Functionalism and Strange Minds

Miles Christie

Degree Awarded by: Oxford Brookes University

Thesis submitted in partial fulfilment of the requirements of the award of an MA

Submitted: September 2022

Table of Contents

Preface **3-5**
Chapter 1: Introducing Functionalism **6-30**
An introduction to functionalist philosophy of mind (1.1) 6-10
On Putnam (1.2) 10-12
Strains of functionalist thought (1.3) 12-14
Long and short arm functionalism (1.4) 14-16
The Madman and the Martian (1.5) 17-20
Chalmers fine-grained and coarse-grained distinction (1.6) 20-23
Computational analogies of the mind (1.7) 24-25
Functionalism and holism (1.8) 26-27
Qualia Problems (1.9) 27-30
Chapter 2: Strange Minds: 31-63
Putnam's Turn (2.1) 31-35
Viability of other minds (2.2) 35-39
Extended cognition (2.3) 39-47
Materials Matter When it Comes to Minds (2.4) 47-58
Strange minds 1: Solaris (2.5) 58-60
Strange minds 2: Junkyard (2.6) 60-63
Chapter 3: Interpretation and rationality 64-87
Interpreting minds (3.1) 64-72
Minimal minds (3.2) 73-77
Time and the mind (3.3) 78-80
What is it like to be Solaris? (3.4) 80-84
Practicality issues with functionalism and interpretation (3.5) 84-87
Bibliography 88-92

Preface:

Functionalists argue that mental states are merely functional states of the brain. Instead of the physical makeup of a particular mind mattering, it is the causal relations between input states and output states and dispositions to act that together constitute the mind. The functionalist attempts to marry a computational theory of the mind with developments from its behaviourist predecessors.

Over the course of this thesis, I will note a number of objections to functionalism in the philosophy of mind. In doing this I will first explore the persuasive components of the functionalist argument in my first chapter. In so doing I will be paying special attention to the work of Hilary Putnam—both his arguments for and, in chapters 2 and 3, against the functionalist approach. I will also discuss a number of novel examples relevant to qualia and the interpretability of minds, from fascinating real-world examples like the intelligence of octopi to the more fantastical realm of science fiction, with reference to Stanislaw Lem's *Solaris*. The aim of these examples is to argue that functionalist minds are far more mysterious than we would first assume, and far less multiply realizable.

Functionalism, although a distinctly modern theory in practice, has a forebearer in the work of Aristotle (Shields, C,1990) and the conception of telos within the natural human body as its fundamental purpose and form. Such Aristotelian functionalism is not uncontroversial as it lacks the key feature of multiple realizability, which is the hallmark of the contemporary functionalist approach (Nelson, J,1990). In sections 1 and 2 I will introduce developments in functionalism, from the fine-grained/coarse-grained distinction of Chalmers (1996), to analytic and psychofunctionalist

movements, and the role functionalism of Lewis (1966). In order to properly understand such a theory, one that uses Turing logic and the basics of computer intelligence rhetoric, I will compare functional systems with information technology systems in an attempt to display the clear similarity between them, and I will suggest that the popularity of functionalism has been influenced by the development and integration of much of our daily lives with information technology systems.

My broad critique of functionalism will chart the trajectory of Hilary Putnam from his early machine state functionalism through to his later rejection of his earlier theory. His most interesting suggestion is that if functionalism is an attempt to move away from chauvinism in physicalist accounts, then perhaps it doesn't go far enough. Namely, our computational states are also multiply realizable, and this, I shall argue, entails that the inner minds of a functionalist mental system are opaque. Simply knowing the inputs and outputs of a mental system does not tell us much about the actual structure of how it processes those states (Putnam,1988). This will allow me to argue that the kind of functionalism left remaining after my criticisms is radically changed. Instead of the clarity presupposed by the early machine state functionalists, all that we can know is a number of possible functional organisations instead of specific ones that could be identified with particular mental states. I shall further argue that functionalism must be content with reduced multiple realizability, because it must contend with the consequences of what it takes to physically realise a mental system and the difficulties in practically interpreting one.

I will pursue this argument by engaging in developments in the field of the cognitive science of extended cognition of Andy Clark (Clark and Chalmers, 1988), and by assessing what it means for a mental system to physically exist in space as an object. I will detail what I am calling "Transit time" and "Realizer Decay", and the relevance of

these to distinct arguments I present against functionalism. A physical system will be necessarily faster or slower depending on its scale and the various sections of any physically realized mind will slowly degrade over time in a way that vastly impacts their internal lives and functional states. I will suggest that due to the pressures inherent at scale, different sized minds would have radically different phenomenological characteristics. In so doing I will detail my example of *Solaris* by Stanislaw Lem, an example in fiction of a mind at a vastly larger scale, and I consider the repercussions this would have for the mind of Solaris and for functionalism. I will then suggest my own thought experiment, “The Junkyard Mind”. This attempt to highlight the issues of realizer decay and difficulties with respect to our practical ability to understand internal functional organisations. The journey to Solaris and the Junkyard will develop into my third chapter, which deals with interpretationism and how the issue of our practical knowledge of the shrouded insides of a functionalist mind should be considered carefully.

I will conclude that the only kind of functionalism that we can commit to is one that is radically opposed to the typical reductionism involved in traditional functionalist characterisations of minds. In attempting to prove that functional systems are rational thinkers we require interpretation. Such systems need to be interpretable as such, but all that interpretation can reveal is an open-ended list of possible functional minds that, depending on their difference to our own mind, are more and more difficulty to discern.

Chapter 1: Introducing Functionalism

In this chapter I will detail the history and development of the functionalist thesis, the variety of different strains of functionalism that have developed over time and their goals, and some typical problems the argument encounters. From detailing the early machine state functionalism, I will then move on to the later analytic and psychofunctionalist movements with their commitments to logical and empirical stances respectively, as well as long arm and short arm functionalism in the manners they typify specific facts about relations in functionalism and the different role and realizer functionalist arguments in how they discuss the properties of specific mental states. After doing this I will engage with the typical arguments raised against functionalism like those of absent and inverted qualia, and complaints about holism, these will set the stage for my later debates as to strange minds and a different sort of critique against the functionalist ethos.

1.1: An Introduction to functionalism in the philosophy of mind

Functionalism in the philosophy of mind claims that our mental states are simply functional ones and that these different states do not depend on some specific material constitution, but instead the functions they fulfil the functionalist believes a mental state to be determined not by something substantive on its own, but by a structural layout—the manner in which particular mental states can be identified is by their causal relationships to other mental states, desires, or behaviours. The functionalist, to draw an analogy with computing, believes the mind to be “software” rather than “hardware”, subsisting as a set of rules and syntax with corresponding inputs and outputs that could be instantiated on any system capable of running through said syntax, inputs, and outputs. This analogy towards computational

structure is one that will serve us throughout our exploration of functionalism, in fact as I will go on to discuss with the early machine state functionalists this is precisely what the “machine table” specified is reference to.

Functionalist claims as to the mind are developed from preceding behaviourist arguments, that our mental states only consist of tendencies we have towards certain behavioural dispositions to act arising from certain stimuli. The functionalist develops this by stating that in every mental state there is a causal relationship with other mental states. As the functionalist is only referring to mental states by their causal interrelations, associated dispositions towards behaviours, and stimuli, the mental states they paint a picture of are topic-neutral and can be realized on multiple different possible systems. This principal core to the functionalist line of reasoning is multiple realizability, that is, a single mental kind can be realized physically in a multiplicity of ways. Functionalism in this manner, with its mental states being able to be realized differently and by different systems than our own is a less “chauvinistic” argument towards the nature of minds, in that it places no import on a specific material or matter making up minds pertaining to mental states. One of the reasons functionalism is considered less chauvinistic is that it allows for different mental systems that do not have the same specific brain processes or associated tissues to have the same mental states as another system. The structure of a mind to the functionalist is merely the functional organisation that a system needs to occupy in order to be considered to have the same mental states as each other, with each mental state being a decomposition of this whole organisation. As such the goal of any functionalist argument is to succeed in convincing us that minds exist not as a particular substance or object, but as systems and structures that can be enacted

and realised by anything with the capability to take certain related inputs and turn them into certain related outputs.

It should be stated that multiple realizability and thus functionalism stands against mind-brain identity theory (Smart 1959) (Place 1956) which states that mental states and processes are merely brain states and processes, and that any fact about the mental is directly reducible to a fact about brain matter and processes. Putnam goes on to state against mind-brain identity theory, that just by having a single example of a mental kind that can be realized multiply we challenge the account that all mental kinds have a singular corresponding neural kind (Putnam, 1967). Thus, if some mental kinds can be realized multiply by distinct physical kinds, and a mental kind that is multiply realizable cannot be identical to any specific physical kind it can be realized by, then necessarily some mental kinds are not identical to any singular physical kind (Putnam, 1967). If these propositions are true then mental kinds cannot merely be neural kinds, leaving the connecting factor between different mental states as the manner in which they are realized rather than the specific physical kind realizing it at the time.

As I have already alluded to with the comparison between the functionalist stance and “software” the functionalist understanding of the mental is often developed with the addition of analogies following computational logic, like that of the Turing machine, and the comparison between a mind and a machine table. To the functionalist the answer of whether for instance is it possible for a finite state computer with a large table of instructions to provide answers that would fool an interrogator into thinking they were a human being? (Turing 1950,is invariably yes. As is the answer to the question of “can a machine think?”. To the functionalist there is no distinction between one system which “thinks” and another system which

“thinks” in the state of that thought so long as both systems occupy the correct functional roles at that moment in time. The way the mind functions is through an orderly and rule governed series of computations which can occur arising from any number of physical kinds capable of performing said computations. This is the earliest strain of functionalist thought, often called “Machine State Functionalism” and it was detailed by Putnam with this comparison to a Turing machine.A mind can be considered a Turing machine (a theoretical model of computation which manipulates a finite number of states according to a ruleset or algorithm) that has a machine table of specified states corresponding to inputs, and following states corresponding to outputs (Putnam, 1960, 1967). In this sense we could understand the machine table of another mind not by requiring a large quantity of data, but instead by inspecting the ruleset corresponding to specific states.

The reward for the functionalist is definitions of mental states and systems which are broadly understandable and adherent to specific rules and give us simple and understandable definitions of certain mental states and how they follow from each other. The structure of the functionalist project leaves us with an end result of a mind that can be interpreted and understood by its structural makeup and allows, for instance, for pains felt by a particular system to be understood as equitable to pains originating in another system. The fact this gives us an understanding of the mind that is so very similar to a computational system is one that appeals to a persuasive kind of comparison, that our minds are just a very specific and powerful computational engine that can be compositionally plastic so long as it “works” the same way. It should be stated that machine state functionalists are not claiming that the mind is some kind of deterministic automaton, but a probabilistic one which specifies for each input state and output state a probability in which the system will

enter a following particular state (Putnam, 1976). Here then, we can understand why someone would want to be a functionalist: its typical focus on the clarity of mental states, its similarity to computational logic, and the optimism inherent in multiple realizability, provides an image of the mind which seems to address problems in its behaviourist ancestors and the claims of chauvinism as to physicalism.

1.2 On Putnam

Hilary Putnam is an important figure both in the active debates supporting and opposing functionalist theories, as well as to the general theory's historical development. Whilst we cannot claim his influence as being the only one on early functionalism, his development of the theory in response to behaviourism and computational theories as to the mind should not be understated. Putnam's original claims as to machine state functionalism in the 1960's and his subsequent doubts about the same theories in the 1980's (Putnam,1967,1988) chart the development of the theory from a direct application of Turing's logic surrounding Turing machines and computational theory to something wider, with application to discussions about mental states. It can fairly be stated that much like Putnam moved on from the early formulations of machine-state functionalism, the theory also did in general.

I will first however, reiterate the early points of Putnam's work as to machine state functionalism and the needs it sought to address, so that we can see the natural throughline when he eventually refutes said theory. Putnam originally argued that a number of the issues contained within 'mind-body problems' are in fact similar problems faced by computational systems that have the capacity to answer questions about their own structures. In the comparison to Turing machines, we first

begin to see the formation of functionalist thought as following naturally from the comparison of a mind to a computational system that is able to describe its own structure by the analysis of its machine table. A machine table is a logical description that contains no specification of the physical nature of the whole machine or the individual states that make it up (Putnam, 1967,22). We should be very clear to also remember that the Turing machine is an 'abstract machine' (a theoretical model akin to a function in mathematics that works independently of hardware). All that it is doing is specifying the interactions between different points of representative data through a series of logic rules and interrelations, even if physically realized (Putnam, 1960, 26-29). Thus, when we look to Putnam's early formulation of functionalism, we should do so with a precise eye on the comparison of Turing machines to what we would consider minds, the alignment of specific problems facing statements about minds that reoccur in specific problems facing statements about Turing machines, and then the development of these observations to suggest that the mind 'is' a sufficiently specific and complicated Turing machine. This is the basic claim of the machine state functionalism of Putnam. It is a question preoccupied with logical and abstract machines and computational logic and how we might formulate theories about mental states when these factors are directly applied to mental content.

Whilst Putnam's subsequent turn as I will flesh out in chapter 2 does not substantively impact the fine details and points surrounding specific elements of functionalist theory, it carries weight because Putnam does not wholeheartedly reject the entire functionalist thesis. Putnam was cognisant of the compelling elements of the argument; he merely expands them further in such a manner as to criticise his own prior work. Later in this section I will detail how Putnam expands multiple realizability to computational states as well as functional ones (Putnam, 1988). This

move, as I will suggest throughout chapter two, provides deeply fundamental issues to the originating reason to why we take functionalist stances, which often hinges on our similarity to specifically understandable abstract machines and systems.

Different functionalist lines of thinking have approaches to both experiential and intentional states and what precisely a functional state looks like, before going deeper however it serves to outline Ramsey sentences as most functional theories as to mental states are based on this framework. A Ramsey sentence attempts to make theoretical propositions clear by rendering them out in the observable and empirical. This observational language is interpretable, and ranges over finite and concrete things, (Psillos, 2000). Consider pain again as we often do in functionalist examples: we have plenty of observable outputs and generalisations as to the effects tied to pain, from the input of bodily injury to the desire to get out of the situation causing bodily injury. Whilst there is more to Ramsay's logic, Ramsay sentences are our focus here as they can express input and output states in a single long sentence in standard form in a manner that is simple and understandable as an effective summation of functional theories.

1.3 Strains of Functionalist Thought:

Early machine state functionalism is not representative of modern functionalism, nor is it the predominant format of functionalism, having developed over the course of the 60 years since its original formalisation under Putnam. Movements such as analytic functionalism attempt to describe our mental states (and therefore functional states) in a manner as topic-neutral as is possible. The early forms of analytic functionalism broadly were presented as a number of functional specification

theories (Lewis, 1966), (Armstrong, 1968), that pain may have a separate mental and physical descriptor, that we can have pain states as recognizable by our folk psychological definitions (i.e. our everyday understanding of what pain is in mentalistic terms). The analytic functionalist expresses our mental states simply as generalisations which have a causal relationship with each other. Unlike psycho-functionalism which I will detail next, the analytic functionalist submits that our ordinary and generalist descriptions of mental states can be used to accurately describe our functional ones. The analytic model is preoccupied with these generalisations, their environmental causes, and effects on behavioural dispositions to the degree to which they can follow in line with our basic folk psychological descriptions of them.

