THE UNIVERSITY OF HULL

Understanding Other Minds
An Interrogation of the Theory of Mind Debate

being a Thesis submitted for the Degree of PhD

in the University of Hull

by

Daniel Dewi Walters, MA Philosophy of Mind and Body, BA
Philosophy (University of Wales, Swansea)

(May 2014)

Contents

Acknowledgments

I would like to thank the University of Hull for awarding me an 80th anniversary scholarship to write this thesis. A special debt of gratitude is owed to my supervisor Professor Kathleen Lennon for services beyond the call of duty at every stage of the process. It has been a privilege working with you. I am grateful to my examiners: Professor David Cockburn and Dr Steven Burwood for their thought-provoking feedback. I would also like to thank Dr Tjeerd Jellema and Dr Søren Overgaard for valuable input at various stages of this project. I would also like to thank all my colleagues in the departments of Philosophy and Psychology for providing me with a stimulating environment in which to research. In particular I would like to thank Professor Paul Gilbert and Dr Demian Whiting for their advice on particular aspects of this thesis and Dr Dawn Wilson. I would also like to thank my fellow postgraduate students for constructive feedback and moral support. In particular I would like to thank Sue Walsh. Thanks are also due to my parents for their encouragement and support. Finally I would like to thank Zoë for her patience and support.

1. Introduction

This PhD will focus on two empirically based theories of how we understand others intentional states: Theory Theory and Simulation Theory. It will provide an overview of the origins and developments of both theories. These theories come in a number of formulations and hybrid positions of the two will also be considered. This PhD will also explore a third alternative to these theories that has been proposed; Direct Perception or interaction theory. Later in the thesis I will contrast all three theories with phenomenological and Wittgensteinian accounts of understanding others. This thesis will locate this discussion in the context of a philosophical debate by considering the stance each of the above theories take in relation to the traditional philosophical problem of other minds.

The problem of other minds

The problem of other minds is a general problem about how one can know other people have minds. Less generally it is a problem about how one can know that others are the subject of psychological states like those one undergoes oneself. It is a problem because (on one picture) it is held that strictly speaking we can't know what others think or feel because we never have access to their psychological states. Other minds are contrasted with one's own mind which one knows directly. Whereas one has direct access to one's own psychological states; access to the states of others

must take the form of a hypothesis. The problem is actually a collection of problems or at least the question can be construed in different ways which call for different types of answer. The question how we know other minds can be asking about a number of things:

1. What are we doing when we attribute mental states to others?
2. What justifies claims we make about the mental states of others?
3. How is attribution of mental states to others achieved?
4. How we could ever perform such a miraculous feat as to successfully attribute psychological states to others?

Question one is asking what the nature of this practice actually is. It asks for a description of what constitutes the practice. This leaves it open what level of description will provide the most salient account of what constitutes the practice. Part of the question is about the meaning of our psychological terms and part is about the nature of the practice of our attributing psychological states to other people. These are apriori questions. The answer one gives to these questions will be influenced by whatever preconceptions one has about the task at hand. I will suggest that the theories that will be examined are in the grip of a particular picture which shapes the answers they give.

Question 2 is an epistemological question both about how attribution of psychological states to others are justified in general and how we justify particular

attributions. This question is central to Mill's other minds dilemma. It is not, however, independent of question 1. For how we justify certain claims depends on what kind of claims we take them to be.

Question 3 is a naturalizing question. It is asking what are the mechanisms employed in attributing mental states to others? What makes the practice possible? This looks like an empirical question.

Question 4 is raising a deeper philosophical issue. It is closely linked to question 1. Certain pictures in play in answer to question 1 seem to make the feat of attributing minds to others miraculous. This thesis will suggest that the two theories with which it is centrally concerned are not able to adequately address this question because they have a picture on which there can't be any answer to it. This is the same picture that dictates their responses to question 1.

The inner

It is important to stress that the problem of other minds is entrenched within a particular philosophical picture. It is a picture of psychological states as inner states of an individual. This picture has a number of features. Psychological states are assumed to be unobservable inner states of a person. A sharp contrast is drawn between the psychological and physical publically available aspects of an individual. Once this picture of psychological states is adopted it entails a particular kind of

approach to accessing the psychological states of others. We must rely on what we can observe to make inferences to what is unobservable.

The difficulties raised by the other minds problem are all related to a greater or lesser degree to this picture of the mind as inner and unobservable. The question about how on earth we could ever attribute psychological states to others is purely philosophical. It is as much the statement of a problem as a question calling for an answer. This is because if one takes seriously the idea that one's own mind is the only mind one reliably has access to then this takes one towards solipsism. Solipsism is the view that one can only have valid reasons for taking one's own mind to exist. The correct attitude to adopt towards the minds of others is a radical skepticism. This can be because solipsism involves the claim that only one's own mind exists so claims about other minds are meaningless. Alternatively it can be because the belief in one's own mind is the only belief that can be epistemologically grounded. We cannot know the experiences of other people or even that they have minds at all. On this second construal it is not the case that the idea of other minds is incoherent but that the idea cannot be justified. This picture does not acknowledge any conceptual link between psychological states and behaviour. Therefore there is no reason to assume the behaviour of others signifies inner psychological states or to bring psychological states into any explanation of their behaviour in the first place. For one thing the movements of other bodies could be explained in terms of mechanical processes without appeal to a priori unobservable psychological concepts. For another thing the application of psychological states, as they are

understood on this picture, is not coherent. One of my arguments will be that, within this picture, question four cannot be answered.

The argument from analogy

Most conventional solutions to the problem of other minds treat the problem as a genuine one. The most well known analogical argument for the existence of other minds was put forward by Mill.[1] For Mill the mind consisted of a succession of conscious experiences. If this is all having a mind amounts to the question can arise what evidence could possibly support our belief in minds other than our own. Mill argued we have 2 sources of evidence for this belief; the antecedent and subsequent conditions that characteristically accompany an experiential episode. He explains:

"I am conscious in myself of a series of facts connected by a uniform sequence, of which the beginning is modifications of my body, the middle is feelings, the end is outward demeanour. In the case of other human beings I have the evidence of my senses for the first and last links of the series, but not for the intermediate link".[2]

We know that other people have bodies like our own, which we each know in our own case, can be acted on in ways that can cause us to undergo particular kinds of experience. Other people also display the outer signs, which we know in our own case to be the product of inner experience. Furthermore, Mill argued, we observe

---

[1] Mill 1867
[2] Ibid P.208

that this sequence is as uniform and regular in others as it is in ourselves. In our own case we know that the first part of the sequence can only produce the last part by means of the intermediate link. We can therefore be sure that, in the case of other people, the sequence must also involve some intermediate link which we cannot observe. This link can either be the same as it is in our own case (i.e. the result of inner experience) or it can be of some other type. Mill argues that the most sensible conclusion is that the link is the same. He observes that there is one body which is connected with all his sensations. Next he observes that there are a multitude of other bodies, closely resembling his own. Modifications to these other bodies do not result in conscious sensations for Mill. As modifications to these bodies do not result in sensations in Mill's own experience, Mill infers that these modifications result in sensations outside of his experience. Each body is inferred to be a subject of experience that stands in the same relation to its modifications as Mill's own bodily modifications stand in relation to Mill's conscious experience.

The argument from analogy represents one possible move in answer to questions about how we grasp others' psychological states. However there are a number of well-rehearsed problems with this move. The best known is that it makes an induction purely from a single case, one's own. The theories under consideration here take themselves to be offering alternatives to this argument. And an issue for consideration will be how far they avoid its pitfalls.

Aims of thesis

Against the background of the classical problem of other minds this thesis has the following aims:

A) To offer an exposition of TT and ST and consider the ways they address the questions identified above.

B) To attempt to disentangle the various moves TT and ST make in response to these questions. TT and ST both offer answers to the questions raised above about how we know other minds. Their answers involve both empirical and philosophical maneuvers. This thesis will attempt to disentangle these two and interconnectedly to decipher whether the debate between TT and ST is empirical or philosophical. Doing this will raise a question whether they are necessarily competing theories.

C) To evaluate the ways in which TT and ST are related to the classic strategy the argument from analogy embodies for solving the other minds problem and how dependant both TT and ST are on the picture supporting this strategy. It is pointed out that they are both dependent on the picture of the mind as inner and (at least in the case of other minds) unobservable.

D) To argue that the framework in which their questions about other minds are formulated is itself flawed. If we abandon this framework we cannot even generate the issues in the same way.

E) To examine Gallagher's interaction theory and related theory of direct

perception. It will consider whether interaction theory is simply offering alternative answers to the same kind of questions as TT and ST or attempting to undermine the basis for asking these questions.

F) To utilise the insights of phenomenology and Wittgenstein with a view to arguing that answers given to question 1, above , by TT and ST make it impossible to give a satisfactory answer to 4.These contrasting accounts give us a different picture of what is involved in attributing mental states to others.

G) To consider the relation between these contrasting accounts and empirical data, particularly concerning the operation of mirror neurons. It is argued that the empirical data does not provide answers to a priori questions concerning our practice of ascribing mental states to others; but may well provide empirical insight into the ways in which that practice is enabled.

H) To argue there is a slippage between the personal & sub-personal in both TT and ST. I will argue that personal level concepts are not simply transposable to the sub-personal level. This is linked to the fact I am giving an account of these concepts that is anchored in the kinds of practice that Wittgenstein and Merleau-Ponty describe, in contrast to treating them as theoretical terms , or as anchored in introspectible data

Some key issues

Personal vs. sub-personal understanding

One of the issues which emerges in discussing both TT and ST is the level of description and explanation at which the theories are operating

Personal level accounts of psychological attribution rely on experiential level concepts whereas sub-personal accounts offer explanations in terms of neurological or functional data. Personal level explanations are made in terms of the language of everyday experience while sub-personal explanations appeal to processes underlying what happens at the level of personal level experience including neurological processes. Personal level states are in principle phenomenologically available, whilst sub-personal states are not. Personal level explanations also make reference to an agent as a whole while sub-personal refers to specific bits of the person.[3] This is linked to the issue of phenomenological availability because it is at the level of phenomenological awareness that we have whole person narratives.

