is, thus, influenced by an understanding of itself. All intentional systems behave in a meaningful way, but only reflexive intentional systems are aware of the meaning. Hence, only the latter are conscious of what they are doing.

1. INTRODUCTION

Classical cognitive science started out from the conviction that the concept of a Turing machine (TM), which was regarded as being well understood, and compatible with natural science, would eventually make feasible a truly “scientific” understanding of mental phenomena. Today, that conviction is increasingly being questioned. Conscious mental phenomena have a peculiar “subjective” quality, which would seem to elude any purely computational analysis.

In this paper, I argue that the concept of a reflexive intentional system could provide a first step towards a full analysis of subjectivity. According to Dennett’s (1981) analysis, both biological and artificial systems may be interpreted at what he calls the intentional level. Such interpretations yield an understanding of the meaning – or the point – of what the system is doing, and thus, by definition, a certain capacity to predict its future behaviour. For Dennett, the “meaning” is strictly in the eyes of the beholder; it is, as he puts it, an ascriptive concept. The ability to predict (within certain limits) is, however, something objective, and testable. Dennett’s “intentional level” is not reducible to either the physical or the computational level, but neither is it incompatible with them; and we can have a full understanding of how the different levels arise (one aim of my paper is to vindicate the epistemological soundness of such “levels” of description and explanation).

While, traditionally, intentionality has been seen as coextensive with consciousness, Dennett’s theory obviously implies a radically different
concept. It may be applied to systems that clearly have no consciousness at all. Against that background we may then define “reflexive intentional systems” as systems that are not only subject to intentional interpretation by an observer, but also able to interpret themselves. This capacity for self-interpretation I take to be one of the constituents of “consciousness”. Thus, whereas traditional philosophy sees consciousness as the source of all meaning, this perspective implies the reverse: consciousness itself arises in the interpretation of a meaning that is already there.

2. THE PHYSICAL WORLD

The division of reality into the three basic domains of physics, biology and psychology has roots that go back at least to Aristotle. To begin with, that division did not imply any obvious epistemological ranking order. With Galileo and Newton, however, a development began that was soon to lead to the present situation, where physics is clearly the exemplary science, biology is tolerated for its apparent nearness to, and compatibility with, physics, but the state of psychology is widely considered a source of embarrassment.

I do not wish to question the present evaluation, but I believe that any attempt, today, to deal with the problem of “mental” phenomena must start out from a fairly deep understanding of the nature and basis of this eminence of the physical sciences.

Galileo’s decisive accomplishment was to make dynamic phenomena amenable to measurement. He thus inaugurated a development which was soon to lead, with Newton, to a replacement of a crude mechanistic notion of causality by that backbone of classical physics, the concept of a “law of nature” in the form of a differential equation. Thus, prediction and control on an altogether different scale became conceivable, and eventually the Laplacean idea of concrete reality as describable and predictable with limitless precision was born. This should be contrasted with antiquity’s notion of panta rei: concrete reality as essentially opaque and subject to unpredictable “flow”, with only the world of platonic ideas really subject to mastery by science.

As Sneed (1971) has analysed in his The Logical Structure of Mathematical Physics, the relation between measurement and physical law is an intimate one. Not only is the meaning of the terms that make up a law of physics tied to how we measure them – Sneed describes measurement as a function from reality to the (real) numbers – the method of measurement already presupposes the validity of a corresponding law. Thus mass, for instance, which on the one hand is used to explain the tendency of all