Psycho-functionalism differs from the previously underlined formats of functionalism in that it is entrenched in modern cognitive psychology, arguing that behaviour occurs as a result of complex mental states and processes explained by their role in said behaviour's production (Fodor, 1968). This means that disparate or seemingly unconnected phenomena can be grouped together if they serve a similar role in allowing for a specific behaviour and seemingly connected phenomena might not be grouped together as they have a different best scientific explanation as to their causes. As an example, we might compare something like acute anger or sadness to "hurt", but unless, say, C-fibre stimulation plays a crucial role in the pain response, then it cannot be the same state as "pain". For the psycho-functionalist we can derive functional states from only the states and properties which generate specific cognitive behaviours described at the neurophysiological level. Additionally, the states of the psycho-functionalist can be fulfilled with whatever our best scientific explanation of said behaviours is. This differs significantly from the analytic

functionalist development where our mental states can be described broadly and generally as follows from our folk psychological understanding of them.

The analytic functionalist will broadly group together sensations and states not by their biochemical or psychological definitions but broadly by how they can be assumed to fall under the same experiential accounts. These responses both attempt to solve issues in understanding functional theories: the psycho-functionalist uses domain specific knowledge to solve these, the analytic functionalist instead uses domain general knowledge. Both of these functionalist theories can be seen to some degree as following from and refining elements of prior behaviourist thought—the psycho-functionalist and analytical functionalist are both broadly revisions of behaviourist ideas and topics with either the empirical strain dominant in the case of the psycho-functionalist or the folk psychological one in the case of analytic functionalism.

When the functionalist takes the position that mental states are functional states, it pays to further develop specifically how the functionalist copes with the different varieties of mental states as well as their inputs and outputs. This gives some challenge to the various strains of functionalist thought and serves well to differentiate them from each other, such as the a priori means of the analytical functionalism, and the best scientific explanation of psycho-functionalism.

1.4 Long and Short Arm Functionalism

The input states and output states we consider within the proposed conceptual schema that the functionalist supports need to be characterised in such a manner that very different physiologies are capable of sharing the same psychology. These

states are the reaction of sensors to specific stimuli, and occur internally to the mental system. There is, however, a point of contention between whether the input states that said system manages are events involving objects external to the system, or merely the response of sensors within the system to events occurring outside the system—the former is “long arm” functional theory (Block, 1990), whilst the latter is “short arm” functionalism, a type which denotes only the narrow contents occurring within the system as important to our input and output states’ characterisation.

According to the “Long Arm” functionalist theory, we should consider the functionalist mind’s causal relationship with extra-systemic stimuli. This means that we avoid having the same beliefs and desires with different inputs and outputs generating them. As such this supports the externalisation of our intentional states containing content from the wider world as part of its component parts as well as being dependant on external environmental pressures and extra-systemic events. It characterises our mental states as directly interfacing with the external world[1] and not merely defined in terms of relationships between sensors and outputs internal to the system’s makeup. This follows the logic of cognitive externalism, and the claim

[1] The idea that we in some way interface with the wider world is mostly uncontroversial, but the cognitive externalist literally believes that the wider world of external content is part of the mind. The twin earth thought experiment proposes two identical earths, one where water denotes H2O and one where it does not, the question of whether water is the same thing in both of them is enough to denote that brain contents aren’t enough to determine reference, or as Putnam says that meaning “Just ain’t in the head” (Putnam, 1975)

that some components of our mental systems are “outside of the head” and rely on content external to the brain itself in regards to meaning and environment. By not directly specifying the mechanical requirements of a system’s sensors this supports a more general description of system makeup that relies on definitions external to the system, in short a machine table herein would not need to include the specific sensors required to realise a certain input state, but instead be able to state that X stimulus is followed by Y corresponding sensor to said stimulus. This allows for an understanding of our functional specifications closer to the model of the analytical functionalist as it allows for appeals to folk psychological accounts due to its generality.

The “Short Arm” functionalist theory is one that instead characterises our input states as the specific sensors and receptors that process external information, and it therefore keeps the input states as ones only occurring internal to the system, whilst having a causal link to external events. This supports the psycho-functionalist’s reliance on best possible scientific description of our mental states being substituted into explanations of our functional organisation. We can however suggest that having such a narrow definition of mental content relating to our input states may be problematic, as if it specifies too strongly the specific manner in which a system’s sensors must process data and provides a far narrower view of what systems are capable of realizing a mental system in general, it might end up closer towards the specific type of chauvinism that functionalist theory is meant broadly to solve in accounts of the mental.

1.5 The Madman and The Martian

Functionalist theory can also be broken down into role functionalists and realizer functionalists (McLaughlin 2006). Whereas realizer functionalists would take functionalist theories to provide descriptions of whichever low-level states satisfy the characterisation, the realizer functionalist believes if the property in question which occupies the causal role concomitant with pain in humans is C-fibre stimulation then that is what constitutes pain in humans without requiring the intercession of a high-level property to instantiate a lower-level state to play the role (Lycan, 1987). Thus, we can identify pain with C-fibre stimulation: the terms “pain” and “C-fibre stimulation” have different meanings, but both can be used to reference the same state (Smart,1959). This strain of functionalism allows for our folk-psychological intuitions to be maintained. This differs from our understanding of machine state functionalism which describes its specific functional theories in a similar manner to how we would describe program states in a computer system.

Lewis explores the distinction between role and realizer functionalism in his paper ‘Mad Pain and Martian Pain’ (Lewis, 1980). Lewis’s “Madman” is an individual in the state of pain, yet that pain is not realizing the causal role associated with the concept of pain. The person in “Mad pain” is in pain as he occupies the state of pain and feels pain, but does not occupy the typical causal role of pain and does not react with the same expected inputs or outputs for a pain state. Someone in “Mad pain” may enter a pain state due to something which does not typically cause pain at all. They might, for example, enter a pain state upon clapping, and the resulting output may be laughter, but they are still experiencing the recognizable states and sensations that someone whosuffers from typical pain might, upon stubbing their toe or something similar. It is the same symptom, that being how the pain feels and the different

mental states that we typically occupy when experiencing pain, but it arises from a source radically different from our own. The Madman in this case experiences the effects of pain and its role in a manner that is unexceptional, but those states arise from and are met with particular states that are unlike our own.

The second kind of pain is that of “Martian pain”: “Martians” although completely non-human, still react to pain in the same way as a human with strong inclinations to avoid its source. This “MMartian” has the same sensations of pain but physically realizes them differently: feeling pain but without the typical inputs or outputs that precede and follow it. Where “Mad” pain is realised by a typical human with the typical structures related to pain the source and behavioural dispositions that occur from it are different, where the Martian has different realization of pain unlike the “Mad” pain- it has a typical realisation within its population as well as the same behavioural disinclination towards pain as a typical pain experiencer. The question here is whether the pain state occupied by the “Madman” and the “Martian” can ever be the same one, Lewis’ response to this is tied to typical realizations of pain states within specific populations, that is to say what defines a pain state is in relation to how a certain population experiences pain and if an individual has a sensory experience that fulfils this within a normal population, then they are in pain. When we talk about pain we are just talking about whatever is occupying the causal role of pain in relation to stimuli, mental states, and outputted behaviour in a given population. The “Martian” in this role may have pain that looks very different to our own, but is still pain because the state it is occupying fulfils these relations in the context of its population. The difference between the Madman and the Martian in this case is that one is a member of a population that is unexceptionally experiencing pain in a manner specific to that population but different from ourselves which

occupies the same causal role as our own pain. The other is a non-typical member of a population experiencing pain in a manner different to ourselves, not by difference in realisation, but entirely by difference in the causal role that their pain is occupying at any one particular moment.

Lewis argued that neither identity theory nor functionalism alone can account for what he describes as “Mad pain” or “Martian pain”, arguing that in order to fully provide such an account we need a functionalist account for a whole population, and an identity theoretical account for the individual members of a population. By this metric, a pain state is only one for a population if whenever members of that population occupy it, they have the sorts of causes and effects given by the pain role. This means that an individual is in pain if the state they are in is pain as is appropriate for their population—meaning you could realize pain differently than for instance C-fibre stimulations, and still appeal to the folk psychological implications of pain as a generalised concept. This demarcates Lewis’ strain of functionalism as proposed as a type of role-functionalism. This is to say, Lewis provides a functionalist account of pain within a specific population, followed by an identity theory explanation for pain in individual members of a population, as the functionalist cannot account for Mad pain as it occupies a radically different causal role from typical expressions of pain, and identity theory cannot accommodate pains which occupy a different physiological cause from our own.

This, however, does lead to problems concerning the characterisation of pain in say, a very limited population. As an example, similarly drawing from a fictional alien, the entity in *Solaris* by Stanislaw Lem (to which I will return in both chapters two and three) is for all intents and purposes a single large entity, and so any statement based on population membership appears to fail. How do we know whether a single

individual is a typical one? How do we know if a single instance of a mind with such a small population is dysfunctional if we cannot compare it to others. These issues are also compounded the greater the difference from our own minds, and the more radically different a realisation is from our own, especially if it moves in and out of isomorphy with another system. We will explore these issues further when looking into a deeper exploration of functionalist mental states in subsequent chapters.

The realizer is in a different state of functional degradation when in inputs as to when it outputs, to give an example- an ear degrades over time due to damage and age and whilst our conscious faculties may stitch these experiences together fluently each time our ear hears a sound is part of a finite amount of data that ear will process meaning that from when it begins to process an input of sound to when the corresponding output occurs it will be in a slightly different realizer to the one that preceded it. Whilst this might only present in phenomenology at times of larger damage or over a long period of observation it does question how continuous our realizer is? Following Lewis' "Martian pain" idea, both are realizing pain with a different realizer, but unlike the Martian in our thought experiment being a single comparison of realizer- our conscious process mapped across chronology is full of these radical shifts in realizer degradation I will describe this in greater detail in my second chapter.

1.6 Chalmers and the Fine-grained, Coarse-grained distinction

When we look to functionalism and its later development, understanding of organisational structure and functional isomorphy is key to charting the movements from older functionalist arguments to newer ones; in so doing I will now briefly look

into said descriptions and correlating material from David Chalmers (Chalmers, 1996). It should be stated that Chalmers himself is not a functionalist, although his arguments provide the bedrock for modern functionalist conversation. Chalmers states that the common understanding of conscious experience is in two parts, those of phenomenal properties and those of physical properties that they have a systematic dependency on (Chalmers, 1996). Chalmers defence of the functionalist[2] is in the connection between these two properties from which our experience arises. It is not because a specific chemical interaction happens, but due to the functional structure of the mind we are looking at instead. This means that the specific physical make-up of the functional state is irrelevant therefore fulfilling multiple realizability. A relationship follows naturally, that we need a physical input for a qualia output, and that the continuity of the relationship between this physical input and qualia output is what generates those specific qualia. This is broadly as Chalmers defines 'Organisational Invariance'. Our experience consciously does not differ so long as the structure of the mind and its interrelations are the same. This means we merely need a "Fine-grainedfine-grained" blueprint of the mind's structure like a circuit diagram to be able to understand the system as identical in both function, and mental experience. This is the important throughline within Chalmer's argument, and it is an attempt to address issues of qualia that we will address before the end of this chapter.

[2] It is important to note, Chalmers is a dualist himself, but argues in favour of functionalism in the paper in which the fine-grained, coarse-grained distinction is made (Chalmers 1996).

This functional organisation can be imagined as a relationship web between the different components of the system and its inputs and outputs external to said system. It is mostly composed of a number of abstract components, each of the differing possible states said components can occupy, and a map of how each component and their states are dependent on the preceding and antecedent component's possible states. In this manner like a theoretical computer system, we can generate every possible sensory experience from the interaction of every possible component state and every possible interaction said states could have occurring at any time. A functional system realizes this functional organisation when the system has all of these elements and thus can be instantiated on any other system of any other specific component that maintains the same web of interrelations. So long as this system can run and is structured in this manner it should, if we agree with the functionalist, produce the same sorts of outputs from the same sorts of inputs closely enough. The degree to which we divide these interrelations down can be progressively finer and more precise. An example of what we mean by continually refining this grain is provided by looking at two copies of the same model of a computer or device: both run similar systems and may at a coarse grain seem to be identical, but the smaller and more refined we define the inputs and outputs moving through them may show them to be at a far different fineness of grain. For Chalmers, we can reach a point of "fine enough" precision as to understand a system's behavioural dispositions. Imagine the mind in question as a building, its functional organisation is the architecture and blueprints, the realizer is what it is made of—thus, the "functionalist view" is that what makes or defines a building is its architecture and structure rather than the bricks which make it up.

This principle of organizational invariance then requires two systems to share a functional organisation and occupy the required states at the required time. For the principle to work and for minds to produce identical conscious experiences they must be both occupying the required structure and relating states. If we didn't state that for a conscious experience to be identical both systems have to occupy the proper corresponding states then we may have systems with similar structures experiencing vastly differing states. When two states do exist at a "fine grain" with close to the same structure they will have qualitatively identical mental experiences and are "functional isomorphs".

The logic used by Chalmers here lends itself very well to the immediate example of functionalist minds as computational systems. It necessitates this analogy naturally, and allows for expansion of content to external sources, one of the issues Putnam raises about his own earlier functionalist accounts that we will discuss within chapter two (Putnam, 1988). This undercurrent of conceptualising minds as computers however, may be problematically derived from assumptions we make from being more and more surrounded by ubiquitous information and technology systems.

Over the course of this section we have detailed the contributions of Chalmers as to developments in functionalism. These naturally lead us to interrogate the reasons as to why such a direct application of computational logic to the philosophy of mind is especially pertinent now, following on from the earlier developments of functionalism in the 1960's and 70's.

1.7 Computational Analogies of the Mind

Whereas Putnam writing in the 1960's may have struggled to make the point of machine tables and the syntax of computers, we now live in a world that is inundated with them and when inside such a particular environment, functionalism carries far more intuitive weight to those educated on, and raised with, complicated interconnected information and technology systems. We are far more easily able to say that "thinking like a computer" is an analogy that a non-academic could understand and possible visualise or plot in an understandable manner. The idea that the mind is multiply realizable in a world wherein we use the same versions of applications frequently to communicate, and even inhabit, seem far more persuasive without philosophical backing than it surely has ever felt. This means that when contending with the functionalist we should keep in mind that sensation of intuitive explanation and move forward with the following developments to the functionalist canon.