The distinction between these two types of description and explanation was initially suggested by Dennett.[4] He distinguished the personal level of whole people, with sensations and engaged in activities, from the sub-personal level of brains and events

---

[3] This contrasts with psychoanalysis which is not necessarily about things which are phenomenologically available but which offers a whole person narrative. In principle it is thought to be possible to make unconscious intentions and desires available.
[4] Dennett 1969

in the nervous system.[5] The former level deals with the "desires, intentions and beliefs of an *actor* in an environment".[6] Dennett was clear that these two types of explanation were to be employed in very different domains and that there could be no isomorphism between these levels. He uses the example of pain. Dennett argues ordinary talk of pain is essentially about something non-mechanical whereas talk of neural impulses travelling across a neural network is essentially mechanistic.[7] The sub-personal level in no way deals with agents or sensations.[8] Sensations are discriminated by people, not by brains.[9] When it comes to explaining what it is about a sensation that prompts us to react a certain way, Dennett claims, all we can say using personal level vocabulary is that a person has the sensation and reacts. If we seek further explanation we must abandon the explanatory level of people, sensations and activities and move to the sub-personal level.[10] However when we do this, Dennett argues, we abandon the subject matter of sensations in favour of physical processes such as the motion of human bodies or the organization of the nervous system.[11] Sensations are non-mechanistic and so any identification with brain processes is incoherent. Dennett cites Wittgenstein's insistence that explanations come to an end as a persuasive reason for keeping the personal and sup personal levels apart from one another.[12]

---

[5] Ibid P.93
[6] Ibid P.164
[7] Ibid P.91-92
[8] Ibid P.92
[9] Ibid
[10] Ibid P.93
[11] Ibid P.94
[12] Ibid P.95

Hornsby claims that "what is visible from the intentional stance is both explicable in personal terms, and is inexplicable in sub-personal terms".[13] She stresses that "what is explained at the personal level *cannot* be explained over again at a lower level".[14] To leave behind the personal level is to move away from its subject matter. When we turn to events in brains and in the nervous system the person is out of the picture. One cannot make an agent intelligible in terms of their inner physical occurrences. Dennett's original distinction involves taking the personal level seriously as a level of explanation. It also involves taking facts at the sub-personal level to be unsuitable for providing explanations driven by personal level phenomena.[15] Instead they are useful for explaining what goes on sub-personally when a person does something intentionally.[16] This includes explaining our capacities to experience psychological states and to recognize them in others.

Hornsby identifies an important shift in Dennett's thought from the way he originally employed the personal/sub-personal distinction to the way he employs the distinction in later work. Later Dennett no longer advocated such a strict division of the personal from the sub-personal. He instead claimed sub-personal theories can be employed to account for personal level behaviour as the product of interactions between the various subsystems.[17] Intentional properties, initially only attributable at personal level become attributed to subsystems.[18] Properties attributed to the

---

[13] Hornsby 2000 P.2
[14] Ibid P.3
[15] Ibid P.6
[16] Ibid P.7
[17] Dennett 1979 P.154
[18] Hornsby 2000 P.8

personal level are equally appropriately attributed to any sub-personal system, be it human or mechanical. Intentional properties are hypothetical ones whose value is assessed by their usefulness in explaining any given phenomena.

This thesis will show that uses of the personal/sub-personal distinction in the two main theories under discussion, TT and ST, is much closer to the way Dennett later employed this distinction and is not compatible with his original emphasis on the separateness of the two levels of description and explanation. This means that it is a lot less clear what is really meant by sub-personal. There are problems about the way in which the concepts and vocabulary are being ascribed to the sub-personal level and the way in which the language of agency has insinuated itself into the sub-personal level. This raises questions about the kind of explanatory links we should expect between levels.

This thesis will argue these theories fail to draw an adequate distinction between personal and sub-personal levels of explanation. The distinction between the personal and sub-personal is a philosophical one which these theories have not clearly made. Considered as personal accounts describing the phenomenological experience involved in encountering others, the theories have no plausibility except in very peripheral circumstances (for example we might adopt theorizing or simulating strategies when we are at a loss of what to make of ambiguous behaviour). To avoid this problem they switch from personal level to sub-personal. However they do this without any acknowledgment about how problematic this

move is. It involves treating the sub-personal as if it is isomorphic with the personal level. These theories rely on a vocabulary that makes sense at the personal level but which this thesis will argue does not make sense at the sub-personal level. It will argue one can't coherently describe personal level states using sub-personal descriptions. This thesis will also argue that the sub-personal level is not appropriately described in terms of personal level concepts.

A consequence of this is that the sub-personal can only be brought in for an account of the enabling conditions as an answer to question 3. Instead of answering the epistemological other minds problem, this thesis suggests, empirical discoveries provide useful accounts of the enabling conditions for personal level practices. They cannot be any help answering question 2. A consequence that follows from arguing that sub-personal activity is not appropriately characterized using personal level concepts is that empirical discoveries about the former do not provide answers to the epistemological justificatory question of other minds.

It will still be an open question whether sub-personal considerations constrain higher level processes. The answer will depend on what conception of these higher level processes one is working with.

Understanding

When TT and ST ascribe psychological states to others there is a question about

what kind of explanatory work they take themselves to be doing. For TT and ST understanding of others is demonstrated through an ability to attribute psychological states to others in order to explain and predict their behaviour. Often prediction is taken to be based on identification of the causes of a person's behaviour. Explanation here means causal explanation. Such explanation can be offered at personal or sub-personal level, though some would argue that certain patterns of mental causation are only visible at the personal level. A second kind of explanation looks for features that make the behaviour intelligible.[19] This can mean rationally intelligible although it does not require that behaviour is fully rational. (For example a person may break something because they are angry even though this will not be constructive). Such intelligibility seems to require the personal level of explanation. And for the later writers I shall be discussing it is interwoven with the practice of attributing psychological states to others. For TT and ST, in contrast, the intelligibility of rationalising link is more likely to be viewed as simply a certain kind of patterning in the causal links, and therefore not something which is necessarily tied to the personal level.

What is at issue in the dispute between TT and ST?

In giving an account of how we understand others, TT and ST theories seem to be raising and answering at least two different questions, though they do not always disentangle them. First, how can we justify our attributions of psychological states to others? (Question 2) Secondly, what enables us to attribute psychological states to

---

[19] See Hornsby and Lennon 1990

others? (Question 3) This is a question about how to naturalise this process, to show how it fits inside a scientific account of the world. This is done by showing that such processes are merely instances of more general types of process found throughout nature.

The conception of the psychological as inner informing TT and ST influences their approaches to the question of understanding others. This picture influences the way TT and ST raise questions both about how understanding is justified and how it is achieved. In order to attribute psychological states to others one must overcome a problem of other minds. How can we justifiably attribute psychological states to others when these are not something we ever actually observe? The two theories take themselves to be in the business of giving an empirical answer to this justificatory question and to be offering contrasting empirically based accounts of how this understanding is achieved. I will try to show that, even viewed from within this framework, that is the framework of viewing the mind as inner, it is not clear that the issue at stake can be answered empirically. This is one of the key claims of this thesis. Even if one accepted their framework it is not clear that it is an empirical matter which theory has the best account.

Moreover it is a central concern of the thesis that the fourth type of question I identified above, is not given adequate consideration by TT or ST. This is the question about how on earth it could be possible for us to attribute psychological states to others. Although these theories are operating within the picture of the mind

as inner they do not give enough credence to the problem of other minds. They have bought into the picture and have set themselves up to philosophically and empirically answer the question of how we attribute mental states to others without having taken on board how problematic things are once one buys into that picture. The 'how on earth is this done?' question is a way of expressing puzzlement how we could overcome difficulties now that the issue has been set up in a way that seems irresolvable. This is different to other types of 'how is this done?' questions like question 3 which merely seek an empirical account of the underlying mechanisms.

In this thesis it is suggested that TT and ST already assume the coherence of the enterprise of attributing psychological states to others and then ask how we do it. There is an assumption that we can at least pick other people out as potential subjects for psychological attribution, whereas the deep problem of other minds raises an issue how this could be coherently achieved. A philosophical account of what our practices are which do not render those practices miraculous is required prior to an empirical account of how these practices get off the ground. In its later chapters this thesis will explore phenomenological and Wittgensteinian accounts of how these practices can get their meaning. It will be argued that TT and ST work with implausibly crude accounts of what the practice of attributing mental states to others consists in. If this is correct and the answer to question 1 which they are working with is not coherent, can we make any sense of the dispute between them? I will suggest that they can only be concerned with the enabling conditions for the workings of a practice. What they do not do is cast illumination on the nature of the

practice itself.

Chapter summary

<u>Part 1</u>

The first three chapters are focused on an analysis of the TOM debate. I analyse
what the important issues are and assess the relative merits of each position. I will
investigate what the theories mean by understanding and explore the empirical and
epistemological issues they address. This analysis is a central concern of this PhD.
Once this analysis has been carried out I will raise more general concerns about the
way the debate has been structured.

One issue that arises out of my examination of the debate is the role assigned to
one's own experience. I suggest that neither theory takes a satisfactory position on
this issue. TT treats all understanding as third-personal and ignores its first personal
aspects whereas ST suggests we start with our own case (I will return to this issue in
relation to Wittgenstein who shows there are problems with getting beyond one's
own case if we begin this way).

Chapter two examines the development of TT. Broadly this is the position that our
knowledge of other people's psychological states is theoretical (and possibly   our
knowledge of our own states too). I examine what is meant by theory. I show that it

can be given a very loose construal or a strict one modeled closely on scientific theory formation. I also examine the closely related issue of what is required in order for someone to be able to be said to grasp or be employing a theory.

Another related issue I consider is how the theory is supposed to be acquired which is a topic of dispute among theory theorists. The main point of disagreement is between learned and modular formulations of TT. I consider empirical evidence that our understanding of others develops in a stage by stage manner characteristic of theories. I contrast child scientist and modular characterizations of TT. Child scientist formulations suggest TT is learned on the basis of experience. This can either be understood as a process of individual learning where children construct and test their own hypotheses or as a cultural process in which they assimilate established theories during their upbringing. Abstract theoretical states and laws are hypothesized to explain others' behaviour. TT views psychological explanation as causal explanation in which the rationalizing or intelligibility links are the basis of the causal links. Modular versions attribute the development of TT to the maturation of a theory of mind module (note that the abbreviation TOM may be employed to refer either to this module or to the debate in general encompassing all positions). This module is often taken to be innate although this is not strictly necessary. This chapter explores the various criteria that have been suggested for modularity beginning with nine criteria originally suggested by Fodor. For Fodor our mind reading abilities are not modular and are incompatible with his criteria. Modularity is confined to input systems whereas mind reading is executed by central systems. I

contrast his approach with Carruthers' conception of massive modularity according to which the mind in its entirety is modular and consists of a large number of functionally distinct processing systems. I examine which of Fodor's criteria have been incorporated in massive modularity, which have been modified, and which have been abandoned. I note that the development of the view that TOM capacities are modular is not entirely linear from its functionalist origins and has more to do with an attempt to naturalize our mind reading capacities by showing that minds in general are modular and therefore continuous with biological systems in general. My review shows that there are a number of dimensions along which versions of TT can differ and that there can also be overlaps between these positions.

