ABSTRACT. It is often thought that the computational paradigm provides a supporting case for the theoretical autonomy of the science of mind. However, I argue that computation is in fact incompatible with this alleged aspect of intentional explanation, and hence the foundational assumptions of orthodox cognitive science are mutually unstable. The most plausible way to relieve these foundational tensions is to relinquish the idea that the psychological level enjoys some special form of theoretical sovereignty. So, in contrast to well known antireductionist views based on multiple realizability, I argue that the primary goal of a computational approach to the mind should be to facilitate a translation of the psychological to the neurophysiological.

1. INTRODUCTION

There is a very widely held view that the psychological constitutes an autonomous theoretical domain that cannot be ‘reduced’ to the level of more basic physical entities and interactions. There are a number of different versions of this view, with differing motivations and background assumptions. Perhaps the most intuitive version rests upon the idea that the psychological essentially involves the intentional, and that the intentional level of explanation is itself conceptually closed and self-sufficient. If actions are caused by mental states identified in terms of their content, then appeal to these content laden states is the salient mode of explanation.

Such an idea seems to underlie our everyday practice of accounting for human behavior in terms of beliefs, desires, fears, and other standard intentional states. In day to day life, the belief.desire model of explanation is deemed adequate to fully account for the various pieces of behavior to which it is applied. For example, if Mary is said to have walked to the bar because she wanted a shot of whisky and she believed that she could obtain one there, then that’s held to be a perfectly full and complete explanation of the action. The basic psychological account is not in need of, nor is it improved by, the addition of further details concerning the neurophysiological substrate of Mary’s beliefs, the biomechanics of limb movement, the psychophysical correlations between the retinal reception...
of electromagnetic radiation and Mary’s visual experiences that enable her to find the bar, etc.\textsuperscript{2}

Classical cognitive science embraces the framework of mental explanation in terms of intentional states and content. And to this traditional framework it adds the assumption that computational analysis will reveal the essential structure underlying these intentional phenomena. Thus the theoretical task of cognitive science is to ascertain the formal procedures that underpin content bearing states and complex cognitive behavior, and some version of this computational paradigm is normally used to identify cognitive science \textit{per se} as a distinct province of research.

The computational paradigm is standardly welded to some brand of functionalism, according to which mental states derive their identity from the abstract causal roles they play in a complex economy of internal states mediating environmental inputs and behavioral outputs. But whatever the details and fine-grained variations, functionalism and/or classic computationalism share the notion that decomposition in terms of input/output profiles and sequences of internal processing states is the salient model for understanding cognitive phenomena. Thus according to the classical view, the level of relevance to the science of mind is the \textit{program}, the sequence of abstract state transitions that yields intelligent responses to external stimuli. This mind/program analogy is one of the basic conceptual legacies of orthodox cognitive science and AI.

Because of this emphasis on abstract internal state transitions, functionalism and computationalism share yet another key feature, which is multiple realizability. The same functional organization can be realized in any number of different physical systems. Thus humans, Martians and robots could all hold the same belief that \( P \), if all three types of system realized the requisite functional conditions. Similarly, the very same program can be run on any number of different hardware systems, from silicon chips to gears and levers to human neurons.

As a direct corollary of the computational/functional paradigm, it is often said that even if one is a hardened materialist, it is still possible to study the mind without really bothering with the brain, since the \textit{mental} level is characterized by algorithms and assorted formal architectures, and these can be instantiated in a limitless variety of physical media. What is then of interest to the science of mind is the abstract formal procedure, not the bowl of porridge in which it happens to be implemented.\textsuperscript{3}

Thus there appears to be a deep compatibility between intentional explanation and the computational paradigm; both of these key aspects of cognitive science seem to imply that the psychological constitutes an autonomous level of description and explanation, and that the organic de-