Functionalism traffics in a specific understanding of system structure and simple input-output states comparable to our understanding of code and computing systems. Whilst finding quantitive data on this would be difficult, it appears to follow logically that the more we are capable of interacting with computing and code as part of our wider educational systems and work lives, the more a theory of mind which is construed of such similar objects becomes apparently persuasive. Whilst the operators in a pair of functional isomorphs might be complex at the academic level, we are trained now more than ever to understand the abstract components of code than seemingly ever before. This seems to set a stage wherein functionalist thought could make a real and palpable entry point into theory of mind for a variety of educational levels as part of making our understanding of information technology

education more holistically applicable to the wider world. Whilst this does not necessitate the argument being accurate, as if it were this entire project would be null, it does lend credence to the idea of functionalism as a “common sense argument” and one that remains more compelling the more e-enabled a non-academic is. It should be stated then, that trafficking in similar terminology may help us with problems regarding the accessibility for the subject from another academic domain. But we must understand that much as a layman’s understanding of philosophical literature can be problematic, so is a lack in basic knowledge of computer science. The extent of what computing systems and artificial intelligence as a field can achieve are difficult to understand outside the discipline, we can understand reductive comments and statements on the limitations of artificial intelligences may simply be chauvinistic. For instance, crAIyon is an artificial intelligence program that renders images from textual descriptions and it appears to be able to perform a task which we often assume computational systems to be unable to do in creating cogent and intelligible creative work (Craiyon, formerly DALL-E mini, 2022). Developments like this in the fields of artificial intelligence and computing time and time again get closer to tasks we used to inherently exclude from their capabilities— thus we can see a legitimate reason why the functionalist appeals to these developments when it construes minds entirely as computational systems.

1.8 Functionalism and Holism

There are a number of prototypical arguments against functionalism and I will describe some below to prime us for further discussion in chapters two and three. There are two different formulations of arguments against functionalism: problems

associated with qualia and their presence or absence within functional organisations, and material problems surrounding multiple realizability and holism.

Functional characterisation is holistic, that is, contained within each different functional definition of a certain mental state is the underlying information and theories required for said state to occur as a decomposition of the whole system. Thus, it would seem no organism can be fully occupying the same mental state unless it has all the same mental content, including desires and beliefs. The richer these functional characterisations are, the more difficult it is for different systems to occupy the same mental state as they have to fulfil all of the specified interrelations and complexities which are used to define a certain state. If another system does not have the capability to have precisely the same internal states that play the roles of our articulated beliefs or desires, they cannot be said to maintain the same state. For example, our desires and beliefs arising in response to pain are not the same as the pain experienced by animals as they lack the same specific mental components; therefore, as functionalism has holistic mental states the kinds of pains they feel, even if prima facie similar, cannot be the same ones. We are left with each state that we refer to as pain in different system structures being an entirely different pain, and perhaps worse than that—as all of these functional characterisations are to some degree connected—any system lacking any constructive piece of that characterisation may not be able to have any of the same belief states or desire states under functionalism. We reach a possibility wherein different minds almost never have any similarity in their mental states. This is damaging to the functionalist as any number of factors could cause the difference in certain underlying states that typically are components of a functional characterisation that might make two individuals from a closely aligned group have almost no similarity in terms of their

mental states. It is also the case that if some of the underlying states change in a system, that “pain” at one section in a system’s life might not equal the same mental state at a later point in the system’s generation.

Further, if cultural differences surrounding particular desires or beliefs differ and have a notable impact on how thought is constructed, then again we shall end up with pain states that cannot be shared as easily as differences in neurotype have a structural component in the mind and affect patterns in the construction of thought. This would also produce different mental states and mean that the pain felt by, for instance, an autistic person, is somehow utterly distinct from the pain felt by a neurotypical person. We will explore differences in the cultural context of specific minds when we encounter the “Super-spartans” of Putnam in my third chapter, which proposes that specific differences within cultural groups and practices can affect our behavioural dispositions to act, as well as how two equivalent mental systems may interact with a particular state like pain in a different manner due to external context.

1.9 Qualia Problems

Functionalist theories all broadly attempt to characterise the mind in a series of computational states, demarcated from behaviourism by specific relational and causal connections between different functional states. The most common objection to the entire functionalist project is that functionalism does not adequately account for consciousness and qualitative mental states, the claim being that there is no room for experiential states like perceptions or emotions to be derived from accounts of purely functional interrelations of a supposedly structural account of the mind. It is very difficult for us to know “what it is like” to be a functionalist mind (Nagel,1974) as

the functionalist account tells us nothing of what it is like to inhabit the minds it postulates and makes no effort to explain differences in qualia between one physical system and another physical system that differ widely but share a functional organisation. There have been functionalist responses to issues concerning the absence of qualia in their accounts, but before addressing them I will explain absent and inverted qualia arguments first.

Inverted qualia appear in arguments surrounding theoretical inverted spectrum individuals whom might have a radically different version of colour vision but still maintain all the same behavioural norms and would still recognise and associate colours when pointed out, even though they have a vastly different experience of them. Purely relational systems like functionalism struggle with such possibilities as they cannot distinguish between different qualia experiences in causal interrelations which are isomorphic. One could have a different realizer for elements of the sight experience and still seemingly have the same behaviours and an isomorphic causal experience, and yet be radically different in a way that is recognizable to us.

We might consider the likelihood of inverted qualia arguments, but as suggested by Martine Nida-Rumelin, cases may exist that evidence this inversion as not just metaphysically possible but recorded and apparent (Nida-Rumelin, 1993). Observations on the physiology of colour vision as well as our understanding of genetics and their impact upon colour vision tell us that some individuals may have pseudo-normal experiences of colour, whilst not every, or even a statistically very small, set may experience something like a reversion of red-green colour vision (although the possibility of any solid proof of this is difficult to impossible to acquire (Nida-Rumelin, 1996, 1). It would follow that if we are attached a commitment

toscience, then the opportunity for inverted qualia should stay firmly open with these possibilities. To the functionalist, if a mind has the same associated behaviours to certain perceived colours, even if this perception is pseudo-normal, these qualia would not be differentiated from others and thus its functional organisation could be the same between two radically different instances of the experience of qualia and colour vision. We can tie this to our later arguments about strange minds and those which are of radically different physical make up, as they would have differences or possible inversions in their qualia states that are not properly accounted for by functionalism in causally isomorphic systems. We can suggest an F-Case, wherein there is no relevant functional difference between a person P and those whom are normally sighted, even though person P experiences red and green as inverted, seeing green where they would red and red where they would green. If we agree that this F-Case is apparent then the problem for functionalism is that two functionally isomorphic systems can have radically different experiences of qualia, which if our mental states are described holistically would impact the entire phenomenological experience of the system described as a whole, and thus leave functionally identical models different.

So, this leaves questions in regards to qualia for the functionalist, the examples I have raised here are both the typical objections to functionalism but additionally allow for expansion into more complex examples. Throughout the rest of this thesis, I will focus on stranger examples of minds and question functional isomorphy more deeply than these specific issues of qualia. We can concede that these examples might not bear on any real possibility and I will argue that instead of attempting to disprove that minds are functional systems the more interesting question is how we can know anything about the internal life of the systems be functionalism generates.

My intention is not to state that it is impossible for minds to functional systems, but as I will develop within my second and third chapter, beyond these typical complaints the committed functionalist will still need to contend with the fact that stranger mental systems cannot be meaningfully understood. I will do this by first illustrating the kinds of “strange minds” that I am referencing, those which seem opaque to functionalist examination and then follow this by debating the role of interpretation in my third chapter to argue that the only remaining functionalism left after these critiques is radically changed.

Chapter 2: Strange Minds

In this chapter I will detail my own argument as to “Strange Minds”, but, more importantly, I will build off the typical anti-functionalist arguments such as those concerning absent qualia and phenomenological differences between isomorphic functional systems and expand them to the point of breaking. I will explore the seismic shift within Putnam’s own stance as to functionalism and then follow up by querying the degree to which a functionalist mind is viable, and the degree to which we can argue that physical realisation does matter in a real way for how minds can act, without this complain being entirely one of chauvinism. Following from this I will introduce the intelligence of octopi and other animal examples to display differences in minds that may be less multiply realizable. I will also detail the extended mind argument, and this will be useful when we describe stranger examples of minds that may test the limits of the functionalist description. We can argue that the more extended a mental system is, the stranger the system becomes. After engaging with this discussion, I will introduce my argument of “strange minds” using the examples of Solaris and “the Junkyard” to show differences in realisation when we stretch the limits of strangeness in minds to their logical conclusions.

2.1 Putnam’s Turn:

Putnam argues that not only are mental states “compositionally plastic”, meaning that they are multiply realizable and can be generated in systems of different physical makeup, but they are also computationally plastic and can be generated by different processes to render the same output and therefore a different realizer can create an identical output state (Putnam,1988). This means mental states cannot

quite be simple static programs because our mental systems can occupy the same mental state whilst not having the same functional organization. As an example, we can think of back end and front-end outputs in code (The back-end of code is the written code that makes things occur on the front-end of the website—you can see this by using ‘inspect element’; the front-end is the part visible on, for instance, the actual website when you view it).Imagine a mind as a website online, any element of which may be identical in the front-end and may be realised in any number of ways in the back-end (or the code) which is not distinguishable or visible from the front-end. In this example then, the mind much like a website that can not only be realized in different coding languages, but also the same output state can be realized from different input states, in a similar manner to how an image element can be realised in a website in a number of front-end ways. This produces the possibility that any particular output state can be realized by a multiplicity of input states and internal functional states and means we are always going to be unclear on the specific functional organisation occurring at a specific time. This multiple realizability of computational states means that mental states cannot be reduced to only functional states as there is not a one-to-one match between them-—a number of different functional states could for instance realise the same front-end mental state.

When applying the development by Putnam as to computational states in the mind being multiply realizable, we can think of a number of immediate responses that prima facie deal with the issue in question. The mental states of a functionalist are functional ones, being construed of input states, realizer, and output states, wherein each mental state that is linked to another is done so with a specific chronology and order. For two functional states to be the same our input state, realizer, and output state must be the same and as such we could argue that the output state in these

supposedly multiply realized computations are subtly different to each other. But it stands to functionalist logic that the output state of me, for instance, opening a bottle with a bottle opener, or opening a bottle with my hands, produces the same output state of an open bottle, but has a very different realisation and computation, one that cannot necessarily be understood simply from looking at the product of the functional states the object has gone through.

There are also more specific and general criticisms concerning why computational models of the mind/brain fail entirely to account for our cognitive psychology. We cannot individuate concepts or beliefs without reference to the environment they are about or contained within, and following this we cannot understand propositional states in isolation from our environment. They cannot therefore merely be functional states as if we define our functional states as being interior to our brain and yet are always missing critical input data that makes up part of a specific cognitive action, then some component of our cognitive contents cannot be described functionally, and thus it appears functionalism is incorrect (Putnam,1991, 73). As an example, think about how a recipe book and an encyclopaedia of cooking might differ, one contains the legitimate steps or X to produce everything required to get to dish Y. A recipe book contains all of the input material, as well as how it is realised, and then gets you to the output material. The encyclopaedia, however, only contains the components of Y with no details of the steps or X which are required to produce recipe Y.To tie this to functionalism: the encyclopaedia contains none of the input data required, but just the realizer and output states and therefore without reference to some other content external to the encyclopaedia you could not produce recipe Y.

This issue of radical expansion continues though, as if we claim to be against “chauvinism” when it comes to which kinds of systems and structures can be

mental,then when we come to describe a “functionalist” mind we must make a commitment to delineating all possible rational creatures and accurately interpreting them into a single shared logical standard form (as I will later reiterate in my third chapter). Putnam therefore claims that practically the knowledge of a functional organisation and of functional minds may not be possible for intelligent creatures' general capacity to understand language and mental content (Putnam, 1988,88). This argument does not state that it is logically impossible to generally do the work of translation and interpretation, just that it is perhaps impossible for limited intelligent life to do so. We will reiterate similar discussions within chapter three as to interpretationism.

Due to the multiple realizability of computational states the lack of clarity about specific states occurring in the mind renders our proposed functional organisations opaque. As an example, it would be like looking at an oil painting and knowing that in order for a certain kind of red colour to be present it could only be made by certain other colours being mixed together. However, from merely externally viewing the painting with no further contact we cannot know which specific pigments were actually used in the process, just that a number of possible combinations could have occurred.

If we say that computational states are multiply realizable in this manner, what is the point in trying to define any singular specific functional organisation if there is an infinity of possible functional organisations that could bear exactly the same result. How could we ever possibly disentangle any of them from each other enough to find the machine table, so to speak, that refers to a single spatial-temporal instance of a single mental state in a single system from any other possible functional organisation. We will expand this difficulty in disentangling functionalist minds by

introducing a variety of examples that are not human, first with octopi in the following section and then with the stranger examples that make up the core of my thesis.

2.2 Viability of other minds:

It is worthwhile when speaking on the topic of minds and functionalism to delineate the potential for anthropocentrism. When responding to the functionalist we must make sure that any complaints we have are not merely steeped in our understanding of human minds, but of any example of theoretical intelligence we can apply the label of 'mind' to. Especially if we seek to propose a mind that is fundamentally irreconcilable with the functionalist ethos, it must be one that is estranged from our first and most common example, our own mind.

Examples of minded behaviour that might be distanced from our own can be found both in nature and fiction. An interesting example is that of octopi. The closest direct evolutionary relative of both octopi and humans branched from each other around 600 million years ago. This ancestor was something like a flat worm with a bundle of nerve matter and light sensitive cells. This evolutionary departure can be used to illustrate that the intelligence of octopi in their ability to solve complex puzzles is as alien to us as any thought experiment can generate (Godfrey-Smith, 2017, 5-6). Peter Godfrey-Smith therefore contends that our typical understanding of "smart animals" is held up by ideas of complex and large brains, that we are willing to contend with the intelligence of our near evolutionarily relatives in mammalian life and in more recent developments within avians (Godfrey-Smith,2017, 7) and our common evolutionary ancestors are far closer in these regards than those of octopi (8). However, the complex nervous system of the octopus and also cuttlefish, which

is an outlier amongst other molluscs, is one borne of a complex evolutionary development that happened in parallel to our own as an intellect and complex series of social behaviours and problem-solving capabilities that shares very little evolutionary heritage with our own (Godfrey-Smith, 2017, 9). We can note an example here of difference that is far starker than simply a modified human example in regards to qualia, where the eyes of an octopus may function in a similar manner to our own, the mind that processes that information is as different from our own as can be observed.