This chapter also examines the developmental evidence that has been cited in support of TT. For child scientist versions this consists in evidence that development in mind reading ability follows the dynamic structure of theory change. Modular versions of TT also attach weight to developmental evidence but the developmental pattern they are interested in is evolutionary rather than that of stages in theory formation. I also consider behavioural evidence for TT in particular false belief tests. These test one's ability to attribute a false belief about the world to another agent. This is designed to establish that a person is capable of employing psychological representations to interpret others. This chapter also explores the implications that have been drawn concerning autism which is interpreted as a TOM deficit.

This chapter also examines Leslie's account of our theorizing abilities given in terms of the underlying causal mechanisms. Leslie posits a modular 'theory of mind mechanism'. This is an inference processing device that processes metarepresentations (involving propositional attitude states such as wanting or believing) in combination with literal features of a situation in order to infer states of mind on the basis of observed behavioural events.

An issue this thesis raises with TT is whether we can really think of our understanding of psychological states as revisable in the face of new evidence. It questions whether it is plausible to think of psychological states as hypothetical entities that could be replaced under a better theory. I suggest that revisions could threaten the very possibility of the psychological practices we have. I also take issue with the highly abstract portrayal of psychological states necessitated by the functionalist approach which seem to make the context of a particular state incidental. I question whether it is correct to think of psychological predicates as naming abstract folk theoretical concepts postulated with the aims of explaining and predicting others. Can they really be contrasted with literal descriptions as e.g. Leslie suggests?

In chapter three I set out ST. This is the position that we use our own minds to simulate the mental states of another person then attribute the results of this simulation to that person. I show how ST moves from an awareness of one's own psychological states to attributing states to others. I examine how the theory offers

different interpretations of empirical evidence of the sort TT appeals to including false belief tests and accounting for deficits in autism. ST offers an answer to a criticism I raised against TT in the previous chapter; that it takes a detached attitude to others. ST attempts to envisage the individual as in more direct contact with the psychological states of others. However I bring out reasons why this attempt is not entirely successful. What we are attributing to others on the simulation story are offline representations which is not the same as directly perceiving actual states.

Later in the chapter I focus on sub-personal versions of ST. I pay particular attention to a sub-personal version of ST called embodied simulation.
Sub-personal accounts are intended to mimic the procedure of personal level simulation but implement it at a sup personal level. Two sub-personal components of simulation are postulated by ST. These are the mirror neuron system and the 'who' system.

Sub-personal versions of ST are greatly encouraged by the empirical discovery of mirror neurons. The discovery that we use the same neurological mechanisms when perceiving the psychological states of others that we use when experiencing psychological states is thought to provide evidence for simulation. It is also thought to provide evidence for direct perception of psychological states because MNs respond to intentional goals of the target rather than physical states. This is seen as offering an advantage over TT which construes psychological states as unobservable hypothetical states.

This chapter also examines the who system which is claimed to take care of the attribution (to self or other) part of the sub-personal simulation process. This implies there is a problem of self-other discrimination the mind reader must confront. An apparent virtue of this theoretical standpoint is that the psychological states of others are meant to be understood as no more problematic to discover than one's own. These formulations of ST are not guilty of thinking we have infallible knowledge about our own states. We recognise psychological states first and foremost as psychological states and subsequently face a problem of who to assign them to (ourselves or someone else). However this setup is mistaken because it construes the neurons as involved in answering an epistemological problem rather than giving an account of how our understanding of others might be facilitated at the sub-personal level.

I pay particular attention to a view expressed by simulation theorists that these mechanisms allow one to overcome an epistemological problem of other minds. I raise a concern about whether this evidence really ought to be interpreted as a simulation process. I point out ways in which it is dissanalogous to classic simulation. In particular some simulationists are suggesting we begin with unassigned psychological states which we subsequently have to interpret as belonging to oneself or to someone else. I argue the idea of unassigned representations is implausible. It gets the phenomenology totally wrong. It is a mistake to try and impose an epistemological burden on our neurology. We should not conceive of neural activations as representational states involved in inference

processing. Instead it would be better to think of them as sub-personal facilitators behind our ability to comprehend psychological states. I also suggest that the idea of shared representations at work is problematic because it glosses over significant differences between the way we experience our own psychological states and those of others. One reason the theory does not offer a satisfactory account of how we could be in direct contact with the psychological states of others is that it does not take into account the rich contexts in which such understanding takes place. I consider attempts by ST to build contextual features of interpersonal encounters into the simulation process but argue they are not satisfactory.

One question that emerges in this examination of ST is what level of description the account is intended to be pitched at. ST is meant to drive empathic experiences and attuned encounters which are primarily features of the phenomenological level of interpersonal encounters. I argue that sub-personal accounts cannot achieve this

This chapter also considers an ambiguity in the stance of ST towards the problem of other minds. Whereas traditional ST offers a solution, sub-personal versions appear to want to argue for a dissolution of the problem. This approach is contrasted with the argument from analogy. A version of this classical solution to the other minds problem is also attributed to TT by simulation theorists who suggest this is a key difference between the theories. The notion of we-centric space is introduced as a way of closing an epistemic gap between self and other. However ST still understands itself as confronting an epistemological problem; that of how to

separate oneself from others.

The chapter also considers empirical evidence offered by ST to settle the TOM question. Whereas ST would predict that we mimic psychological states of the other, TT would not and might predict mind reading capacities to be executed by a self contained module. However I will suggest that it is problematic to interpret MNs as involved in mimicking others.

Another issue raised by this overview of ST concerns the appeal to MNs. These might appropriately be appealed to in an account of how we understand others if we take the question to be about empirical rather than justificatory mechanisms. However before the empirical merits of such evidence can be assessed we need a satisfactory account of what understanding others involves. The stipulation that it is an inferential process is problematic. The evidence is being used to state what the practice of understanding others involves rather than to show how the practice is supported at the sub-personal level. However if we do not accept the stipulation of the practice it is less clear what we should do with the evidence. I will be arguing that empirical evidence does not determine how we are to characterise the practice of understanding others.

An issue I raise with ST is that it takes self understanding as given. In order to see itself as providing a response to the epistemological problem of other minds one would have to think of this as a form of knowledge. ST also takes our access to the

psychological states of others to be identical to our own. However an experience of pain is quite different from an experience of observing someone else's pain. In later chapters I will focus on why it might not be coherent to envisage the straightforward extension of one's own psychological states to others. An issue touched on at various points in this chapter is whether ST can be supplemented by theorizing or whether they are mutually exclusive. This issue is examined in more detail in the next chapter.

In chapter four I examine hybrid positions that have emerged in the TOM debate. The chapter begins with questions about whether simulation necessarily requires some theoretical assumptions, for example about the similarity of others to oneself. If so this will mean the boundaries between TT and ST will not be as sharp as first envisaged. The chapter next reviews the dimensions along which versions of TT and ST can vary.

The debate about whether TT and ST are compatible rather than diametrically opposed begins by examining the idea that ST might collapse into TT. According to this way of thinking simulation could be employed heuristically by a theory-processing device. This is because it is hard to see why any process should not theoretically be divisible into discreet stages. The idea is that these could function as representational states causally mediated by theoretical processing. If this is the case an argument can be made that a tacit theory is being employed by the simulation process. This chapter interrogates the notion of tacit theory. It notes a danger that

the notion of tacit may be being construed so broadly it is in danger of losing its usefulness. I suggest that even if ST has tacit theoretical underpinnings this is not the type of theory we have been discussing in chapter two. It does not employ an abstract set of folk psychological states and interconnecting laws.

I examine a suggestion by Heal that the threat of collapse can be avoided if we think of the TOM debate as an a-priori debate about how abilities at a personal level are interrelated rather than as an empirical one about how our ability to predict others' psychological states is sub-personally implemented. She looks at possible ways to construe the debate as an empirical question but argues there is no reason to assume we can expect a decisive answer.

The chapter next turns to Goldman who argues that if there is a collapse what we get is a hybrid of the two theories rather than the collapse of ST to TT. He recommends a stricter criterion of theory be employed in the debate. Employing less general criteria for theory will enable demarcation between the two theories. Not just any theoretic inference can count as evidence of TT. For example Goldman argues a theoretical rule linking individuals is not the type of rule TT employs. He also recommends distinguishing between process led and theory led formulations of ST.

The chapter next explores explicitly rather than accidentally hybrid positions that have been adopted within the debate. The issue now becomes how the two positions can interface with one another. Heal recommends a position where simulation is

supplemented by general psychological concepts and principles that are theoretical in nature. Botterill and Carruthers take the converse position that simulation supplements and enriches TT. It also enables the mind reader to cut down on potentially exponential theoretical processing. Goldman discusses different forms a hybrid theory could take. One strategy could implement the other, the two could cooperate or they could operate in autonomy from one another.

Discussion of the various forms which a hybrid theory could take raises a need to reframe the debate. Goldman's recommendation is that we construe it as a debate between ST and simulation neglecting TT. Carruthers objects that this is biased in favour of ST. His recommendation is that the central concern should be with which type of operation is most central to mind reading. He thinks making this the debate central would render the discovery of MNs unproblematic for TT, providing it assigns a secondary role to simulation. He identifies the key issue as whether first personal knowledge has a central role in mind reading.

One aim of this chapter is to show that later refinements in the TOM debate rather than settling matters have brought ambiguities with the central terminologies of theory and simulation into sharp relief. Furthermore the new question of which strategy is more fundamental is even harder to construe as an empirical question than the initial question about which strategy was best supported by empirical evidence. The issue is not whether either process can receive empirical support as empirical support for both positions is being taken for granted. However what type

of evidence one will class as fundamental will be influenced by one's philosophical sympathies. I suggest this raises the question of whether the debate can really be settled by empirical means. It is also argued that a concern with defining the positions in order to preserve mutual exclusivity is also conceptual rather than empirical.

Another issue considered is that the empirical evidence for mind reading is concerned with highly trivial cases. One thing that makes them trivial is the lack of context built into the experimental setting. We will see just how important this is in later chapters. Wittgenstein shows that in many cases the context itself is what determines meaning.

Summary of part 1

From this examination of the TOM debate this thesis will conclude that neither TT nor ST can adequately account for our grasp of psychological states. Whereas TT treats all access to psychological states as third-personal ST is unable to escape the first-person perspective. It also treats self-understanding as given. It will also be noted that much of the evidence appealed to in support of these theories was open to different interpretations. Some of this data is in fact argued to support both positions. The way in which the data is being used is also problematic. A criticism is that sub-personal data is being offered as an answer to an epistemological question about how we can know the psychological states of another rather than simply featuring in

an explanation of how this answer is facilitated at a sub-personal level. Both theories state that psychological attribution is an inferential process.

It is concluded that TT does not have a plausible notion of what is involved in having psychological states. One problem is that it views the meaning of psychological states as fixed by their place within a causal explanatory theory. This leads to the implausible consequence that we might abandon talk of psychological states in favour of a revised theoretical vocabulary. Another problem is that the focus on the functional role of a theoretical state within a folk psychological framework cannot account for the role context might have in making the state what it is.

Although ST attempts to improve on TT by allowing us to have a more direct perception of psychological states this attempt is undermined by the use the theory makes of intermediate representations of psychological states.

It is also noted that evidence offered in support of ST does not in fact support a classic simulation process. What is offered in its place is an account of how unassigned representations get attributed to self or others that is highly implausible from a phenomenological perspective. It cannot take into account the differences in the ways in which psychological states of self and other are experienced. ST is criticised for being unable to take into account the rich contexts in which psychological attributions take place and therefore unable to appreciate any role these might play in facilitating the attribution itself (in fact this flaw also applies to

much of the data employed by both TT and ST). A rejection of ST as an answer to question 1 raises a question whether the data they present for MNs can still play a naturalizing role in a more philosophically robust account of our practices of attributing psychological states to others. It is argued that data does not determine its own interpretation and a satisfying interpretation still awaits a satisfactory answer to the question of what we are doing at the personal level.

There are many ways in which a hybrid of TT and ST might be spelt out and a number of these were examined. However there is little prospect that any will be universally adopted. It is also seen that the TOM debate has developed in a way which gives rise to further ambiguities concerning the central empirical claims of each theory. The question which position is more fundamental takes the debate in a more conceptual direction rather than an empirical one.

Part 2

Chapter five introduces a new theory offered as a third alternative to TT and ST providing an account of how we can access other people's intentional states. This is Gallagher's theory of direct perception or interaction theory. In fact there are two ways of reading this account. Firstly as offering a solution to the epistemological problem about knowing other minds; as engaging in just the same enterprise as TT and ST. Secondly as a way of showing that there is no epistemological question to answer. One issue raised in this chapter is whether this theory offers a rejection of the epistemological problem or a solution to it. It will be argued that elements of both approaches are found in Gallagher's work.

Direct perception is contrasted with perception plus some further process such as simulation or mind reading. The main idea is that perception provides all the necessary resources to understand other people's psychological states. Perception utilises physical social and cultural features of the environment to contextualise gestures and expressions. It also draws on dynamic features of one's own interaction with another individual. Direct perception is a personal level account and may be underwritten by complex sub-personal processing. A feature of Gallagher's account I take issue with is that most of what he has to say about direct perception concerns perception of objects rather than psychological states. This is a problem if Gallagher is failing to account for what is distinctive about our perception of mental states.

Gallagher argues rival theories adopt a third-personal 'observer position' in relation to others when trying to decode their psychological states. He emphasises the links between one's interaction with the other and understanding. He also criticises the idea that mind reading strategies are employed to understand others. He suggests mind reading is a lot less common than these theories suggest.

This chapter also introduces Gallagher's interaction theory (IT), which attributes a central role to interacting with others in allowing us to achieve interpersonal understanding. Intentional states manifest themselves in bodily behaviour. Gallagher appeals to developmental evidence to support interaction theory. This evidence concerns our primary and secondary intersubjectivity. Primary subjectivity provides an interactive attunement to others that begins at birth. Because others are engaged through interaction their behaviour and one's own is synchronised. Gestures, expressions and actions are perceived as meaningful. It is these qualities that facilitate a direct perception of another's psychological state. One's own involvement in interactions helps to decode the meaning of others' expressive gestures. Gallagher argues this capacity remains with us and does not give way to theorizing or simulation strategies. Secondary intersubjectivity enables us to begin to co-constitute the meaning of the world with others. Gallagher claims we are involved in interactions with others from the very start. These interactions are instrumental to one's development as an agent. One reason this is seen as important is that it does not treat the agent as an isolated individual engaged in trying to understand others. It is just such treatments that encourage the use of analogical

arguments beginning from one's individualised experience as a strategy for understanding others. Although Gallagher is opposed to both TT and ST this chapter is primarily concerned with the significance of his ideas for ST. He locates ST within this analogical approach and offers criticisms of the strategy. It treats one's own states as too obvious and those of others as too obscure.

I also consider another strategy Gallagher employs to undermine sub-personal versions of ST. This is to argue that it is based on its explicit counterpart and so vulnerable to the same criticisms that can be levelled at standard accounts of ST. He shows how ST tries to interpret neural activations involving mirror neurons as exhibiting a multistage inferential process modelled on explicit ST. He argues that the reason sub-personal versions of ST feel obliged to adopt this approach is because they share a way of framing the task of understanding others as that of accessing hidden inner states on the basis of observable behaviour. One reason for this is because perception is treated as an overly passive process rather than the interactive one suggested by Gallagher's own position. While I take a sympathetic view of Gallagher I also question the coherence of any attempt to model sub-personal processes on personal ones. This point mirrors a point made in my discussion of ST.

Gallagher also offers a rival interpretation of MN activity. He claims MNs can offer empirical support for DP. We can agree with Gallagher up to a point. There is nothing to block the possibility that MNs facilitate a direct perception of others. I agree that our perception is direct. However Gallagher appears to be suggesting a

relationship between personal level accounts and sub-personal level ones which would allow sub-personal discoveries to dictate the account we give of personal level practices, and this is something I would call into question.

My discussion of Gallagher concludes by evaluating the case for and against reading interaction theory as involved in the same kind of epistemological enterprise as TOM; as providing an answer to question 4 rather than rejecting the framework in which the question can be raised. I conclude there are reasons for thinking Gallagher's account of direct perception is in the business of providing resources for answering the epistemological question. However this is not to say his position does not have many advantages over TOM approaches. These positive features of Gallagher's work anticipate aspects of phenomenological and Wittgensteinian approaches that I will explore next.

The next section of the thesis focuses on phenomenological accounts of our understanding of others psychological states provided by Merleau-Ponty. I begin by looking at his critique of the classical model of understanding others. There are parallels between the classical model and TOM approaches so Merleau-Ponty's objections will have consequences for these approaches. Merleau-Ponty offers an objection to intellectualised accounts of understanding others; understanding is not based on finding a common meaning in one's own experience and that of another. Merleau-Ponty also offers reasons to reject the idea that psychological phenomena could be inferred from physiological data. Merleau-Ponty criticises scientific

approaches for entrenching a theoretical distinction between a subject and object. He argues this encourages the use of analogical strategies and it does not allow psychological subjects to be encountered as such. It also leaves no room for the idea that consciousnesses are plural as consciousness is something residing in an individual. Merleau-Ponty opposes individualistic conceptions of the psyche. He also opposes the idea that psyches are incommunicable and only accessible indirectly by others.

Merleau-Ponty fleshes out the consequences of this framework. We could not justify attribution of psychological states to others. One reason is that this framework has reduced our own psyche to a mass of sensations. This is a highly individualistic way of conceiving things. It leaves us without a common medium of representation. We then face an implausibly intellectualised task of learning to associate our inner bodily sensations with the visual experiences we have of others. Merleau-Ponty points out that we are able to recognise psychological states in others before we have developed sufficiently to be capable of performing this kind of translation. Drawing analogies with others presupposes a common medium of representation to draw them in but we do not have one. This raises a problem for the strategies employed by TOM positions. Merleau-Ponty also argues against ST by resisting the need for a comparison with oneself.

Next the chapter explores Merleau-Ponty's account of our direct perception of the psychological states of others. For him a psyche is already a relation to the world

and this is already a world populated by other psyches. This means it can experience

the things it is directed towards, including other people, in a meaningful way.

Merleau-Ponty offers an account of our development in which we first of all

encounter conducts or actions, be they our own or those of others. The intentional

level is operative even before we are able to perceive people as individualized

selves. The level of description at which we decode people's intentions is the

personal one. On this picture, unlike on the classical model, subjects are identified

holistically. We do not require a story where we link motor and visual sensations on

a point by point basis. This makes it easier to appreciate conducts are transferable

across agents. Expressive actions resonate with one's own bodily schema in a way

that mere visual stimuli could not.

Merleau-Ponty characterises our experience of the body as a corporeal schema. This

is an experience of behaviour with two dimensions; one's own behaviour and that of

others. This functions as a whole. Understanding of self is facilitated by one's

interactions with other psyches. Simultaneously the actions of others are understood

because they offer themselves to one's own motor schema.

Expression is assigned a key role in understanding others in Merleau-Ponty's

phenomenology. Psychological states are manifested in expressions not

subsequently attributed to them. Gestures are aimed at and executed in the world.

Expressive features are bodily features and therefore the body is also extremely

important. It is through this that we can perceive the bodies of others. These appear

as familiar modes of dealing with the world. They open us up to a social world by placing us in a particular situation within it.

Like Gallagher's account, Merleau-Ponty's account of understanding psychological states is led by personal level description. Behaviour needs to be understood at the level of cultural factors rather than biological ones. There is no need for analogical reasoning. In fact before we can acquire the ability to make comparisons between others' expressions and our own we already have to perceive others as psychological agents.

A feature of Merleau-Ponty's account emphasized here is that he does not assign an epistemic privilege to the first person. This distinguishes his approach from the one adopted by ST. We start out without an awareness of ourselves and others as separate. Merleau-Ponty gives a phenomenological account of experience in which experience of one's own psychological states is intertwined with experience of others. He offers an account of how this separation from others is eventually accomplished. One key stage in this accomplishment is recognition that one's own mirror image is also available to others. Merleau-Ponty also argues the separation is never actually completed and is transcended in adult sympathy.

Merleau-Ponty shows that awareness of others enters the picture prior to the possibility of drawing inferences between sensual and visual data. Our initial contacts with others come via introceptive rather than visual channels. Merleau-

Ponty argues against an oversimplified picture in which our sense organs provide mere sensory data. Other people feature not just as part of the contents of our experience; they also structure the content of experience itself. Merleau-Ponty appeals to the ability to imitate others' expressions and conducts which he refers to as mimesis. He argues people possess this ability in virtue of the power they have to control their own body. He argues that without this ability perception would not be possible. Merleau-Ponty also discusses a phenomenon called transitivism which denotes an initial inability to distinguish self and other. To begin with there is no absolute distinction between self and other. He argues the conception of a separate self which one eventually arrives at is an abstraction from this initial mode of being. In Merleau-Ponty's framework there is no room for the possibility that first person experience can be privileged in giving an account of understanding. To have a world in the first place is to share it with others.