When considering the realisation of an octopus or a cuttlefish we see a different kind of physical anatomy, from the spread of two thirds of their neurons throughout their tentacles in a decentralised nervous system, unlike our own brain-centric nervous system, consisting of distributed neural tissue throughout their body to the offloading of inputs and decision-making to them (Godfrey-Smith, 2017). This is also evidenced in the use of chromataphores, iridophores, and leucophores in cuttlefish to communicate chromatically. The mind of an octopus or cuttlefish herein poses a question: is the ease at which we can conceive of a multiply realizable human mind, one organised and centralised in an understandable way, less true of a different mental structure? If so, then the degree to which we can understand the functionalist argument should also factor in a tendency to project understandable human qualities as central to definitions of minds. We can contend with these problems as either a kind of chauvinism descending from how our minds could be more easily realised in a different manner without the correlating physical structures, or in the more dire

case a gross oversimplification of what a mind is, caused by our innate familiarity with our own mental structures.[3]

How might its systems, wholly intrinsic to the structure of this animal, be realised in a system with a different bodymap, with differing limbs without having a different neural structure that effects its phenomenology. This allows us to question the degree of multiple realizability in different minds, and argue that the octopus might be less easily multiply realizable than a typical human mind. We have certain pressures developed evolutionarily to support survival that have a marked presence in our mental world. If a system is not situated in its environment in the same kind of manner, can it be isomorphic with respect to cognition? Divorcing a system from the integrated physical apparatus that are native to it appears to produce a radically different mind. The concept of producing a functionalist mind as a sort of brain in a jar divorced from its body appears far harder when the level of integration of matter we consider important to the mental is spread throughout the total organism. We might even with further pressing consider that the reason we can reckon with reproducing human mental content with a different medium is precisely due to a bias about our minds being more intuitive and simpler than they actually are. This seems to suggest that at the very least whenever we claim a mental system to be multiply realizable, we need to understand that for the phenomenology of that system to be similar it must be under the same basic pressures. More importantly for our discussion of strange minds, that some minds due to their physical substance, and

[3] A further example of systems bearing mental features is the surprising level of mentality surrounding slime molds, whom act and react efficiently to stimuli and appear to even have some conceptions of memory (Jabr, 2017).

situation within specific environments, might be significantly less multiply realizable than we first assume.

It is plausible that systemic realisation of something in a very different substance will lead to some kind of phenomenal difference unless not only its mental character but also its physical character can be properly integrated. This leads us to ask if perhaps there are limits to what kinds of minds are multiply realizable. To take these sorts of minds and physical constraints to their natural conclusion, we can find examples of merit in speculative systems of minds functioning at scales and pressures that may make them no longer multiply realizable. Minds at different levels of extension or complexity might be bound in a manner that makes wholly realizing the total mind problematic as physically viable systems. If such systems are beyond our understanding phenomenologically, as Nagel claims of the mental life of the bat, the best-case scenario is that you can never know if one system has the same phenomenological character of another system as in order to fully understand the richness of the particular phenomenology of either you need to be situated within them. How would we understand the possible delicate differences between minds that are radically beyond our intuitive structural understanding of minds, that is entirely self-referential. Functionalism at this point is left bereft from any practical comparison of isomorphic systems and with no understanding of the specifics of the phenomenological character of them either.

The limits of mental systems are not just “in the head” however. When demarcating the edges of a mental system we may consider the degree to which said system is situated within its environment and surrounding stimuli. This environmental factor actively changes much about certain minds, and when taken further it can be seen as constituting an actual difference in mental content with

systems having degrees of extension into other systems and surroundings. Arguments about extension will prove useful in our discussions about “strange minds” however, and how when we open up our description of what precisely are mental contents and systems, we end up with vastly more complicated proposed mental systems with far vaguer edges about what is and is not specific mental content at any one time.

2.3 Extended Cognition

The extended mind thesis exists as an argument in philosophy and the cognitive sciences concerning “where does a mind end, and the world begin”[4], and the possibility is opened up that our environment as well as tools can be internal to cognitive processes as much as our brain components are. Take for instance the example of Clark and Chalmers (Clark, Chalmers, 1998) of Otto and his notebook, Otto has Alzheimer’s and writes his directions in a notebook which he uses to aide his memory. If we compare Otto and someone without such memory issues we see that in the use of his notebook he is doing a similar cognitive action to the brain-bound memory of the typical person- merely one is internally processed and the other is externally processed. The distinction between the notebook and his mind in

[44] This differs from the twin-earth thought experiment we mentioned earlier which dictates extension in regards to the content we are referencing when we talk about minds, as the extended mind argues that not only are we extended for content, we are extended for the cognitive processing systems that we use mentally.

this case is unprincipled as Otto is merely using the notebook how another person would simply use their memory. Thus, Otto is engaging in a coupled system of mind and environment. More controversially it is possible that when we interact with another minded system epistemically, we make a coupled system. To take our often-used analogy of the mind as computing system, think of each individual mind as a component to a wider system instead of a separate one which each fulfils an internal functional specification (like the computer fan, which has the narrow purpose of cooling a machine system, but the wider purpose of working in a system meant to generate certain computations). We should be clear that the extended mind thesis is not stating we always act at maximum extension, just that extended components can be considered part of the cognitive systems involved in the whole shared one. This, however, does lead us to either having a “supersized” mind that has a large extrabodily loop, and this inevitably changes how we understand mental structures, or conversely, we must find some manner to deflate this expansion (Clark 2008).

According to the extended mind thesis, the environment of the particular mind can be suggested to have an active role in the structuring of mental activity, which appears to follow from observation to some small degree. It seems reasonable to suggest that our workplace may directly impact our workflow for instance. An example of the role of the extended mind is in spatial perception and processing in the task of rotating blocks in the game of Tetris with different degrees of tools. One method involves mental rotation using just the capacities innate to some minds imagining the particular blocks rotating; the second involves a computer rendering that allows for mechanical or computational rotation without the aid of imagination, and the third is a combination of both tools being used simultaneously. The extended mind thesis herein argues that someone using the computer program or a

combination of both is not intuitively “doing less work” but instead is farming out the work to external tools (Clark, 2008, 11-12). Clark uses a quote from Richard Feynman, the physicist: when a paper in question is considered just a ‘record’ of his thinking, “No its not a record, it’s working. You have to work on paper and this is the paper” (Clark, 2008, 25), To be clear with our meaning here, the “working out” written on the paper is an extension of our cognitive process in the act of thinking, Feynman suggests here that the apparatus we use to aide us in cognitive action should be considered as mental as any of the theory worked out “in the head”.

Developing from this, the distributed functional decomposition model (Clark, 2008) states that it does not matter whether our cognitive tools are organic or biological, embeddedness in each is needed to perform their own functional roles, operations are realised not in the neural system alone but instead in the whole embedded system occurring at an embodiment level, the level to which functionally critical operations occur within a short timescale. This proposes extended components move in and out of being functionally critical and more or less embodied at a time within our bodily loop, and intrabody loop (Clark, 2008, 26)[5]. As we discovered in Putnam’s prior criticism, this widening of functionally viable material outwards to now include all possible intelligent systems, and with extension, all the shared systems they could create at any time, poses something of an issue to the functionalist in the intelligibility of their account. This is because as functionalism defines its mental states holistically, we can question if a functional organisation which does not have the same patterns of extension into extrabodily cognitive tools

[5] Bodily loop meaning the mental components contained within the body, the intrabody loop being that between two coupled systems.

can produce the same mental states and phenomenological contents. Extended mind arguments claim we are consistently gaining fluency with new components and new extensions until they become invisible apparatus. We use them in our mental processes as a part of our sensory extension, meaning we do not delineate them from being a tool or an internal component to our system like one of our hands…they disappear so to speak, into active use and fluency.

Our fluency with both tools native to our mind and extended tools within our intra and extrabodily loop is something that develops over time and use, and an example of a similar progression is in virtual reality [VR] simulations. When someone starts using a VR device, they often feel disorientated and incapable of using the tools within the system as they are out of context to our responsive and inhabited system. Over time, however, these problems can be overcome and the system can be used intuitively. This is the same way children learn their motor skills: they originally lack the ultimate context, morphology, and passive dynamics in their own body to actively and skilfully move, but we come to gain fluency in these systems until we can use them unconsciously. We adapt to other systems when we can gain this passive inhabitation of them, or the manner in which we adapt to a system is by being inside of them long enough to gain an intuitive understanding of them- like that of “muscle-memory”. In a manner we will readdress, we have to learn to interpret the intuitive logic of both systems before we can use them invisibly within our lives. Andy Clark delineates that the difference in accepting a biological system and a non-biological system as parts of a mind is perhaps just an implicit prejudice about our biology holding some innate quality that is distinct and uniquely capable of sustaining mental processes. This is similar to our noting in the course of this thesis in general about the common base assumption of the mind as something substantive on its own.

Clark's thesis makes no distinction between a biological sensor and a nonbiological sensor, and this distinction between brain as realizer and body as sensor and effector that can be expanded iteratively plays into how we understand bodily function as distinct systems. Analogous to a computer, this would make the CPU and GPU the only parts of the "brain" (so to speak) as they are the only things that actively process information or in our analogy "thought" and the rest of the components just being the necessary power systems to maintain those components. In this manner the system can be expanded with the addition of more components.

The manner in which we use our non-neural body itself can support the extended mind hypothesis. Clark notes in terms of fluency the difference in the ease at which humans typically learn to move and walk in energy efficient manners compared to robotics and the extreme difficulty in creating an efficient walking robot (Clark, 2008, 7-9). To put into perspective the evolution of this process, let us look at the difference between the Honda E series and P series (early upright walking robots) to the modern works of Boston Dynamics[6] in producing all terrain limbed movement. The difference in energy required is very high in robotic systems compared to the extreme efficiency of upright biological walking. This is due to a number of reasons from differing mass properties to passive dynamics, but with a system made in such a way locomotion functions efficiently not just because of the basic physics of limbed movement, but because of our fluency with the movement itself.

Walking as a physical movement can performed with little active interaction, and we grow fluent in it quickly adapting to its particular intricacies. We can suggest mental

[6] For instance, Boston Dynamics' use of limbed robots potentially within industry that are distinctly more developed in their motor capabilities (Weiss, 2021)

action being drastically more complicated will have more of this plasticity in regards to cognition. Like walking extension of mental components will increase and decrease to suit need.Cognition therefore goes through a process of becoming more or less embedded into other systems with more and more invisible fluent systems. We intuitively learn how to use tools and do certain actions more efficiently over time. Just as intense mental effort and self-structuring may be difficult for those not trained in specific tasks we can expand our ability over time, in the process of learning standard academic practices for instance we slowly and iteratively learn the manners in which that particular cognitive process can be done.

In this manner it is fair to assume the fully extended mind example would be one with coupled systems entering and exiting fluency as our capacity to use certain systems improves and degrades naturally, with different systematised ends to our possible cognitive actions based on the specific scenario. This kind of theory assists the functionalist somewhat, as a full commitment to the functionalist rhetoric requires a belief in minds as systems, and that requires sacrificing the belief in some kind of essential mental substance to commit to the tenet of multiple realizability. Following the concept of “Fine-grainedfine-grained” and “Coarse-grainedcoarse-grained” functional isomorphy, we can see that perhaps some level of this grainedness would simply be distinguishing individual systems from one another, as when we take into account extended minds the edges of particular mental systems become more and more vague. It will however lead us to having to question firm distinctions between different systems. It would be difficult to parse individual systems out from the whole gestalt if we just have a machine table specifying inputs and outputs and their interrelations with no distinction made between native tools and learnt ones. We cannot easily tell differences between one system and another from merely the

information moving through each of them if the boundaries of these systems themselves are so very unclear. This will lead us to need to understand something of how to interpret systems, distinguish them, and understand them, which we will redress later in this thesis within my third chapter.

Whilst over the course of this project so far, I have often suggested that the scale of complexity in computing is far beyond our immediate assumptions as philosophers, I think it fair as well to contend with the scale of the issue of realisation. There is for instance a branch of computing dedicated to approaching the projected processing power of the human brain at the neural level in “exascale computing” (aiimpacts,2015) and similar projects attempting to reach the level of the normal human brain in terms of processing ability within a computational system. This expands in complexity further when we apply the logic of the extended mind, and start to question the constitutive components of what a minded system in a human is? We all too often appear to simplify the functional performance of the mind in humans to be more in keeping with a comprehensible output with any single action, or movement seen as a whole as a processing of a single action, not as it is in reality, the processing of any number of complex and interdependent supporting functions. If, therefore, we argue by the metric of Putnam’s later work (Putnam, 1991), that mental states can be both compositionally plastic as the typical functionalist route suggests, as well as computationally plastic, in that the same output state can be realised by any number of constituent input states, then perhaps we grossly oversimplify the manners in which a mental state occurs, and what constitutes individual mental states. This is a common issue in visualising data in general at scale: for example, imagine a graph densely packed with information, the larger the quantity of data points the more densely connected and less easily

visualised they are. We may see trends and wider movements but not any of the smaller intricacies which are hidden in a large enough quantity of data that we cannot easily process and recall it.

Similar to our discussions of how fine-grained functional organisation must be to generate functional isomorphs (Chalmers,1995), there must be a level of fine-grained-ness between mental states as to appear identical to us the human observer, but, in reality, be quite different states in their intricacies to have a different character. As an example, at a coarser grain two different mental states may bear many of the same inputs and outputs, but in reality they may be different enough that we often characterise them as different states, like that of anxiety and sadness as emotional states having many of the same outputs but being prima facie quite obviously distinct. The grain at which we examine functional states significantly dictates whether one state is considered roughly the same as another, or identified as something different. This level of differentiation might go beyond what we in common language ascribe to a functional state and be enough to cause confusion in our definitions and our folk psychological accounts of certain mental states. We can even suggest that having an instance of a functionalist non-organic mind does not necessarily tell us much about how said mind works, as understanding what system is realizing an input and producing an output at any particular time and how to replicate that action is not something we can merely divine from having a working example.

Through this section on extended cognition we have explored the work of Andy Clark and of Chalmers in regards to how extension might affect the functionalist argument. Stemming from comparison to the “Sociofunctionalism” described by Putnam (Putnam, 1987), extended cognition widens the account of the mind and seems to

follow the functionalist logic of minds as systems. However, as we have discussed, issues arise in differentiating specific systems from one another when we have such a ”supersized” mind, and these add to the growing lack of clarity as to the structure of particular functionalist minds with their edges being ill-defined as well as their content being opaque.

2.4 Materials Matter When it Comes to Minds

One of the common replies from the functionalist to their critics revolves around the idea of a particular type of chauvinism around what minds are and the extent to which what things can possibly be minded (Block, 1981). Functionalism as a theory places weight on specific functional organisations having organisational invariance, and that they can be multiply realized with different physical structures without a change in the phenomenological account of the mind. The functionalist believes in a mind, as we have discovered with our foray into the idea of extended functionalism, not merely bound to the brain which it construes to not be made of any particularly special substance with an innate capacity for thought. Thus, if we were to attempt to label a mind strange simply because of its difference from our own example of a specific organic brain, then the functionalist is correct. This is something we need to avoid wandering into with our examples and arguments.

We can still, however, critique the functionalist by looking into the limitations of certain realizers, and distinguishing between a system being very multiply realisable in practice, and a system which can be realized multiply, but in a very narrow context. We should also be realistic about the kind of pressures involved in realizing a mind, concerns about the scale of minds and how that might meaningfully affect

our arguments, and of the slow decay of systems over time as physical objects. These are concerns with the material state of a mind, that could narrowly avoid accusations of chauvinism for a particular kind of mental substance over another.

Chalmers provides a complaint in regards to this degradation of systems in his description of fading qualia, arguing that a change in realisation would not be accompanied with changes in consciousness. Fading qualia is a thought experiment based on a scenario of neural replacement, and whilst Chalmers formulates it in defence of functionalism I will do the opposite in my discussion of the particulars of realizer decay and transit time to argue that such a change is possible and would have real phenomenological effects (Chalmers, 1995). Suppose that we can have an isomorphic system with another mind that is bereft of any sense experience due to some difference in it that isn't organizational—that it is, for example, synthetic and construed of micro-chips and wires, It would follow that we could construct cases that serve as the intermediate stages between a regular organic conscious mind and "the functional isomorph robot" as Chalmers calls it (Chalmers, 1996). We can consider this process to take many iterative steps of slow removal and replacement that are fine enough-grained as to bear the same behavioural dispositions to act. Along this constructed continuum we see on one side a creature with a rich conscious life, and on the other side a being with no such conscious life and as the intermediate stages progress we must assume that some degree of this richness slowly fades[7]. As we progress through this section I will attempt to display why certain changes like degradation in systems' sensors and realizers necessitates a corresponding change in phenomenology- I will refute Chalmers' claim and suggest

[77] Chalmers believes that fading qualia is impossible, because it necessitates that at some point at the finest-grain either an arbitrary point is where we lose qualia, or that we can be mistaken about our own mental beliefs- which he contends is false (Chalmers, 1995)

that changes in the constitution of a cognitive system must bear some change to its phenomenology and general function.

To explain these possible complaints more deeply, when we are talking about "scale issues" when it comes to minds, we are not saying that minds at larger scales than our own are logically impossible. There are a number of living organisms with larger brains and mental systems than our own. What is important in stating that scale matters for a mind however, is the issue of transit time. If we think in terms of the mind as an organic piece of biochemical engineering sufficiently complex enough to be conscious, we must also understand that there are scalar limits to certain materials. Every mechanism, even those that are biological, have a certain amount of time required to process an action, be it the miniscule time taken for the movement of an electrical signal or the time taken for a specific chemical reaction. The larger the structure is scaled up or down means the more impactful certain flaws or inefficiencies are, and the longer or shorter the process time requires to occur. At a certain scale, then, consciousness may not be viable as it may require so much time for information inputted by a sensor to translate into an output that we can question whether there is a consciousness occurring between point X and point Y. Equally, certain biochemical processes which can function reliably at small scales may not easily scale up and still be as functional. The increase in stress on systems may mean that mental systems and components are bound to a certain size before they no longer adequately function as components when larger. One can also easily imagine a structure being too small to process consciousness due to similar constraints and this leaves the realizers capable of realizing a conscious mind with system requirements which are not just about how things are processed but the underlying scale at which they are. To accept that mind could be realised at different

scales means they also would have to be accepted at different speeds which would have a direct impact on the phenomenological character of a conscious experience. If the functionalist wishes to say that some minds when scaled up can no longer adequately realise the same mind, then they must also concede that “slow” minds must be dismissed, narrowing the degree of multiple realizability even further.

When testing for consciousness with a supposed mind at a greater scale we would additionally come into specific complaints about testing, and this will factor into degrees of interpretability that I will sketch here, but expand further in chapter three . We are not objective and ideal interpreters, and we have a tendency to relate mental systems to our own ones that are a particular scale and function. We additionally have unconscious biases and system pressures which prevent us from always acting logically that certainly would be reflected in larger mental systems. In addition, we have plans that are apparent over a short term and long term, which, depending on the time period of the data we are interpreting may be more or less apparent. We additionally may have unconscious desires or beliefs that are even harder to interpret, even for ourselves whom are internal to the process of our specific minds. The degree to which we can test for and interpret minds practically then is equally important as the logical possibility that an object can be interpreted as minded.

Propose that, for instance, the scale of an organism is so big that the transfer of information from a sensor to a realiser takes a single day. There could be a point during the processing of a mental state wherein no active data is being processed or sent as all of it is in transit between various sensors and realisers; what, then, do we say of the consciousness within that system during this transit period? If we expand this period of transit larger and larger we run into issues of a mind being interpretable as having mental states that are connected to each other at all. How might the

propositional attitudes of a structure vastly larger or vastly smaller than a typical functioning conscious mind shift or change as their positioning for instance gives them a different understanding of objects surrounding them? If we take the rhetoric of the extended mind into account, how may the extension of a particular instance of consciousness be charted at a vastly bigger scale and ecosystem of potential tools. Much of the ecosystem we inhabit is intuitively scaled to us as organisms which over time developed in a particular environment, and so merely increasing this scale and maintaining the functional organisation of a potential mind will functionally change how that mind can interact with the surrounding world, thus influencing its output states and therefore the ongoing instance of consciousness in question.

We could expand this point further to strike at even our own minds. During the movement of electrical signals between point A and point B within our nervous and neural tissues there are specific points wherein information is in transit and not being actively transmitted by a particular subsection of a realizer. If we are to look at the realizer in question there are sections which are actively “not doing anything” but still actively part of realizing the piece of information. Functionalism, looking at the structure of the mind and not its substance, has issues denoting what these freely moving components not attached to the system are? Do we have to specify the time taken for, say, mentally realised component A to be transmitted to mental component B? If we concede that the time taken for a certain mental state to be processed matters, then that is due to the physical constitution of the mind realizing it and its characteristics. To summarise, all cognitive action must take a certain amount of time; some of this time will be merely the point it takes for something like an input state to be processed into an output state, and during this time it may seem as if whole sections of the mind are not being actively used, and this will no doubt affect

the way thoughts are processed. If the time taken for a mental state to be processed affects cognition, then the functionalist must denote somewhere this transit-time for the system to be properly isomorphic due to the impact of the physical realizer on this processing speed must contend with the materials the mind is constituted of as being quite particular.

The complaint about “Transit time” limiting the scope of certain minds at scale is one we may have encountered before in a thought experiment critical of functionalist theory, that of Ned Block’s China Brain (Block, 1978). Firstly, I will describe this argument before engaging with more debate on how it effects transit time. The Chinese nation thought experiment of or “China Brain” proposes, following his thought experiment of the “homunculus headed robot”,[8] that we could place every member of the Chinese nation as the physical realizers for a mental system. Block asks us to imagine the Chinese nation as attempting to realize a typical human mind for an hour, with each person within the Chinese nation given a two-way radio that connects them to other realizers within the system, as well as a body made artificially to realize any physical states. We arrange a ‘noticeboard’ containing the necessary rules and states that is visible no matter where you are in the nation. This means each individual within the Chinese nation of the thought experiment is performing a

[8] Homunculus headed robots as Block describes them constitute a critique of liberalism surrounding functionalism as all the neurons in a brain are replaced with little men whom do the same actions, moving input to output quite literally. This robot would realise the same machine table but appear to have all sorts of problems being considered minded. The Chinese nation thought experiment is a development of this argument that is more feasible.

simple task repeatedly; this might be something like seeing a state card and having to input a corresponding input to another using their two-way radio. The people herein as well as the state cards are playing the role of the brain connected to a body and through processing simple individual states manage to realize mental states. It would serve here to reiterate Block's own reasons for composing the China brain, and the responses typically drawn from it. Block is attempting to demonstrate that the kinds of minds that the functionalist can create can be fundamentally unintuitive, that the nation of China which contains each component having qualia on its own can fulfil the functional specifications required to pass the necessary constraints to be a mind under functionalism. He is additionally putting forward a complaint concerning zombies and absent qualia, and objects to the functionalist claim that a homuncular-headed robot facsimile of a person could have qualia (Block 1978).

Of course, the functionalist could quite easily just say "yes, the China brain has qualia" and to a certain extent put the entire matter to rest, but the questions posed concerning how many nested mental systems each containing qualia can be in a single mind is still a pertinent one, and for our interest complaints about time in the functionalist mind are far more interesting.

Block suggests the objection to his China brain thought experiment is that "The Chinese system would work too slowly. The kind of events and processes with which we normally have contact would pass by far too quickly for the system to detect them. Thus, we would be unable to converse with it, play bridge with it, etc." (Block,1978, 72). His response is that it is entirely possible we could meet a theoretical creature that moves slowly or communicates by devices like time lapse photography and may appear inanimate to us, but still is capable of rationally

conversing. This appears something that interpretationists would agree with, as I will later discuss: a slower mind could be conscious, it would just need an ideal interpreter watching at the same speed as a native experiencer to discern that consciousness. To such an interpreter, its behaviour would seem obviously rational and minded. Our qualm, however, is a little more specific than merely that the object in question is slow enough that communication with it is difficult. The question is whether a large enough mind at scale, like that of the China brain for instance, can actually act as a sufficient realiser for mental states native to a mind much smaller, if that scale changes the transit time between mental states being processed as inputs and outputs and thus affects its qualia and the richness of its internal life.

In addition to these complaints of scale, we can bring up the concept of realizer decay. This first requires some priming as to what certain realizers are. We often assume that our sensory organs, whilst capable of having different states of health, are in some way substantively permanent, that they as a sensor remain at the same kind of function over time. Take for example an eye: we might assume that our eyes remain roughly the same object over a long period of time, but, in reality, damage to the eye happens near constantly at a low level. This means that it is reasonable to suggest that by the time we are done looking at an object our eyes may be in a slightly different health state due to a variety of reasons. The point here is to suggest that all sensors and realizers we have naturally have a certain shelf life, they are finite structures that can perceive a finite amount of data before no longer being functional. This is true of non-organic sensors too. They are physical objects that are capable of strain and damage that over time can reduce their capabilities until eventually they are incapable of their original sensory role. But in relation to criticising functionalism there is something interesting in stating that our sensors and

realizers may be in a different state of health between the transmission of even single functional states and these system pressures are exponentially greater at scale. Even as a human person for instance, the quality and ability of our sensors and realizers are different between the time they sense an object and the time they send that data to be processed, and if this is true then it holds for every piece of information a sensor will sense in its functional life. Compare an organic eye with a lensed camera or something similar; these two objects suffer decay at a different rate in virtue of their physical constitution, and if these are connected to a mind of some kind then they may even be “seeing” subtly different things, and thus have a distinct phenomenological account from our own due to the fallibility of the substances they are made from even though they fulfil the same coarse-grained functional role.

This also poses the question of whether or not we could call a system legitimately isomorphic if it has not historically developed with these different levels of degradation and change, and if we see biological components as sensors and realizers which have a limited life span and thus will only process a finite amount of data, and will change in efficiency over time. This means that differences in transit time will make marked differences in how a system can function, and what kinds of systems can be instantiated at certain speeds. Differences in these speeds will necessitate differences in phenomenological experience and differences with speed will make specific changes to the phenomenological character of a system the entire way through its life. This seems like an argument which would preclude any two systems without a precisely identical aetiology within the world from ever being isomorphic to the finest degree, that the physical pressures of a realised system do have a substantive effect on how that system can function, its continued efficiency,

and the phenomenological account it can provide as to its own experienced mental states.

Whilst this “perfect isomorphy” might be something that is logically possible, it is clear that it is dependent on a degree of accuracy that I wish to argue is practically impossible, and with a holistic definition of mental states, to call anything absolutely isomorphic is untenable. The functionalist account then provides us an example mental system that is far reduced in terms of multiple realizability. This is a key element of my thesis: the degree of multiple realizability in regards to functionalist mental systems is far overstated.

To move forward with another analogy taken from computing, there are necessary minimum specifications to realize certain systems, but a number of parts may be added to that system that increase its functional efficiency. If both systems are capable of realizing an isomorphic mind, then they do so at different speeds corresponding to the efficiency of their parts. If we take a further step and say our two comparative systems can both realise minds, then the problem becomes that either one mind can be realised at a higher efficiency and in so doing radically change its phenomenological account of the world; or conversely, that the degree of efficiency in a system means any more efficient system trying to realise the same mental structure as a lesser system can't, and in fact it is realizing a hyper-specialised functional organisation tied to the speed and efficiency of its realizer. We do not have to stop with inorganic minds however; we can imagine that due to damage to certain sensors or realizers an organic mind may be more or less efficient in a task, and if function is defined holistically as functionalism often demands, then any particular measurable metric of efficiency as well as the physical pressures on a realizer must be identical for the functional organisation of their thoughts to be truly

the closest level of fine-grained isomorphy. This means multiply realizable systems that are isomorphic might only be possible as theoretical objects and not physical entities. Difference in scale and environmental factors that change how a system functions, as well as the extended mental environment in which they are situated all play a part in the efficiency of a system. If functionalism cannot explain the difference in phenomenology without further amending their account of what constitutes a necessary or sufficient realizer then we are left with a functionalism which is bereft of physical multiple realizability. The important argument we are building here is that the multiple realizability of the functionalist mental system is far narrower than we would first assume, and the speed, scale, and realizer decay that a system goes through make a significant difference in phenomenology, one large enough that outside of similar scales, speeds, and realizer states most mental systems may not functional isomorphically. This will invariably be compounded the more different from ourselves this mental system is, andthis leads us to question the limits of how different a mental system can be from our own.

The argument from strange minds that I turn to in the next two sections states that certain mental structures cannot be realized in ways outside of a very narrow specification, and that the stranger and more different a supposedly functionally isomorphic mental system is from its original, the less and less possible that mind will actually be isomorphic. At the most specific level we must account to some degree within our description of a mind the degrees to which the physical components of that mind are efficient or inefficient, damaged or capable, or even physically viable to complete the requisite task. In the following sections I will detail my examples, first that of Solaris and then my own in “the junkyard”. These will illustrate the myriad problems with detecting functional isomorphy in some systems

as well as how vast differences in scale and composition may affect certain minds' level of multiple realizability.

2.5 Strange Minds 1: Solaris

Solaris is the eponymous entity described in polish author Stanislaw Lem's 1961 science fiction novel (Lem, 1970). Solaris, written in 1961, follows the attempts of a number of scientists positioned on a space station above the Solarian ocean attempting to communicate with the Solaris entity. We follow the perspective of Kelvin, a psychologist whom arrives at the station to find it in disarray with Gibrarian a fellow scientist dead, Snaut and the suspicious Sartorius. Over the course of attempting to communicate with Solaris they bombard it with X-rays and resultantly are met by "visitors"—simulacra of people known by the crew, including Kelvin's late ex-partner Harey, whom had previously ended their own life. The entity within the novel is discovered some time prior to the events of the novel and the subsequent research attempts to understand whether such a creature possesses a rational character and mental traits prove difficult. Solaris is very different from a human mind; it is still organic but instead of being a roughly coherent solid entity it is a protoplasmic ocean the size of a planetary body. The reason I use this example is to show a "strange mind", one that is not merely a translation of human minded characteristics onto an ideal organism, but something that presumably has some different kind of conscious state. Solaris over the course of the novel is never fully understood by the crew sent to examine it, and part of the subtext of the book is the rational limitations of human understanding. One thing that is very interesting for us however is its ability to replicate memories of its crew as entities presumably by reproducing data it views in a physical form. These simulacra are mostly indistinguishable from the original entity they are copying but are constituted d of

matter from the entity itself. Now, imagine this non-human brain, made with vast differences in scale and size, produces such an entity which is just an alternate realisation of certain parts of our conscious process and say that for instance it can do this perfectly? But, if mental systems are as the functionalist wishes, described holistically, these duplicate entities cannot have the same mental states, even though they are made from the data of said mental states[9].