The final chapter examines Wittgenstein's ideas about understanding others, which parallel those of Merleau-Ponty. Wittgenstein's ideas are put to work in two ways. Firstly they are employed to show that if we start off with an account of psychological states as inner we will not be able to reach a satisfactory account of our ability to attribute psychological states to others. These remarks have implications for ST. The chapter begins by fleshing out what is involved in the inner outer picture Wittgenstein is undermining. If psychological states are classified introspectively it is very difficult to give them an external application. Furthermore this conception of inner states is internally linked to an unsatisfactory picture of the

physical bodies of others which is divorced from the psychological.

Wittgenstein shows that introspectively classified mental states will not be transferable to other people. As well as offering a challenge to the idea that psychological states are in fact anchored in first personal experience Wittgenstein also provides ammunition against the idea such states could be utilised in a solution to the epistemological problem of other minds which is question 4. A major failing of this conception of psychological states is that it does not take into account what is distinctive about other agents and our mode of access to them. On this picture there is no intrinsic connection between psychological states and living bodies. Wittgenstein brings out the limitations of relying on one's own experience as a model for the experiences of others. At best it will give us an idea of having our experiences in their bodies. Furthermore he shows that such an extension would not even be coherent. There is nothing special about other bodies on this picture which would make them suitable recipients of such psychological attributions.

Another aspect of Wittgenstein's resistance to any attempt to anchor the meaning of psychological terms in first personal experience concerns the notion of a private sensation. If by private sensation one means known only to the individual Wittgenstein provides two types of objection. Firstly it is wrong to claim other people cannot be said to know about one's sensation. Secondly it is incoherent to claim one knows about one's own sensations.

These arguments have a clear application to standard ST. They bring out the

incoherence of assigning a particular role to the imagination in attributing

psychological states to others. Next the chapter focuses on the application of

Wittgenstein's arguments to sub-personal versions of ST. It argues that an analogue

of Wittgenstein's objections still applies even though no role is being assigned to the

imagination by this theory. One reason why is because the theory interprets the

neural activity as having a representational format modelled on the imaginative

process found in standard simulation. I also argue that if sub-personal versions of

ST are in the business of offering a solution to the epistemological problem of other

minds then they fall foul of Wittgenstein's objections to analogical arguments. As

well as being an exceedingly weak analogy it does not provide us with what we

need, it merely allows us to extend our own inner states to other people.

This chapter next considers some reasons why it might be unfair to claim sub-

personal ST is operating within the framework of the inner that Wittgenstein has

criticised. In particular I consider reasons for thinking sub-personal versions of ST

do not conceive of psychological states as exclusively first personal and that our

access to other people is not limited to mere physical bodies. I argue that

nonetheless interpreting neural activity as having a representational format leaves

these versions of ST open to the difficulty of how to justifiably apply them to others.

Sub-personal versions of ST claim to operate on shared representations of

psychological states rather than exclusively first personal representations. They are

couched in a developmental story reminiscent of the one Merleau-Ponty gives in

which we do not initially distinguish between states of ourselves and of others. This chapter shows that closer interrogation of the notion of shared representation appealed to by ST does not stand up. Neuronal activation would more appropriately be described as agent neutral. Sub-personal ST ignores a strong grammatical connection between sensation and an experiencing subject that Wittgenstein draws attention to. A problem is that ST attempts to anchor the meaning of these states in sub-personal activity which brings it into conflict with Wittgenstein who has argued that meaning must be anchored in public practices.

Another partial defence of sub-personal ST considered is that, rather than suggesting we confront mere bodies, sub-personal ST suggests we automatically respond to perceived intentional qualities. Other people are suitable subjects of psychological attribution. However, while we have a picture which explains how psychological representations might causally be extended to others, ST has nothing to say about how such extensions could be justified. This picture does not make allowances for the intrinsic connection between a psychological state and an experiencing subject.

The second half of this chapter puts Wittgenstein's ideas to a different use; to highlight some oversimplifications in Gallagher's alternative direct perception account in order that a more nuanced account of direct perception can be formulated. It begins by fleshing out his alternative to the model of the inner in which he brings together physical and psychological states as different aspects of one phenomenon. Wittgenstein shows that psychological expressions are not inferred on the basis of

physical characteristics but instead permeate the psychological subject one confronts. Like Gallagher Wittgenstein argues our perceptual capacities are not focused on the most basic physiological level and that they are sensitive to intentional patterns. However, whereas Gallagher tends to treat perception of psychological states as on a par with perception of objects, Wittgenstein emphasises important differences. Perception of psychological states is significantly different from perception of physical objects. Psychological states are tied up with shared practices. These practices do not facilitate an epistemological solution to the other minds problem. Rather our attitudes towards others embedded in these practices are themselves fundamental and belong to the practices themselves. Wittgenstein also draws attention to the normative character of these practices. The role a given psychological state has in these practices is necessary for its identification as a psychological state. This differentiates one's practices involving other agents from those involving objects. Wittgenstein also argues we are not able to perceive people as mere objects during our interactions with them.

The philosophers examined in part two provide further material to develop criticisms made of the TOM debate in part 1. Gallagher is particularly concerned with ST. He diagnoses a tendency to construe the mind reader as an individual involved in an attempt to understand others. This encourages the use of analogical arguments which Gallagher argues take self understanding for granted and treat the task of understanding others as too difficult. They also involve an impoverished notion of perception, which is construed as a passive process.

However dissatisfactions will also be raised with Gallagher's own philosophical standpoint which is not immune to some of the criticisms this thesis levels at TOM. In particular it is argued to be involved in an attempt to answer an epistemological problem of other minds despite the fact that elements of his philosophy can also facilitate in dissolution of the problem. His account also treats perception of psychological states analogously with object perception which does not allow the distinctive character of the former to emerge.

Merleau-Ponty's critique of intellectualised accounts of understanding others as involved in a search for a hidden meaning has direct implications for TOM. Successful analogies with other agents require a common medium of representation but Merleau-Ponty argues that one will not be found. He provides further ammunition against ST by denying we have epistemically privileged access to first personal states.

Wittgenstein presents arguments in a similar vein that also have implications for ST. He argues that we will never be able to construct a coherent account of our ability to attribute psychological states to others if we begin with a conception of psychological states as inner because such states cannot be given an external application. He also argues this conception of inner states goes hand in hand with a conception of the bodies of others which is entirely unsuited to psychological attribution because it does not include their distinctive characteristics which facilitate an accompanying range of practices. He further argues that the idea of a

private inner sensation is not in any case coherent. It is both incorrect to claim

knowledge of one's own sensations and wrong to claim other people have no access

to them. Wittgenstein's argument that the imagination will not be capable of

fulfilling a key role in attributing psychological states has a clear application to

standard ST. This thesis argues an analogue of this argument also has an application

to sub-personal versions of ST, which bestow sub-personal mechanisms with a

representational format modelled on the imagination.

Wittgenstein is also used to bring out what is problematic about the notion of shared

representations found in sub-personal versions of ST. It ignores the grammatical

connection between a sensation and an experiencing subject. It also attempts to

anchor the meaning of a psychological state in sub-personal states rather than in

public practices.

The philosophers examined in part two also help construct a more philosophically

satisfying account of how psychological states get their meaning.

To fully appreciate the direct nature of intersubjective perception requires scrutiny

of our psychological practices. This will reveal atypical features of intersubjective

perception (such as a sensitivity to normative characteristics) and undermine an

assumption that it will be continuous with perception in general including perception

of objects.

The attempts by TOM to solve an epistemological problem are undermined by

showing that we are engaging with other as psychological agents before such questions could arise. An understanding of oneself already presupposes a world shared with others. For Merleau-Ponty psyches are relations to the world and for this reason experience the things they encounter as meaningful. Merleau-Ponty argues self understanding is itself facilitated by interaction with other agents. They not only figure as contents of experience (as in the other minds picture) but also as part of the fabric of experience.

Whereas TOM was criticised for adopting a detached attitude to others the philosophers examined in part two emphasise the interactive nature of intersubjective perception. This also explains how context is built in. Both Gallagher, and particularly Merleau-Ponty, emphasise the bodily nature of our intersubjective encounters. This contrasts to the intellectualised process depicted in personal level accounts of TT and ST. It is the expressive qualities of bodies that manifests psychological states and it is through one's own body that one perceives the bodies of others. Bodies are presented to us as elements in the social world and as familiar modes of inhabiting it.

Wittgenstein fleshes out an alternative to the model of the inner in which physical and psychological states are united. Psychological characteristics permeate the physical body. Rather than picking out physiological features of bodies from which we must infer psychological ones we are sensitive to intentional qualities. As well as being in opposition to the model operative in TOM this model can also

sophisticate Gallagher's position in that Wittgenstein brings out important

differences between perception of psychological states and of objects. Psychological

states are bound up with characteristic practices and it is the attitudes embedded in

these practices which constitutes the crucial difference. One feature of this is that

these practices have a normative dimension.

2 The Theory Theory of mind

This chapter will review progress and development in TT. It will also bring out diversity of ways in which the term 'theory' is employed. It will consider evidence for the acquisition of these alleged theoretical capacities for understanding others. After offering an exposition of the main forms of TT it will evaluate this theory in relation to aim B. It will argue that the theory offers answers to some of the questions about how one can know other people have minds but does not address other questions I raised. I will make a case that it offers an answer to Question 1; what are we doing when we attribute mental states to others? and to question 3; how is attribution of mental states to others achieved? but has little to say regarding questions 2 or 4. It will then consider what can be done in regards to the further aims of this thesis. In particular it will point out instances where a picture of the inner is being adopted in which psychological states are conceived as unobservable inner states. It will attempt to disentangle the philosophical and empirical moves made by proponents of this theory.

Origins of TT

The term theory of mind (TOM) was introduced in 1978 in a paper by Premack and Woodruff. TOM gives a certain kind of account of our ability to ascribe mental states to ourselves and to others. A person has a theory if he attributes mental states

to himself and others by employing a system of inferences.[20] Premack and Woodruff argued that this should be understood as employing a theory because "such states are not directly observable, and the system can be used to make predictions about the behaviour of others".[21]

Premack and Woodruff were primatologists who were interested in whether chimpanzees possessed a TOM. They were conducting research to see whether chimpanzees were able to attribute states of mind including intention, purpose, knowledge and belief to a human actor.[22] However the term TOM is now found in a number of other disciplines including philosophy, cognitive and developmental psychology and neuroscience.