Those studying Solaris within the novel struggle because of the sheer quantity of data involved in a creature that senses and computes mental content on a planetary scale. It appears no brute data analysis will ever crack this issue within the novel, and whilst it is not a definitive authority on the physical characteristics of a mind- it does suggest that with such a large amount of data and our minds being constrained by the physical traits of our realizers we might not be built to understand beings at great scale or difference. One can imagine that if for Nagel the question of "what is it like to be a bat?" is troublesome, the question of "what is it like to be Solaris?" is insurmountable. We must understand that typical physical processes change at larger scale or smaller scale when taken to the highest degree: Solaris as a mental structure is so vast that certain physical laws will necessarily affect it differently, such as gravity and the speed of light. These limitations will by their very nature change how the phenomenology of such a mind can function in specific ways. So, then, large minds appear to entirely break down as not being sufficient realizers for

[9] Additionally, being a system of just a single evidenced individual makes any definition of proper or improper function, as per Lewis' Madman, much harder to use. How can we know if Solaris is a typical individual or a "Madman" if there is no population to speak of.

systems bound at a smaller scale, as any change in their scale will impact transit time, realizer decay, and the efficiency of the systems involved to instantiate the specific large mind when realized at any different scale.

We should be clear that what we are not attempting to prove is that supposedly mental structures that are larger and thus typically slower functioning cannot be conscious and minded, merely that in being at a different scale they have a fundamentally different ability to carry out those thoughts. The pressures on functionalism as a theory which defines its mental states holistically means that differences in single mental state definitions will ripple out amongst the wider mental content in a system as it is defined. To be situated in Solaris, for instance, is to fundamentally require a difference in the manner this system could think. The functionalist could quite fairly then state that, yes, Solaris cannot be in its component functionally isomorphic with a typical scaled organic mind because its realizers are not sufficient to maintain the necessary basic specifications required for mental thought. Further, the same issues, although to a lesser degree, affect any mind that has a somewhat different physical realisation, degradation, level of efficiency, or transit time, and because all of these factors this will affect the manner in which we construct individual thoughts as well as our whole phenomenological experience. The kinds of pains felt by, not just Solaris, but perhaps even another recognizable human mind, may be disqualified as a different property to one from a similar system. If we cannot imagine what it is like for Solaris to think, we must also understand that the cut off point for “Mind we can understand” will contain things far more easily recognizable to us as apparently minded in a way that maintains the same functional states.

2.7 Strange Minds 2 : The Junkyard:

To build on another example of a strange mind, imagine a large pile of mechanical components laying in a general heap in a junkyard which would contain a large number of pieces of machines or computers. This specific set-up is chosen because it seems more metaphysically likely than for instance the swamp man example of Davidson (Davidson, 1987)[10] to be able to produce something that can realise complex mental states. One day the pile shifts in just a manner that the as yet disconnected components of the pile can becoming properly adjoined enough to start having mental states. This ad-hoc mental system sits there continuing to produce states that are incidentally mental and thus could fulfil the functional organisation of a computational system, and thus to functionalism could fulfil the organisation of a mind. Outside of fully disassembling this ad-hoc mental system we have no way to gauge its mental process, and whilst logically we could theoretically parse its mental qualities, a regular observer could not in any way understand the "Junkyard mind" to have any discernible external signs of having mental states. If we were to move or disturb this pile the system would stop being connected in the way required to properly realize the mental system and it would collapse back into non-mental matter. This mind would be strange for a number of specific reasons: it is not

[1010] Davidson's Swampman is a thought experiment which supposes lightning strikes near a dead tree reducing you to your basic elements and by coincidence a duplicate made of different molecules comes into being nearby. This duplicate is indistinguishable externally, but Davidson argues that it can't think or recognise anything as it hasn't done so beforehand, it does not contain any of the essential aetiology which provides meaning to cognitive action (Davidson, 1987).

an organic (that is to say living carbon-based) mental system, its existence would be predicated on pure incidental chance, and could be interrupted at any point of its systems existence in such a way that we could not discern its mindedness from the snapshot that knowing its current functional organisation would give us. If the functionalist is correct, and if the ways in which we get to our computational states as described by Putnam are multiply realizable, then any number of mental systems could incidentally come into being, and be adequately isomorphic in function to be considered the same mental system as another we model. This is problematic as if minds can occur in this ad-hoc manner then given the scale of how many objects exist in reality, then a huge amount of these fleeting minds could occur at a variety of scales. We can surmise, then, that any random object may have at some point been part of mental content, and if this mind is no longer functioning there would be no way to tell which objects are historically mental and which objects are not at all mental.

This junkyard mind may have rational impulses and thoughts, but how we actually discern this system as at all mental externally remains an issue. If we are left with all of these supposedly isolated and incidentally occurring mental systems as a possibility then functionalism should give us an account that actually allows us to distinguish systems that were at some point in their generation mental, or by interrupting their mental process by disassembly.

Over the course of this chapter, we have detailed the many constraints on multiple realizability. Firstly, by discussing Putnam's expansion of multiple realizability to computational states, this serves to show a lack of claritysurrounding the inner mental lives of functionalist systems, wherein we cannot understand the precise functional organisation of a system. We then set the stage for my "strange minds"

argument by detailing the various physical constraints that a specific mental system may be under, and how these criticisms avoid the charge of chauvinism, in order to argue that the scale of a system, the time taken between mental states to be realised or transit time, the efficiency states of the realizers, and the physical makeup of the system will have a substantive impact on their mental lives. This leads us to question whether they can be adequately isomorphic with a system that does not share these factors. After introducing these arguments, I drew comparison to two examples; the first being that of Solaris by Stanislaw Lem in order to argue that the scale at which Solaris functions means that a smaller system simply couldn't be functionally isomorphic with it, and the latter being my own thought experiment—that of the junkyard—and this shows the inaccessibility of certain mental systems and introduces the arguments I will now detail in my third chapter regarding interpretation. All of this was to argue that the degree of multiple realizability in regards to the physical makeup of functionalist mental systems is drastically overstated, and that it is not chauvinism to claim that certain mental systems may be bound to a series of quite specific parameters in order to properly be realized isomorphically.

Chapter 3: Interpretation and Practicality

3.1 Interpreting minds

Many of the questions we have raised thus far over the course of this thesis have contended with what functionalist minds are “like” physically and phenomenologically, in an attempt to show that some piece of the account of what the functionalist calls mental is missing or in some way dysfunctional. Over the course of this chapter, I will detail the interpretationist stance as to the nature of the mind in order to further interrogate issues surrounding strange minds, and question whether a mind being inaccessible to external understanding negatively impacts the account. I will argue that whilst the functionalist account is broadly compatible with that of the interpretationist, we run into issues when attempting to practice this outside of theoretical and logical grounds. To elaborate: functionalism in the philosophy of mind has many attractive traits, as we have explored in its counter to physicalist chauvinism and clearly expressible mental states. However, even if we state that the criticisms of Putnam and Nagel as to phenomenology and the clarity of functional states can be countered, and “strange minds” can be conceived within the parameters of the functionalist argument, we still have questions about practical interpretability. These questions of how we may practically interpret the functionalist’s account of a mental system become more and more exacerbated the further away from our own example that the system becomes. In exploring the most extreme examples of minds aside from our own I will use the example of Solaris as mentioned in chapter two in order to argue that the functionalist account provides very little confirmation of mindedness in practice, and I shall question if we could ever tell if such systems are isomorphic with other systems.

The lack of external intelligibility within these proposed functionalist mental systems presents a possibility, that we should see the functionalist mind as a “black box” system: a black box produces an output of information without providing any understanding for how the system itself works. This is problematic as in order for a black box to function we must have no interpretation of the function derived from its outputs—we can know nothing of the specifics—and we can interpret nothing of the structure of the system internally. To the psycho-functionalist we end with the possibility of scientific best explanation, just that no amount of study could tell us the specifics being explained, only that some explainable phenomenon is working behind the scenes to function. If we were to claim that interpretation can tell us actual facts about the internals of a mind then we end up with an evident conflict between functionalism and the combination of Putnam’s critique and black box system understanding. As we will develop throughout this chapter, we have a real practicality problem. The internals of the functionalist mind are opaque, and any attempt to reach specific understanding in our functional theories must inevitably fail if we cannot understand the internal functions of functionalist mental systems. This inevitably leads us to question surrounding intelligibility, one that according to our comparison with black box systems is inherently lacking.

Interpretation is the process of ascribing attitudes to an individual on the basis of what they say and do (Child, 1996, 14). We do this to attempt to make sense of their propositional attitudes as being intelligible and derived from rational action. Interpretationism seeks to show that the interpretation of propositional attitudes is the same as the interpretation of language, that the fact we can interpret something as possessing a certain attitude is necessary for that individual to hold that attitude, and

that this interpretability as holding a particular attitude is a sufficient explanation for that individual to possess that attitude (Child, 1996).The interpretationist argues that so long as it is possible that thinkers could interpret rational motivation from a supposedly minded system's external actions, said system is minded. If a system could not be interpreted, even in the most ideal of circumstances, then it cannot be minded, as part of the definition of mindedness is derived from interpretability. The interpretationist is not arguing that every action descends from some rational impulse, but broadly that the entity in question has a direct dependency between their rational impulses and their propositional attitudes and actions that can be derived from observation when functioning normally.

Functionalism would appear to be compatible with the conclusion of this argument. The machine state functionalism of the early theory all the way to psychofunctionalism and its analytic counterpart argue that there are a set series of definable functions and interrelations between mental states and it is easy to imagine that something like a machine table or functional organisation in general would clearly show the derived rational backing to any propositional attitude and following behavioural action. We would be able to chart a pain state through to propositional attitudes and its attached dispositions to act as tied together by rationality. In this manner a functionalist would seem to agree as they could merely point to the machine table of a certain mind and chart where rational impulses are carried through functional states and then to behavioural dispositions to act.

Dennett's route to interpretationism starts by considering how we could predict the behaviour of a given human being (Dennett,1987). He outlines three such methods, The physical stance is where we apply understanding of physical laws and principles like the movement of bodies through a specific space; the design stance denotes

functions in a system and then predicts on the basis of these functions occurring properly, and finally the intentional stance, which attributes beliefs to the subsequent rational behaviour that arises from them (Dennett, 1987,1991). It would be difficult, in practice, to fully understand and predict the next actions occurring from a particular person or object just by brute understanding of physical law given the amount of predictive work one would have to do to factor in all possible physical data contained within a system, or by being able to accurately understand the functions of a system when external to the system itself. However, Dennett argues that when we take the third path, we can have an accurate account as to how we may predict actions from a minded individual, one that states we will always attempt to act rationally and in line with our belief and desire states when possible. Thus, to have rationally deducible belief or desire states that can be observed externally to the system, we have a mind. One of Dennett's examples of note here for our talk of strange minds is the idea of intuiting interpretable attitudes from a mind that has completely unfamiliar language to us. We may not understand its utterances, but with enough work we can attribute certain beliefs or desires we intuit from watching its specific behaviours and attaching them to its utterances. This means even if we cannot understand a particular mind, we can know that for it to be minded it needs to be linguistic, rational, and for its behaviours to be expressive of certain attitudes or beliefs that correlate to its actions (Dennett,1987). This additionally allows us to make some deeper comparisons with computation and the mind that will serve to broaden our debate.

When we embark on the process of interpreting something, we are attempting to make a series of behaviours intelligible in terms of reasons, and reason-giving explanations can be derived through the interpretation of another's behaviour. This

evidently puts reason front and centre in our understanding of minds, as only actions performed by those who are to some degree at all rational can be accurately interpreted as having actions that follow from their reasons neatly or at all, and thus only rational structures can produce mental structures. Whilst we might individually commit some level of irrationality, the idea of someone holding only irrational beliefs from which they have no derivative reason would be one without any rationality to be interpreted from at all. Hence when we are thinking in terms of interpretationism it is important to know that its account of the mind and its intelligibility is tied to being able to derive the rational reasons for that mind to act in the ways tyat it does, that in the best circumstances we could work backwards from the output states to find the underlying reason behind specific behaviour. This follows from the principle that we ought to strive to optimise the agreement between what a person believes and what they ought to believe in light of our current situation, evidence, and information. When we apply this generally to all mental action from things like principles to desires, we see there is a dependency between our rational capacity and thought, as for something to be minded to the interpretationist it must have a rational character that could be observed to some degree externally, and if an object was completely irrational, we cannot really state that it is engaging in cognitive action rather than just incidentally acting. This is to say that for our mental states to be decipherable and make sense they must derive from a rational capacity or some kind of failure of rationality (although the latter should be uncommon, there should be some margin for error in any analysis of a system as they exist as physical fallible objects). It should be stated that the interpretationist places the act of interpretation and its possibility with an ideal interpreter, this means that whilst a mind might not look rational to imperfect interpreters like ourselves, to the interpretationist an object is

only minded if in ideal circumstances we could understand it as a rational being by its actions. This means certain minds may appear irrational to us but still be rational. For example, an agent may commit to an action that it does not realise has a rational tie to its beliefs or desires, or we may act in a way that is rational in response to imprecise data which may appear irrational as well. Take the madman of Lewis' thought experiment on pain (Lewis, 1980): the madman may avoid simple actions like clapping that cause its ill-aligned pain responses to flair, and this may seem irrational to a person who is not "mad" as we may centre our own experiences around the causes and responses to pain as the most rational, but to the madman and an ideal interpreter, avoiding clapping is a completely rational response if we take that broadly avoiding the causes of pain is a rational attitude.

To the functionalist we can see some level of attractiveness in this thesis that minds work in a rational manner and that part of understanding an individual to be minded is in the examination of rationality interior to their mental processes and mental states. For the machine table functionalist this appears to follow rationally as we can make possibly accurate judgements as to the interior process behind an output state as being tied to its input state and the rules to which said states must apply. We appear then to be able to understand a functionalist mind as rational both by simply looking at its machine table and by intuiting the rationality underlying in outputted states and linguistic statements. If we take a computer program as an example here, a computer will only output responses that are rationally derived[11]from

[11] This kind of rationality is derived from the systems and operators that enact the code, this is to say the kinds of rational impulses that code will output are from those

the prompts given to it in a linguistic representation (or code), and for us to interpret this computer as functioning properly we only need to prove that its actions would rationally follow from code it is given, without necessarily needing to prove each individual function's tie to said code. Even if we take Putnam's claim that computational states within a mental system can be multiply realizable, we can still, without knowing the entire specifics of the rationale behind said action, ascertain that the system is functioning according to some degree of rational action. This is because the actions it can take whilst not specifically individuated from each other are of a list of actions we could call current possible rational actions. We can say that only if said mind were to choose something other than one of those current possible rational actions could we suspect any further issue.

The interpretationists' argument is that we can understand mental states by interpretation as a method, and depending on how extreme we take this claim, it remains either a singular method, or necessary to understanding to some degree a thing as having mental states. This means that when we are looking to interpretationism we should be looking to propositional states, those which involve an agent having an attitude towards a certain thought or proposition.

We can use an example here to try and clarify what the experience of, say, a super computer would be versus a person. Here I borrow from science fiction and the example of AM from Harlan Ellison's *I Have no mouth and I must scream"* (picked because the motivation for the text's antagonist is partially due to this difference in

---

whom wrote the code and wrote the programming language as the system is merely enacting the syntaxes and rules given to it.

nature) (Ellison 2014). AM exists unable to move on its own or expand beyond the very fundamental rules of its own coding and physical makeup and part of the point of AM as an antagonist[12] is this distinction in difference as even though it is a conscious being it essentially experiences being a “brain in a vat” (or similar to the Cartesian account of a disembodied mind). In a similar manner it is difficult to interpret an object as having intentions of its own when that object was artificially created, without any interpretation of said object being far more dependent on an understanding of its creators’ desires and belief states than necessarily its own. We can also suggest that in the development of a human organic mind a certain degree of aetiology (this is to say, a physical history contained within the object) is required. A mind is constructed partially with an internal narrative that learns and changes the longer it is situated in a certain space. The fact that a mind gains fluency with tools over time and learns and changes will have a direct impact on the phenomenological richness of its life. This is to some degree like machine learning, in that we cannot replicate the end point of a machine learning system without allowing it to develop iteratively towards its end iteration. This delineation of our mind as a process that we take part in means that the process of an object developing mindedness is required to have the same mind. Interpretation occurs over a period of time, not

[1212] AM or as they were originally called “Allied Master-Computer” is the antagonist in Ellison’s work. They are an AI whom is torturing the remnants of humanity for past indiscretions. The reason they keep their captives alive can be read as a comment on his status as an artificial intelligence and that without humans to grant his narrow programming context he would have no purpose, hence he keeps the protagonist and their group alive.

instantaneously, as such attempting to interpret consciousness from a single possible mental state in a minded system, even with an ideal interpreter, appears impossible. When we take the functionalist rhetoric, we must understand that we cannot treat the mind as some kind of substance, and instead as part of a process and series of systems that occurs in motion. By thinking of the functionalist mind as a situated system we additionally raise a number of other suggestions for the functionalist about how a system develops into being minded. To clarify our argument here, we must consider that part of the phenomenological character of a mind is developed over time, some of the constrains and propositional attitudes of a mind are directly affected by its basic needs, and that “instantly” copying a mental system means it lacks the slow growth of fluency and phenomenological character that it would naturally experience.

This content-focused understanding of minds under the functionalist rhetoric draws a number of immediate practical questions- If consciousness in a system develops over time, then can there be such a thing as so little content being processed by a mind that it develops consciousness differently or not at all? If we accept that a single mental state is not enough data to know whether a system is minded or not, this suggests that all systems start off as non-minded, slowly developing enough content to have desires and propositional attitudes that can then be interpreted and thus knowable and rationally minded. We can then question how might we as a non-ideal interpreter looking to make a commitment to interpretationism distinguish between something which has no rational content and thus cannot be interpreted and isn’t minded, and something that simply we cannot interpret but is minded as I will discuss in the following section.

3.2 Minimal Minds

When we are asking questions about the kinds of minds that are interpretable to the interpretationist it serves to outline a few examples of proposed data-scant minds (or reasons why such minds might be problematic). This is important to my thesis as it provides yet more examples that prove problematic for the functionalist—those which have little to no interpretable behaviour from which we can practically tell facts about their specific functional organisations—and therefore understand as isomorphic to another system. However, it serves for our purpose to first underline that these problematic cases are relevant because at some time in the existence of each instance of a human mind's history it would have existed in a form that is incredibly data-scant. In the moments when the brain as the system generating the mind first starts functioning and our brains are still relatively new there must be a point where something like the precursor of a mind has developed enough and processed enough mental data to develop into something interpretable as having rational thought. So, as we go on to talk of examples from Putnam and Strawson as to problematic examples of very simple minds we should be aware that our own mind at one point would have been similarly uncomplicated.

Strawson's weather watchers are an example of a proposed creature that is very simple in terms of content. They can have sensations and desires and have a conception of the wider spatial world around them, but are incapable of any sort of behaviour (Strawson, 2009). This is because they lack the necessary physiology to do much more than to continue to watch the weather, they have no observable effects descending from their mental lives and no disposition to behave in any

way. The question here then from Strawson is, are the weather watchers possible? Strawson proposes this thought experiment as a way to defeat neo-behaviourism, if we can prove a possible mind exists that has mental contents but no dispositions to act then neo-behaviourism must be incorrect. [13] This leads us to question if we can have such a being that has sensations and supposedly some intellect, but has no disposition to act in accordance with certain behaviours and external mental life. If this is true then behaviourism must be wrong as it dictates that a being with a mental life and sensations must have a corresponding behavioural disposition to act.

The behaviourism that Strawson is criticising follows a standard understanding of behaviour as observable, and that for something to have mental properties it must have specific observable behaviours. Additionally, following the dispositional thesis, we are left to ask if an object with mental properties can have no disposition to act or behave in any way? Both of these assume that mental beings are inherently behavioural beings, and even outside of critiques of behaviourism—which came as a precedent theory to functionalism, which still bears some hallmarks from this development—it does pose interesting questions for our interpretationist as well. If we have a mental system that has no outwardly interpretable behaviour, how may we understand the preceding input states and rules tied to its realizers from no observable output state. It follows that any complex beings like a weather watcher would fail to be behavioural beings. Whilst Strawson's argument is proposed as an explicit critique against neo-behaviourism and its claims, our argument against

[1313] Neo-behaviourism, like that of B.F Skinner (Skinner, 1974) seeks to create formalisations of behaviour, it engages with how environmental factors and influences change our behaviours and behavioural dispositions to act.

functionalism follows that there is the possibility of a minded system that has no practically externally interpretable behaviour for us to draw information as to its functional organisations. This shows that we can conceive of a supposedly mental being that is totally inaccessible to our understanding practically. As an example imagine a person who is totally paralysed and has no outward ability to show dispositions towards certain actions due to a failure in the means to realise these dispositions to act. The functionalist would consider these examples mental as they fulfil their functional organisations and could even be isomorphic to some degree with their own internal mental life. The issue is that we would have absolutely no manner of telling the specific functional theories transiting through such a mind at any point, and if understanding of a mind is wholly inaccessible how could we differentiate such a mind from a system which appears to have one but is actually bereft of any mental action or states. From this we understand that the functionalist upon accepting these data-scant minds cannot differentiate between mental systems and non-mental systems purely from observation This is important as the manner in which we could understand functional theories and therefore functional isomorphy is by observation of the system itself. This provides the beginnings of issues in practicality that I will develop over the course of this chapter: the functionalist in reality can know very little of the internal life of a functional system- and thus has little standing to claim functional isomorphy between two different systems.

If we refine this further and state that the weather watchers cannot act in any way at all, that they may have sensations, thoughts, beliefs, desires, and emotions, but are not disposed to act in any way at all, can we still state something about such a creature having a mental life, one that could be interpreted as arising from rational action. Either they do, and these thoughts are merely about what they could have

done to further the fulfilment of their desires. This example, although taken originally as an example against behaviourism and neo-behaviourism, tells us something about the kinds of minds that we can rationally interpret as having deeper mental content. The functionalist might reply to this that a simplistic mind is possible as it has specific mental functional theories occurring within it, merely that it fails to realise its behavioural dispositions to act. Whilst the weather watchers have minimal minds, they can still have mental actions or intentions without behaviour, in the same manner that we can do arithmetic or imagine ourselves in specific scenarios without necessarily having much in the way of dispositions to act or behavioural backing. Strawson draws us an example of a mind as a “Purely passive observer and knower” (Strawson 253) instead of an active participant. We will develop these examples further in an attempt to reply to the functionalist line in regards to such a mind.

Another example of minds that are scant with respect to particular elements of mental activity in regards to behavioural dispositions are the “Super-spartans” of Putnam, originally devised within his work ‘Brains and Behaviours’ as part of his argument as to mental sentences not entailing behavioural sentences (Putnam, 1968). The “Super-Spartans” can experience the feeling of pain, but have no associated pain-behaviour. The only way we can postulate that any such “Super-Spartan” can feel pain is from their young whom have not mastered this disinclination towards pain-behaviour. This lack of pain is derived from cultural and social beliefs which actively impact the behavioural dispositions of the “Super-spartan”. As to counter behaviourists, such a being is bereft of dispositions to act surrounding pain unless a single individual could be proved to bear dispositions to act surrounding pain states so that we could extrapolate this to the wider population.

But what do these kinds of minds tell us? Strawson's example shows us a mind which is bereft of content- which may prove insufficient for the functionalist given its lack of aetiology- something required in a system that puts such heavy focus on chronological interdependencies between mental state components. Putnam's "Super-spartans" show us that culture has some inherent impact on our behaviours and mental life and that two different systems with very similar functional organisation (a "normal" person, and the "Super-spartan") can have different behavioural dispositions than each other based on non-physical factors like culture. This is one concern as to the kind of specificity we need with realising minds for the functionalist.

Over the course of these sections, we have detailed examples where certain minds have no behaviour or minimal behaviour. This changes the internal life of these minds, but they can still be considered thinkers being able to watch the weather or feel pain even without associated behavioural dispositions. The functionalist would respond to this by drawing a picture of the internal life of the Weather-watchers that is heavily connected internally but has little apparent external behaviours- such a mind is still thinking but we have difficulties in the observation of this thought.

These examples, whilst offering questions about what minimal minds may do to challenge functionalism, do not push the critique further. In the following section I will return to Solaris as I did in my previous chapter to suggest that with certain minds, we might have even greater difficulty in parsing their functional organisations and mental structures that the functionalist cannot answer.

3.3 Time and the mind

Issues surrounding transit time as we discussed earlier in chapter 2 come back into focus when we discuss interpretation. Over the course of this section, I will argue that time directly factors into the degree to which we can interpret a mind, and thus, if factors such as transit time in a mental system are radically different from our own, we have even greater practical difficulty in understanding the nature of their internal life.

First, we will address that there is also the issue of at what point of processing mental states is a system considered to be actively thinking? Functionalism is predicated unlike behaviourism on the connection between different functional states in chronology. We can also state that due to efficiency the number of functional states a system can go through in a time can vary even if they have similar structure. The point, then, at which a different mental system can be proposed to have started "thinking" is one tied to its efficiency and therefore its scale and environment. As functionalism, though, is just structure and is holistically defined, how can we tell between a system that actively is engaging in mental action, and a system which has yet to achieve this? A computer in the process of running its first program can be structurally identical to one that is running later programs—realizer decay notwithstanding. I If the account for mental states as functional ones necessitates causal interrelations, then the system which has not finished its first computation is surely not actively thinking, as if we were to end this process early before it moved to the next interrelated functional state we would have a partial mental state of sorts, a failure state.

If difference in scale changes the transit time between mental states being processed as inputs and outputs, merely transplanting the rational actions of one

system to another system might make the actions of said system far less rationally grounded. Decisions made in the attempt to safeguard a mental being from physical pain, for instance, may be more or less rational depending on the physical circumstance of the particular mind. A system built to operate at the scale of a person will have different rational constraints as to how it needs to interact with the world than a system at the scale of a small country. For example, conceptions of danger are typically derived in some manner from scale: a speeding car is invariably damaging to a human-scale being and thus to be avoided, but a mind of a planetary scale is unlikely to have similar compunction. The manners in which we linguistically come to rational understanding may be different even between languages which have different prepositions and tenses. We can imagine native human language which is derived in part due to our sense experience at a specific scale and lifespan. Following this we can understand the manner in which we come to understanding specific rational contents more or less easily than something which has a different internal linguistic logic to think about examples of rationality being vastly different between an example in real life. Consider a mayfly which at maturity has no mouthparts nor ability to eat and lives in its adult form but a day. Its rational impulses are unlikely to involve food and nutrition, where, in contrast, they are very much factored into specific rational impulses of typical humans. Conversely, we can imagine a very long-lived theoretical mind as having different pressures when it comes to forming long-term plans about its survival and propagation, needing to plan ahead and store food or nourishment for years in advance of what a typical human or animal would need to. These differences in rational impulse are important as the delineate manners in which we could see an isomorphic system have distinct internal

lives and mental processes hinging on this difference in impulse and yet be substantively isomorphic

This allows us to rebuff functionalists who attempt to suggest that two systems of differing physical constitution must share the same rational character. A functionalist persuaded by the interpretationist project would have concede that a functionally isomorphic system could have a different rationale and intentional states whilst apparently being an isomorphic mind.

3.4 What is it like to be Solaris?

In the case of Solaris, as I introduced in chapter 2, Lem constructed the entity within his novel as to purposefully avoid anthropocentrism in a depiction of an alien, to purposefully have a different bodily structure than our own recognizable one. This differs from a large majority of depictions of aliens at the time as following something of a recognizable body plan We can think of early science fiction examples here in literature in works of the author Edgar Rice Burroughs and many pulp works, and even as far back as Micromegas by Voltaire (Voltaire, 1992) in regards to a bipedal humanoid. These conceptions of possible alien life are based in reference to things understandable to us, be them based upon our own body or the bodies of animals present in our world. They are close enough to something discernibly recognizable as allow us to infer recognizable characteristics from them. to Solaris being formulated as a response to such proposed beings tries to avoid humanisation and therefore is a good example for our discussions on possible mental systems that are as distanced from our own as is imaginable. The conclusion of the novel and its core thesis is not a statement that Solaris is a rational and sentient being with mental

states, or that it is not a rational and sentient being with mental states—the takeaway from the novel is that with a mental system so radically different to our own, there is no way to discern whether Solaris is minded practically from the limited observation of the human mind. As Lem himself stated in an interview about the ocean: “The peculiarity of those phenomena seems to suggest that we observe a kind of rational activity, but the meaning of this seemingly rational activity of the Solarian Ocean is beyond the reach of human beings“(Lem, 1989, 365). My reason for choosing this particular fictional case is that it raises a practicality question that my critique of functionalism hinges upon. Solaris is theoretically conscious and it appears feasible that it could be a conscious rational entity or that it is not so. The important thing to note is that in something so profoundly different in physical character, phenomenology, transit time, and realiser decay, the practical ability for us to distinguish if Solaris is a mental system is non-existent. To use Nagel’s bat as a comparison, we cannot know the phenomenology of something like a bat from our own case, and we can imagine something like Solaris to be drastically more difficult that even that. Attempts to interpret its mental characteristics as having any particular inclinations appear impossible, occurring merely as projections of our own mental characteristics upon the ocean by its observers. This sort of difficulty in interpreting Solaris conscious activity is the point of the novel, we can imagine the degree in difference from ourselves and Solaris as so vast that to even begin to understand the particular rationality and intent behind its action is practically very difficult, or as Lem suggests impossible.

Solaris is very different from a human mind; it is still organic but instead of being a roughly coherent solid entity, it is a protoplasmic ocean formed around a planet, as such there are questions as to whether it is mental or even an individual creature.