TOM is often used synonymously with the term 'Theory Theory' (TT) which is itself one possible explanation of how mental states are ascribed to others. The term TT was probably first introduced in 1980 by Adam Morton to classify a converging trend in philosophy and psychology.[23] Both disciplines were aiming to construct plausible accounts of how we acquire knowledge of one another's motives, beliefs etc.[24]

---

[20] Premack and Woodruff 1978 P.515
[21] Ibid
[22] Ibid P.518
[23] See Stich & Nichols 1998 P.422
[24] Morton 1980 P.8

<u>Morton's version of TT</u>

Although responsible for first introducing the term, Morton was not arguing in favour of TT but sketching out what he saw as the main competitor to his own 'scheme theory'. According to Morton's conception of TT, when trying to anticipate the actions and reactions of others, and to understand the reasons for their behaviour, people are making use of a common stock of beliefs which take the form of an implicit theory.[25] Morton interpreted recent work in both philosophy and psychology as attempting to elucidate the concepts and principles belonging to such a theory.

The philosophical trend begins by emphasising the dependence of particular instances of psychological ascription on something more general.[26] One learns to recognise particular patterns of behaviour as standard signs of different kinds of mental state. Next one learns to view one's own states in the same terms. One then learns to ascribe states to oneself based on one's own behaviour. These skills are acquired as a unity.[27] The attribution of an individual state is not independent of other states and it is only large combinations of states that allow us to correlate patterns of behaviour.[28]

TT embodies a move in philosophy away from a behaviourist approach to understanding others towards a functionalist one:

---

[25] Ibid P.7
[26] Ibid P.9
[27] Ibid P.9
[28] Ibid P.10

49

"One standard philosophical account now is that psychological terms are understood as occurring in certain typical roles in the production of behaviour, and that, added to this one has a large number of beliefs about how combinations of states result in action. In short there's a theory of mind that we all have".[29]

One question this raises is how we acquire this theoretical knowledge about mental states and how to attribute them. This is something that we will see theory theorists disagree about. According to Morton's conception of TT we each begin to learn this theory very early during the course of socialization.[30] He claims the evidence for this is varied and includes the frequency with which generalisations about actions and motives hold. One observes how one's elders attribute psychological states to others and this can serve as correctional to one's own routines for attributing states to oneself and to others, which Morton thinks can be idiosyncratic.[31]

In psychology the journey to TT also stems from an interest in the role 'inference to best explanation' plays in attributions of psychological states.[32] Originally researchers had searched for characteristics of static facial expressions that could be correlated with various emotions. Subsequent research explored how longer patterns including available background information and extended periods of observation of

---

[29] Ibid P.10
[30] Ibid P.8
[31] Ibid P.11
[32] Ibid

the target influence the attribution. In the same vein attention started to be paid to the interactional dimension of attribution.[33]

The conclusion for both philosophers and psychologists is "that there is little that is psychologically unique in the attribution of states of mind".[34] The relations that hold between mental states always conform to characteristic patterns. The focus of interest therefore turns to the nature of this structure. The standard view which Morton terms TT is that they are structured as:

 "an 'implicit theory of personality'. That is, as a body of beliefs, most of which everyone has, concerning the relations between different attributes. The dominant opinion is that the core of the theory is a set of roughly probabilistic correlations expressing the relative likelihood of a person satisfying one trait satisfying another".[35]

Theory

Morton's construal of theory was relatively broad compared to subsequent modifications. However he did have some stipulations. He defined a theory as:

---

[33] Ibid
[34] Ibid P.12
[35] Ibid P.11

"a body of assertions whose terms refer to individuals and properties, and which is transmitted and evolves in accordance with the intention that it assert the truth about them".[36]

This body of assertions can be used to explain various things and predict various other things because it possesses a certain homogeneity.[37] It can be used to make determinate claims about certain phenomena. A theory consists of conjectures about the best explanations for these phenomena. These explanations are accepted on the basis that no better explanations or refuting evidence has arisen.[38]

A theory does not require a precise formulation. To subscribe to a theory does not require an understanding of everything the theory commits one to.[39] However subscribing to a theory entails accepting that its terms are intended to refer to objective realities and that modifications to the theory are made with the intention of increasing the ratio of true assertions about reality over false ones.[40] To subscribe to a theory only requires explicit knowledge of a few of the theory's assertions. However doing so puts one in touch with a large number of facts about things as one can look to fellow subscribers as repositories of assertions one cannot produce for oneself.[41]

---

[36] Ibid P.5
[37] Ibid P.4
[38] Ibid
[39] Ibid P.5
[40] Ibid
[41] Ibid P.6

52

Many of the characteristics Morton ascribed to TT have been adopted by its proponents and it is worth briefly summarising these characteristic features. The primary purpose of the theory is for explaining and predicting behaviour (the behaviour of others and, more controversially, of oneself). Psychological episodes are particular instances of general patterns. Psychological concepts only have meaning within a theory. Morton also raises the issue of how a theory is acquired which continues to be a central concern. Theoretical reasoning involves an inferential process to best explanation and in this sense our access to the phenomena it attempts to account for is indirect.

Folk Psychology; Eliminativism and Functionalism

TT was heavily influenced by the idea that our everyday understanding of the psychological constitutes a rough and ready body of knowledge commonly referred to as folk psychology. The paradigmatic terms of folk psychology are belief and desire. These are often referred to as propositional attitudes. It is a folk theory that we use to explain and predict the behaviour of other people. The term 'folk theory' is used to contrast this kind of understanding with formal scientific knowledge. If our propositional attitudes constitute a theory, an important consequence is that they can be true or false; we could turn out to be mistaken about folk psychology in the sense in which other earlier theories such as folk physics turned out to be mistaken. We may discover that states such as belief and desire which are posited by the theory do not actually exist. A new theory may instead postulate a very different set

of entities to explain human behaviour.[42] The fact folk psychology is meant to be a 'folk' theory makes it particularly vulnerable to the possibility of replacement by a more rigorous scientific contender.

Eliminative materialism is the position that we should in fact cease to employ the vocabulary of folk psychology in our explanations of one another. Churchland suggests that although it has been some use in enabling us to explain a limited domain we have not been able to extend it into new areas. Another reason that critics have suggested for thinking we in fact ought to abandon folk psychological explanation is that folk psychological entities do not smoothly reduce to neurophysiological laws. They are ontologically discontinuous with the rest of physical nature.[43]

TT was influenced by the ideas of David Lewis. Lewis argued that the meaning of theoretical terms can be defined functionally by reference to their causal roles.[44] In the case of psychological terms, a psychological term can be defined as the occupant of a certain causal role which is causally connected to sensory stimuli, motor responses, and other psychological states in specifiable ways.[45] Theoretical terms are introduced into vocabulary by the theory they belong to.[46.] A theory assigns its terms an implicit functional definition. The terms name the occupant of a specified causal role.

---

[42] See e.g. Churchland 1981
[43] Greenwood 1991 PP.5-6
[44] Lewis 1972 P.249
[45] Ibid P.250
[46] Ibid

54

Theoretical or 'T' terms are contrasted with all the other terms of a language which Lewis refers to as 'O' terms.[47] O terms are already understood by native speakers prior to the introduction of the theory. The theory is initially presented in a sentence called the postulate of T. This postulate states that the entities postulated by T terms stand in certain specified relations (including causal relations) both to other T terms and to O terms. One can see why psychological terms are being thought of as theoretical. T terms have been defined as the occupants of particular causal roles specified by the theory.[48] Further they have been specified as entities that stand in specific relations to other T terms and O terms.

Lewis suggests we think of our common sense psychology as a primitive scientific theory. In order to form this theory one must collect all the platitudes concerning all the causal relations that hold between mental states, sensory stimuli, and motor responses.[49] Next we must form a conjunction of these platitudes. This will be the postulate of the term-introducing theory. The names of psychological states will be the T terms.[50] These are defined by their relation to their causal relation to stimuli, responses and to each other. We learn what mental states are when we learn what states occupy these causal roles. Such states must uniquely fill these slots.[51] Lewis argued that if psychological state terms are theoretical terms then they will not name anything unless the cluster of platitudes which constitute the theory is more or less

---

[47] Ibid P.250
[48] Ibid P.254
[49] Ibid P.256
[50] Ibid
[51] Ibid P.251

55

true. The meaningfulness of psychological terms stands or falls together.[52] Their

meaning is not given independently of the theory. Particular psychological terms

could be replaced by formal variables (such as X, Y, Z) as long as we preserve the

functional relations that hold between these Terms and other O terms. This places

his thinking close to the instrumentalist view of the meaning of theoretical terms.

According to instrumentalism a body of theoretical terminology should be assessed

in term of its effectiveness in explanation and prediction of the phenomena in

question. Instrumentalist accounts are contrasted with realist accounts of theoretical

terms which claim that these terms pick out things in the objective world that exist

independently of the theory that captures them.

Lewis's account of theory closely resembles the ideas of TT advocates Stich and

Nichols. Stich and Nichols see TT working in line with what they term the dominant

explanatory strategy in cognitive science.[53] This is to posit an internally represented

knowledge structure, such as a body of rules, principles or propositions which have

the function of guiding the capacity that science wishes to explain. Stich and Nichols

suggest that this body of rules can usefully be termed the agent's theory of the

domain. They argue such a theory can be partly accessible to consciousness but that

in most cases the agent will have no conscious access to the body of rules that guides

his behaviour.

---

[52] Ibid P.258
[53] Stich and Nichols 1992 P.123

Again we observe a number of key characteristics of TT. Psychological concepts only have meaning within a theory which is being employed to explain and predict behaviour. More specifically, Lewis suggests psychological laws are defined by the functional roles they occupy within a theory. Folk psychology is envisioned to be a rough and ready scientific theory. The theory is largely implicit.

Development of TT

This thesis will now explore ways in which versions of TT diverge from one another. Proponents of TT have modified and developed the theory in a number of different directions. The term theory has been used to designate something that may be innate and modularised, learned individually or acquired through a process of enculturation.[54] Since Morton the term theory has been used by proponents of TT both in a very general sense[55] and in a very strict sense corresponding to the structure of a scientific theory[56].

TT has also been developed simultaneously for a number of different purposes. Although the overall research question is how we are able to ascribe states to ourselves and to others this question is closely bound to other questions. Such questions include; how do we come to develop this ability? And how do we account for the deficit found in people unable to ascribe mental states to others?

---

[54] Carruthers 1996 P.1
[55] See Stich & Nichols 1992
[56] See Gopnik 1996

A recent work by Goldman divides TT into two subgroups; the child scientist theory and the modularity theory.[57] The child-scientist view implies that we have to learn our theory of mind whereas modular theory often presupposes that this knowledge is innate. There are many further distinctions that could be drawn such as between learning by theorising and by enculturation. Furthermore it is hard to fit much work presented under the banner of TT squarely in either camp. Some theory theorists claim theories are modular but not innate or vice versa. It is often difficult to find a proponent of TT who fits clearly in one categorization whereas proponents who hold a hybrid of both views are common. Nonetheless this distinction will be employed here as it provides a useful handle on the debate.