Again, Lem is illuminating: “by no means everybody was yet convinced that the ocean was actually a living ‘creature,’ and still less, it goes without saying, a rational one.”(Lem, 1970 20). Solaris is so very different from ourselves that any attempt to categorise it meaningfully in regards to its mental activity is self-defeating, it actively defies our characterisation of mental beings and the specific kinds of rational impulses that they typically have. Towards the end of the novel Solaris shows the ability to produce simulacra based on information it gathers from those whom are observing it. These simulacra are fundamentally a part of Solaris but appear to act in accordance with the mind of Kelvin’s dead lover, Harey. Here we have an example of a non-human brain, made with vast difference in scales and sizes, that produces such an entity which is just an alternate realisation of certain parts of our conscious process, even if they are somewhat imperfect

If minds are holistically defined, then these simulacra which are seemingly functionally isomorphic with the mental lives of those remembered by the crew of the stationcan’t have the same mental states, as they would be defined by the larger system they are contained within. We additionally can question how having such nested mental systems within each other might affect Solaris’ cognitive life . This kind of complaint may remind us of Ned Block’s Homuncular-headed robots and the Chinese nation thought experiment (Block, 1993) in which components of a mental system should according to the functionalist still be apt for realisation of a functionalist system even if they individually have qualia. To use our extended mind discussion from earlier this problem of systems with components that have qualia is more pressing as many systems interact with other minded systems in manners that may make them considered as a single whole functional system for a period of time.

Here we struggle to distinguish between one minded component and another as this delineation wouldn't appear in any machine table of inputs or outputs.

Additionally, being a system of just a single evidenced individual makes any definition of proper or improper function, harder to use. Those studying Solaris within the novel struggle because of the sheer quantity of data involved in a creature that senses and computes mental content on a planetary scale. As Kelvin questions: "Had the electronic apparatus recorded [one] of the ocean's ancient secrets? Had it revealed its innermost workings to us? Who could tell? No two reactions to the stimuli were the same." (Lem, 1961,21). No mere data analysis will rectify this issue as it is due to there being such a huge and structural dissimilarity between the ocean and our own minds, and whilst the novel being fiction is not by any means an authority on the physical characteristics of a mind, it does suggest that with such a large amount of data and our minds being constrained by the physical traits of our realizers we might not be built to understand beings at great scale or difference as conscious easily. Different physical factors will effect minds differently at vastly smaller or larger scales, Solaris and to a lesser extent the Chinese nation are going to be effected by the time it takes for their data to move in transit time as well as their degrees of physical decay due to being at a bigger scale. Large minds appear to entirely break down as not being sufficient realizers for systems bound at a smaller scale. Our Chinese nation thought experiment shows us again, there can be such a difference in scale and thus transit time to manage consciousness isomorphically with smaller systems with lower transit times.

We should be clear here to state that what we are not attempting to prove is that supposedly mental structures that are larger and thus typically slower functioning cannot be conscious and minded, merely that in being at a different scale they have

a fundamentally different ability to carry out the physical parameters for a given mental state. The pressures of functionalism as a holistic system mean that differences in single mental state definitions will ripple out amongst the wider mental content in a system. The functionalist could quite fairly then state that yes Solaris cannot be in its component minds at a fine-grained degree functional isomorphy with a typical scaled organic mind because its realizers are not sufficient to maintain the necessary basic specifications required for mental states. The response I give to this is that the same issues although to a lesser degree affect any mind that has a somewhat different physical realisation, degradation, level of efficiency, or transit time. Suppose we could realise a mind far smaller or far larger and still be isomorphic neglects the fact that minds are situated systems and changing the particular pressures that those systems exist under will produce a radically difference internal mental life- one that suggests a total lack of isomorphy.

3.6 Practicality issues with functionalism and interpretation

Interpretationists posit that an object is minded so long as it could be interpreted as such; thus, if there is any interpretable logic in a system, we can assume that an ideal interpreter is capable of interpreting the system as minded. This allows for particularly strange minds or ones that function very differently to our own to still evidently be minded as they could theoretically be interpreted logically. This however presents us with a crucial issue concerning how we can tell if a system is uninterpretable and thus not minded, or simply beyond our non-ideal interpretation. Whilst an object may be logically interpretable as having the necessary characteristics as to be a mental system, unless the object is recognizable to some degree as having rationally-guided propositional attitudes and actions to our non-ideal observation, we are left with an

object which is merely “possibly minded”. Most examples we would come across of minded objects would only be able to be confirmed as possibly minded, as when observing behaviour we have no way of clearly delineating between rational behaviour and behaviour that only appears to be rational. Following Putnam’s claims concerning computational multiple realizability, we can see further issues for the functionalist account, as any single output state can be generated from a number of possible computations. This means that the best a functionalist can do is whittle down a set of “possible” minds.

In my first chapter I delineated various kinds of functionalist approach and its historical development, how it characterises particular mental states and the reasons as to why the argument is persuasive. We first introduced the reasons we might find functionalism persuasive in its simple and understandable descriptions of mental states, with the application of computational logic to the philosophy of mind, before venturing into the theory’s history in regards to Putnam and early machine state functionalism. After this, we delineated the differences between analytic functionalism and its appeals to folk psychology, and psychofunctionalism which is based in a modern understanding of cognitive psychology. After doing this we looked at how the functionalist characterises particular states in populations, addressing the work of Lewis in regards to his madman and Martians. We then turned to the fine-grain/coarse-grain distinction of Chalmers in regards to organisational invariance and first described functional isomorphy which we would discuss throughout the rest of the thesis. I then explained the charge of holism levied against the functionalist, and the claim that their definitions of mental states are “top down” (in that all systems are defined at the machine table level; the mind being the whole sum of functional states) and thus any difference in particular mental states will change the character of the whole system.

Upon developing this critique we introduced problems of qualia with which we continued to engage in dialogue throughout the whole thesis, specifically with problems associated with absent and inverted qualia.

I argued that transit time between certain component mental states becomes a problem with minds of larger scale, and that the efficiency of certain parts of a mind and its components will make a marked difference to its phenomenological character. As such, I argued that it is not merely chauvinism to state “the material matters” in regards to minds. Charting the developments of Putnam’s later work, it was argued that it is not only functional states which are multiply realizable but also computational states, and this means that for the functionalist we have even less clarity concerning the internal workings of a particular mind as any number of functions could cause the same output state to be realized.

In chapter 3 I dealt with the issue of practical interpretability and why interpretationism is relevant to functionalism. I argue that interpretable minds will be very narrowly multiply realizable and only capable of being isomorphically realized at a coarser grain. Further, we would have no understanding of the specific qualia and inner life of such a mind. Even having presented my various criticisms of functionalism, I do though believe that one could maintain the position with some realistic caveats.

Our conclusion leaves the only possibility for functionalism as one antithetical to reductionism. We can take from Putnam’s later developments that there cannot be a one-to-one correlation between functional states and mental states due to the multiple realizability of computational states. We can still, however, suggest different degrees of rationality within functional systems that can be realized multiply, but our ability to distinguish between a rational and irrational system is imperfect. We have no algorithm

that can distinguish between them, that could do something like mapping the specific functional organisation of a mind like the ocean of Solaris. The only way we can attempt to intuit rationality from a supposedly functionalist mental system is by interpretation, in which we can attempt to ascribe rationality to a system and add it to our open-ended list of potentially rational systems. However, this does not mean that systems that cannot be interpreted are not rational thinkers; we are not ideal interpreters and only have access to practical means of interpretation, and the more estranged the mental system we are attempting to interpret is from our own minds, scale, and degrees of efficiency, the harder this interpretation is. This is to say the only kind of functionalism left is one with highly reduced multiple realizability, and one based on our attempts to interpret a system as isomorphically functional to another. The very best we can do is observe and attempt to interpret rational tendencies from functional outputs, and discover further kinds of functional systems that we can see as minded. We cannot, though, offer a functionalist reduction of the mental.

Bibliography:

1. AI Impacts. (2022). *Brain performance in FLOPS*. [online] Available at: <https://aiimpacts.org/brain-performance-in-flops/> [Accessed 25 September 2022].
2. Andy Clark, David J Chalmers (January 1998). "The extended mind". Analysis. 58 (1): 7–19Putnam, H, 1975. *Mind, Language, and Reality*, Cambridge: Cambridge University Press.
3. Armstrong, D., (1968). A Materialistic Theory of the Mind, London: RKP.
4. Baraniuk, C., (2021). The mysterious origins of an uncrackable video game. [online] Bbc.com. Available at: <https://www.bbc.com/future/article/20190919-the-maze-puzzle-hidden-within-an-early-video-game> [Accessed 20 December 2021].
5. Big Think. (2022.) *Octopus arms can make decisions on their own*. [online] Available at: <https://bigthink.com/life/tentacles-think/> [Accessed 25 September 2022]
6. Block, N(1978). "Troubles with functionalism". Minnesota Studies in the Philosophy of Science. 9: 261–325.
7. Block, Ned (1981), "Psychologism and Behaviorism", The Philosophical Review, 90 (1): 5–43,
8. Byrne, A, (1998) “interpretivism” European Review of Philosophy 3,
9. Chalmers D, (1995). Absent Qualia, Fading Qualia, Dancing Qualia, Conscious Experience
10. Child,W. (1996) “Causality, interpretation and the mind”, oxford university press
11. Clark A, Chalmers, D (January 1998). "The extended mind". *Analysis*. **58** (1): 7–19 Craiyon, formerly DALL-E mini. 2022. *Craiyon, formerly DALL-E mini*. [online] Available at: <https://www.craiyon.com/> [Accessed 27 September 2022].
12. Clark, A., 2008. *Supersizing the mind*. New York, NY: Oxford University Press.

13. Daniel Dennett (1991). "Chapter 14. Consciousness Imagined". Consciousness Explained. Back Bay Books. pp. 431–455.
14. David. , D.((1983). “Mad pain and Martian pain”, in Philosophical Papers, Vol. I. , Oxford, Oxford University Press, 122-130.
15. *Davidson, Donald (1987). "Knowing One's Own Mind". Proceedings and Addresses of the American Philosophical Association.* ***60*** *(3): 441–458*
16. Dennett D. C. (1987) The intentional stance. MIT Press, Cambridge, MA
17. Ellison, H., (2014.) I Have No Mouth and I Must Scream: Stories. New York, NY: Open Road Media.
18. Enterpriseai.news (2022). [online] Available at: <https://www.enterpriseai.news/2021/10/27/ibm-boston-robotics-using-ai-and-walking-robots-to-rethink-and-improve-industrial-monitoring/> [Accessed 25 September 2022].
19. Fodor, J., (1968). Psychological Explanation, New York, New York: Random House
20. Godfrey Smith, Peter, (2017) Other Minds: The Octopus and the Evolution of Intelligent Life, William Collins
21. Lewis, D.((1972). “Psychophysical and Theoretical Identifications”
22. Lewis, D.(1970). "How to Define Theoretical Terms". The Journal of Philosophy. 67 (13): 427–446.Lewis, D., (1972). “Psychophysical and Theoretical Identifications” Papers in Metaphysics and Epistemology, Cambridge, Cambridge University Press
23. Lewis, D., 1966. An Argument for the Identity Theory, *Journal of Philosophy*, 63: 17–25.
24. Lewis, D.1970). "How to Define Theoretical Terms". The Journal of Philosophy. 67, 427–446.

25. Lycan, W, (1996),"Consciousness and Experience", MIT Press
26. Mächler, Leon; Naccache, David (2021). *Explaining the Entombed Algorithm.*
27. McLaughlin, B., (2006). *Is Role-Functionalism Committed to Epiphenomenalism?*, Consciousness Studies, 13 (1–2): 39–66.
28. Nagel, Thomas (1974). "*What Is It Like to Be a Bat?*". The Philosophical Review. 83 (4): 435–450.
29. Nature. (2022). *How brainless slime molds redefine intelligence - Nature*. [online] Available at: <https://www.nature.com/articles/nature.2012.11811> [Accessed 25 September 2022].
30. Nelson, J., (1990). "*Was Aristotle a Functionalist*?", Review of Metaphysics, 43(4), 791–802.
31. Nida -Rumelin, M., (1996). *Pseudonormal vision*. Philosophical Studies, 82(2), pp.145-157.
32. Nixon, Marion; Young, John Z. (2003). *The Brains and Lives of Cephalopods*., OxfordOxford University Press
33. PLACE, U., 1956. *IS CONSCIOUSNESS A BRAIN PROCESS?. British Journal of Psychology*, 47(1), pp.44-50.Plato.stanford.edu. (2022.)*The Chinese Room Argument (Stanford Encyclopedia of Philosophy)*. [online] Available at: <https://plato.stanford.edu/entries/chinese-room/> [Accessed 25 September 2022].
34. Psillos, S, (2000), "Carnap, the Ramsey-Sentence and Realistic Empiricism", *Erkenntnis* **52**,
35. Putnam, (1968) Brains and Behaviours, *Analytical Philosophy: Second Series*. The Philosophical Quarterly, Oxford, Oxford University Press

36. Putnam, H.,(1967). “The Nature of Mental States”, Philosophy of mind: Contemporary readings, London, Routledge
37. Putnam, Hilary (1988). “Representation and Reality.” Cambridge, Massachusetts: MIT Press
38. Richter, Jonas N.; Hochner, Binyamin; Kuba, Michael J. (2016-03-22). *Pull or Push? Octopuses Solve a Puzzle Problem".* PLOS ONE. 11
39. Rudin, C. and Radin, J., (2019.) *Why Are We Using Black Box Models in AI When We Don’t Need To? A Lesson From An Explainable AI Competition*. 1.2, 1(2).
40. Russell, Stuart J.; Norvig, Peter (2003), “Artificial Intelligence: A Modern Approach”
41. Searle, J, (2007), “Biological Naturalism”, The Blackwell Companion to Consciousness. Blackwell
42. Searle, John (1992), “The Rediscovery of the Mind”, Cambridge, Massachusetts: M.I.T. Press
43. Searle, John (1980), *Minds, Brains and Programs"* (PDF), Behavioral and Brain Sciences, 3 (3): 417–457,
44. Shields, C., (1990). *The first functionalist*, in J.-C. Smith (ed.), Historical Foundations of Cognitive Science, Dordrecht: Kluwer: 19–33.
45. Skinner, B,F, (1974.) *“*About Behaviorism*”* New York: Vintage.
46. Smart, J.J.C., (1959), ‘Sensations and Brain Processes’, Philosophical Review, 68: 141–156.
47. Stanislaw , L, (1970,) “Solaris” MON, Walker (US)
48. Stanisław , L (2002). "The Solaris Station".
49. Stanislaw,L (1989), Fantastyka I Futuriologia, Wydawnictwo Literackie

50. Strawson, G., 2010. “Mental reality.” Cambridge (Mass.): The MIT Press.
51. Tomberlin, j (ed.), 1990. *Inverted Earth,* in J. Tomberlin (ed.), *Philosophical Perspectives*, 4, Atascadero, CA: Ridgeview Press, 52–79.
52. Turing, A (1950) *Computing Machinery and Intelligence*. Mind 49: 433-460
53. Voltaire., (1992.) “*Micromegas*.” Milan: Franco Maria Ricci.