The child scientist theory

According to child scientist versions of TT the theory is learned on the basis of experience.[58] Carruthers points out this version of TT itself divides into two versions:

"In one version the child is pictured as a little scientist, constructing and revising theories in the light of incoming data. In another version the child is seen more like a scientist student, picking up the FP of its culture through interacting with and listening to the talk of its careers and older siblings".[59]

---

[57] Goldman 2006
[58] Carruthers 1996 P.4
[59] Ibid P.4

Gopnik is one of the clearest and most important proponents of the child scientist view. She argues that understanding what the comparison between children and scientists entails requires a grasp of how science is supposed to proceed. Gopnik illustrates the features of scientific procedure she takes to be relevant. However she does not presuppose a comprehensive understanding of development in science is required in order to make a comparison between scientific endeavour and the way children learn to grasp mental states. Gopnik argues that understanding how children acquire a TOM can itself teach us something about theory acquisition in science:

"The moral of my story is not that children are little scientists but that scientists are big children. Scientists and children both employ the same particularly powerful and flexible set of cognitive devices. These devices enable scientists and children to develop genuinely new knowledge of the world around them".[60]

Comparison between children and scientists

Gopnik argues cognitive development in children and development in science are both underwritten by an abstract set of representations and of rules that operate on these representations. The capacity to create and manipulate these representations is an evolutionary one designed to allow children to gain a veridical view of the surrounding world.[61] Scientists continue to employ the same capacities:

---

[60] Gopnik 1996 P.486
[61] Ibid P.489

"the core similarity that we capture in the scientific analogy is a similarity in the rules and representations that allow scientists and children to make cognitive progress".[62]

Scientific progress and normal cognitive development in children involve regular causal relations between input and representations.[63] We also possess cognitive processes that transform both the representations and the rules over time.[64] These representations and rules are abstract structures that can combine logical and psychological characteristics.[65]

Gopnik acknowledges that there are obvious differences between theory formation in children and scientists. However she makes a case for thinking these differences are superficial and can be accounted for within a TT framework. One difference is that scientists self-consciously reflect about theory-forming and testing whereas children do not explicitly engage in theoretical discussion. Gopnik's response to this potential counterevidence to her position is twofold. Firstly she suggests that more of scientific endeavour is inaccessible to conscious reflection than one might think.[66] Second she argues that a child's theoretical investigation is not as isolated as the above image can seem to imply. The child has a social structure that provides "contradiction, instruction, and the linguistic transmission of information".[67]

---

[62] Ibid P.490
[63] Ibid P.501
[64] Ibid P.487
[65] ibid
[66] Ibid P.491
[67] Ibid

Another difference is that science occurs in an institutional setting. It involves formal interaction with other scientists and a division of labour whereas conceptual change in children occurs within an individual child. Gopnik argues these structural differences stem from the different types of problem been tackled by science rather than signifying a deeper difference in its approach to solving problems. Whereas the evidence necessary to tackle the child's problems is usually easily available, science is applied to highly specialised problems which call for a division of labour.[68] Transition between theories in science may take years whereas the speed of progress is much quicker in infants. Once again this is argued to be due to differences in the nature of the problems rather than in the methodology employed.

A final striking difference that Gopnik tackles concerns the idea of convergence. Children usually converge on the same theories at approximately the same age[69] whereas this is not something we observe in scientific practice.[70] Gopnik argues that this pattern in the child's cognitive development in fact provides the best reason to suspect that the same general cognitive structures are actually being employed by both scientists and children.[71] It is an assumption of TT that, all things being equal, we would expect to find the same pattern in scientific development: "theory theory proposes that there are powerful cognitive processes that revise existing theories in response to evidence. If cognitive agents began with the same initial theory, tried to solve the same problems, and were presented with similar patterns of evidence over

[68] Ibid P.492
[69] Ibid P.487
[70] Ibid P.494
[71] Ibid

the same period of time they should, precisely, converge on the same theories at about the same time".[72]

Although it is much less apparent in science than in developing children, convergence is in fact a feature of scientific theory change. It is less apparent because scientists, unlike children do not approach their problems from equivalent starting points. However "when the assumption of common initial theories and common patterns of evidence, presented in the same sequence, does hold, scientists, like children, do converge on a common account of the world".[73]

A general assumption is that the answer to question 1 will be a scientific answer and that the developmental evidence provides an example of the scientific method rather than posing question 1 without this assumption.

<u>What are theories for Gopnik?</u>

Gopnik is sympathetic with Churchland's suggestion that our everyday understanding of the mind is analogous to a scientific theory and thinks this idea receives empirical support from studies of child development.[74] Before briefly reviewing this evidence it is necessary to understand what Gopnik understands by a theory. She has a much more rigid view of what is meant by theory than Morton did. Gopnik is as opposed to modular accounts of mind (which I will examine next) as

---

[72] Ibid
[73] Ibid
[74] Ibid P.505

62

she is to simulation accounts. Theories, as she defines them, behave differently to modules.

Gopnik wants her definition of theory to be as general and uncontroversial as possible. She focuses on those features of theory that she takes to be generally accepted by everyone.[75] She begins by detailing the structural features of a theory. Theories are "systems of abstract entities and laws that are related to one another in coherent ways".[76] Theoretical concepts are abstract in that they are apprehended in a vocabulary that is different from the vocabulary typically used to describe the observed phenomena. They make reference to a set of entities underlying the phenomena one seeks to explain.[77] Theoretical concepts stand in law governed causal relations both to other concepts in the theory and to the evidence. The causal structure that underlies a theory is usually thought to explain the superficial regularities we observe. Gopnik does not specify what this underlying vocabulary is but presumably has in mind something like a language of thought.[78] This would consist in a system of mental representations physically instantiated in the brain. Causal operations only act on syntactic features of representations. They are not sensitive to semantic features.[79]

Theories also entail ontological commitments about what we observe and how it will behave. These commitments will encompass counterfactual situations. They can be

---

[75] Ibid P.495
[76] Ibid P.496
[77] Ibid
[78] Bishop & Downes (2002) P.124
[79] Aydede 2010

63

seen in the way we respond to apparent violations to the theory. Normally violations will initially be greeted with disbelief.

Next Gopnik sketches the functional features of theories which follow from these structural features. Theories allow us to make predictions about new evidence. What makes theories more than mere empirical generalisations is that they entail predictions about a wide variety of evidence including evidence that had no influence on the formulation of the theory and this may be their most valuable feature.[80] Theories also allow us to interpret evidence. They even exert an influence over which pieces of evidence are considered to be salient. A final functional feature of theories is that they enable us to explain evidence. They have an explanatory force which distinguishes them from mere generalisations.

Finally Gopnik sketches the dynamic features of theories. Characteristic intermediate processes facilitate the transition from a theory to its successor.[81] We saw earlier that the initial reaction in the face of counter evidence to one's chosen theory will be disbelief or denial. The next stage will involve the introduction of ad-hoc supplementary hypotheses intended to account for the counter evidence within the scaffolding of the original theory. Over time we come to see that such hypotheses undermine the coherence of the initial theory. The next stage is to formulate an alternative theory for explaining the evidence. This is followed by a

---

[80] Gopnik 1996 P.496
[81] Ibid P.498

period of intense experimentation and observation. Gradually the evidence will causally bring about a shift to the new theory.

The above features of theories have been described by Gopnik using the vocabulary of philosophy of science. She also describes theories in the vocabulary of cognitive science; a theory is "a particular kind of system that assigns representations to inputs".[82] These representations will have distinctive structural features owing to "the specific abstract, coherent, causal, ontologically-committed, counterfactual supporting entities and laws of the theory".[83] These representations are operated on by rules resulting in new representations. This generates predictions.[84]

Theories can not only predict data but also explain and interpret it due to distinctive functional relations holding between the representations themselves and the input.[85] The theoretical interpretation assigned to an input interacts with other representations belonging to the theory in particular rule governed ways.

The patterns of representations that occur may even alter the nature of the representational system itself. This is because it can affect the nature of the relations between input and representation. As new inputs can result in new representations, the rules that connect input with representation themselves change. Eventually this

---

[82] Ibid P.499
[83] Ibid
[84] Ibid P.500
[85] ibid

65

can result in a new set of representations and relations between representations and inputs.[86]

Empirical evidence

Gopnik claims there is substantial empirical evidence to support the claim that children possess theoretical structures like those described above. Crucially their understanding of others appeals to abstract theoretical entities and laws related to one another in coherent ways. They postulate mental entities such as perceptions, beliefs and desires, and psychological laws in order to explain human action.[87]

As well as these structural features of theories children also share the functional features. Four year old children are able to make consistent and largely correct predictions about a variety of new events including events quite different from any they have previously experienced.[88] Children also use their understanding to interpret action in terms of underlying psychological entities.[89] If provided with descriptions of human behaviour in neutral terminology they automatically interpret and describe it in terms of psychological predicates. Children justify their predictions by appealing to causal explanations. The explanations appeal to an underlying framework of mental entities and psychological laws, and they relate different mental entities to one another in a coherent way.

[86] Ibid P.501
[87] Ibid PP.505-506
[88] ibid P.505
[89] Ibid P.506

66

Children's understanding of the mind also appears to go through the dynamic stages associated with theory change. Gopnik argues evidence that four year old children's understanding of the mind is constructed from an earlier theory can be found by examining younger children. Three year olds also consistently predict interpret and explain the behaviour and mental states of others. However their predictions, interpretations and explanations are quite different suggesting they are the product of an earlier theory.[90] The transition between these theories has also been demonstrated to bear some resemblance to scientific theory change. They first apply the new theory only in specific cases where they are faced with counterevidence to their current theory. They also appear to undergo a period of intense observation and experimentation.[91] Gopnik also appeals to tentative evidence that the transition between theories results from the accumulation of evidence and counterevidence. This succession of theories is argued to go "all the way down" beginning at birth.[92]

In summary, Gopnik argues for a strong resemblance between the development of a child's theory of mind and the development of mature scientific theories in terms of both structural and functional features. Both are in the business of constructing inferential links between perceived cause and effect. Both evolve and change over time. In fact TT would be better characterised as a succession of theories. Both employ abstract rules and representations. Gopnik stresses that theoretical concepts are abstract; formulated in a vocabulary different to the one usually used to describe the phenomena.

---

[90] Ibid PP.506-507
[91] Ibid P.508
[92] Ibid P.510

67

Modular TT

Other advocates of TT such as Carruthers are able to agree with Gopnik on certain central issues. Our understanding of psychological notions such as belief, desire, perception and intention is primarily given by the place those notions occupy within a folk-psychological theory of the structure and functioning of the mind.[93] To understand a folk psychological notion requires some implicit grasp of how it is related to other entities and laws within the wider framework. Carruthers is also in agreement with the view that children's developing competence employing psychological terms results from progressing through a series of increasingly sophisticated theories.

However Carruthers does not take Gopnik's position that it is 'theories all the way down'. He argues that the core of one's TT is innate, rather than acquired through a process of theorising, or learning.[94.] The different psychological theories that young children progress through should be understood as different stages in the maturation of a TOM faculty. This innate mechanism probably results from a process of ontogenetic development which may also require certain external experiences to trigger its development.[95]

Carruthers argues this modular version of TT has more plausibility than the child scientist version. One reason why is because it is hard to believe child scientists

[93] Carruthers 1996b P.22
[94] Ibid
[95] Ibid P.23

should all converge on the same beliefs at the same time. On the other hand,

Carruthers argues, it is entirely plausible this understanding should be innate given

its vital role in facilitating communication and social co-operation.[96] By way of

providing empirical backing Carruthers suggests that this idea also fits well with

what has been learned about the development of social competence in apes and with

what is known about the absence of mentalistic abilities in people with autism.[97]

Autistic subjects are thought to lack sensitivity to psychological states of other

people.

Origins of Modular TT

To understand the notion of modularity that operates in TT it is necessary to first

make a detour to explore Fodorian modularity. This notion of modularity predates

TT and differs significantly from it. Nonetheless it has been very influential in TT

and a grasp of this conception of modularity is essential if we are to gain an adequate

grasp of modular TT. The foundations of modularity theory were laid by Fodor in

the early eighties.[98]

Fodor postulated a set of nine criteria for modularity. His view was that for

something to count as modular most of these criteria should be present to some

degree. A number of these features have been adopted by more recent TT theorists.

[96] Ibid P.23
[97] Ibid
[98] Fodor 1983

69

Others have been disputed. Some features were seen by Fodor as more important

than others and it is worth examining these features more closely.

Fodor thought input systems were modular and his discussion is primarily concerned

with such systems.[99] These are contrasted with central systems responsible for

thinking and general intelligence. The properties in virtue of which input systems are

modular are properties which are not shared by central cognitive processes.[100] Other

TT theorists such as Carruthers have extended the range of modularity to include

central systems. This is something that will be discussed later.

Domain specificity refers to the range of information which is able to be processed

by a modular device. To put it another way it is a restriction on the range of content

such a device may take as input.[101] We should expect this range to be relatively

narrow. If we wish to establish whether a system is modular, a key characteristic of

modularity to look out for is whether the system operates in an eccentric domain.

This is because a strong rationale for making such a system modular is that the

associated computations it performs are idiosyncratic.[102] Fodor postulates the

maxim: specialized systems for specialized tasks.

---

[99] Input systems present the world to thought (p.40). They are specialized computational mechanisms which have the function of performing inference-like operations on representations of impinging stimuli in order to generate hypotheses (p.73). There were at least six input systems for Fodor; one for each of our five senses and one for language.
[100] Ibid P.47
[101] Carruthers 2006 PP.3-4
[102] Fodor 1983 P.52

<u>Mandatoriness of operation</u> refers to the fact that the module functions automatically rather than being under conscious control.

<u>Limited central access</u> means central systems have limited access to representations within the module. The intermediate representations a module creates and processes prior to producing an output are not accessible to consciousness.[103] Accessibility is being understood here as availability for explicit report of the information encoded at these levels. There is virtually no access to the lowest levels. Even if one believes that perceptual processing is bottom up, accessibility by central systems is top down. The further one gets from concrete lower levels towards more abstract levels of encryption, the more accessible the information becomes.[104]

<u>Speed of processing</u>, because the processing is highly specialized and working with limited input it will be comparatively fast compared with paradigmatic central processes such as problem solving.[105] The speed of these processes is likely to be related to the earlier criteria of mandatoriness. Making a process automatic saves on computation about whether a process should be performed, and if so how, thereby also saving time.[106]

<u>Information encapsulation</u> concerns the information potentially available to a module. This criterion was particularly important to Fodor. Informational

---

[103] Robbins 2010
[104] Fodor P.56
[105] Ibid P.63
[106] Ibid

71

encapsulation says that the module does not have open access to information

contained in other systems. This would include relevant background information.

Fodor also states that certain sorts of feedback are incompatible with a system's

being modular.[107] This is feedback from higher levels of representation, for example

expectations or beliefs.

Informational encapsulation, like mandatoriness of operation, also helps to account

for the speed of modular systems. Only a small portion of the possible information

that might be analysed will be considered.[108] Speed comes at the expense of

intelligence.[109] This is not necessarily a criticism though, because the task of

modular systems is only to provide representations of very specialized inputs.[110]

Another way Fodor makes this point is to say modular systems are impenetrable

where penetrability is susceptibility to top down effects during processing.[111] Putting

the point this way lets us see one reason why information encapsulation is seen as

such an important criterion of modularity for Fodor; cognitive penetrability is a

salient feature of central systems.[112]

Shallowness of outputs is the result of processing being computationally

undemanding along with the impoverished nature of the input.[113] Fodor's arguments

---

[107] Ibid P.66
[108] Ibid P.70
[109] Ibid P.80
[110] Ibid P.70
[111] Ibid P.74
[112] Ibid P.83
[113] Robbins 2010

for this criterion are closely bound up with what he has to say about informational encapsulation. The above discussion of informational encapsulation might be taken to imply that modular systems have complete autonomy from the rest of the organism, an implication that would be totally implausible. Clearly modular systems have to interact with background knowledge and other processes at some stage. Fodor's solution is to draw a distinction between the outputs of modular systems and the interlevels of representation at work within the system. He argues no interactions with other systems take place internally.

The output of modular systems is likely to be shallow because they are informationally encapsulated. As Carruthers explains, outputs are shallow in the sense of being non-conceptual. They generate information but do not issue in thoughts or beliefs.[114]

Instantiation in specific hardwired neural structures. Modules are realized in dedicated neural architecture. Fodor argues there is characteristic neural architecture associated with each input system.[115] Again this leads us back to an earlier criterion of modularity: "the intimate association of modular systems with neural hardwiring is pretty much what you would expect given the assumption that the key to modularity is informational encapsulation".[116]

---

[114] Carruthers 2006 P.4
[115] Fodor 1983 P.98
[116] Ibid

Neural architecture is thought to be a natural concomitant of informational encapsulation.

<u>Susceptibility to characteristic breakdowns</u> Modules can be selectively impaired with little or no effect on the operation of other systems. One reason a modularist would predict these characteristic breakdowns is because input analysis is principally affected by specific, hardwired neural circuitry.[117] Furthermore pathologies relating to input systems are caused by disruption to these specialized circuits.[118] The association of input mechanisms with hardwired circuitry is one way in which they contrast with central systems, which are neither associated with specific circuitry nor subject to such well defined breakdowns.[119]

<u>A characteristic pace and sequencing of ontogeny</u>. This final criterion is suggested speculatively by Fodor. Modules develop at a characteristic pace and sequence in accordance with specific, inwardly determined patterns under the influence of environmental triggers.[120] This is taken to imply that they are innate.[121]

We have examined nine criteria for modularity suggested by Fodor. Given the extent of Fodor's influence about what it is to count as modular, together with the fact that Fodor did not think mind reading abilities were modular, it may be surprising that TT claims mind reading tasks are executed by modular systems. In fact theory

---

[117] Ibid P.99
[118] Ibid
[119] Ibid
[120] Ibid P.100
[121] Robbins 2010

theorists adopt many of Fodor's criteria of modularity. Although Fodor has been a big influence on modular TT he was not of the opinion that our mind reading capacities were modular. In fact Fodor's conception of modularity makes such abilities archetypal non-modular processes.[122] The mind is envisioned as a general purpose computer with modular input systems.[123] Fodor distinguishes input systems from central systems and argues that only the former are modular. TOM abilities belong to central systems.

Massive Modularity

We now come to the way modularity features in TT. It is offered as an answer to question 3. It is a competing answer to the one given by child scientist versions of TT. In direct opposition to Fodor's conception of the mind as a general purpose computer, Carruthers argues that the mind is a massively modular device.[124] As evidence against a general purpose system he points to the variety of challenges that humans and other species face, for example, the various computational challenges involved in learning. Although in all such cases the challenge is to extract information from data, it is implausible that a single general learning mechanism could underlie vision, speech recognition and mind-reading to name just a few. Each, Carruthers argues, would require the existence of a distinct specialized mechanism.[125]

---

[122] Carruthers 2006 P.5
[123] Ibid P.2
[124] Ibid P.40
[125] Ibid P.29

A second line of attack Carruthers pursues is to appeal to evidence demonstrating that animal minds consist of large sets of belief generating, desire generating and emotion generating modules. The systems responsible for the selection, organisation and control of action are also constructed on modular lines. Carruthers argues that there is nothing left that might be termed non-modular. Next he argues it would be very surprising if the same did not turn out to be true of human minds. Almost all biological structure has been preserved in the evolutionary transition from apes to hominids as it is in most evolutionary transitions.[126]

To understand what is being claimed it is best to start by looking at what is meant by a module here as Carruthers' conception of module differs from Fodor's. Carruthers' usage of this terminology is closer to that found in biology and in artificial intelligence.[127] For him a module is a functionally distinct processing system of the mind. Its operations are at least partly independent of the operations of other modules. Its existence and properties are also at least partly dissociable from other modules. A module will be frugal in its use of information and other cognitive resources and its operations will be largely inaccessible to other systems. Carruthers also cautions against thinking of modules as analogous to material objects. This would encourage the view that modules are physically discreet from one another and difficult to modify.[128] As biological systems, modules are likely to be constructed by connecting resources previously available to other functions in new ways. For this reason they are likely to share parts. This is consistent with their being functionally

---

[126] Ibid P.34
[127] Ibid P.xii
[128] Ibid P.21

specialized and independently changeable. They are likely to exhibit a high degree

of flexibility in development and in response to environmental change

Massive modularity is the position that a mind consists of a very large battery of

functionally distinct processing systems, possessing the properties of a module

described above. These processing systems stand in multiple input and output

relations to other modules.[129] Many of these systems will themselves be composed

out of further arrangements of modules.[130] In the case of the mind these modules

exhibit a hierarchical organization.[131]

Carruthers locates his conception of massive modularity somewhere between the

weakest sense and the strongest sense. The weakest sense of module is 'dissociable

functional component.[132] The weak view simply claims that the mind consists of

entirely distinct components, each of which has a specific job to perform in the

functioning of the whole. The strongest sense of modularity is one that includes all

the properties of a Fodor module.[133.] However we will see that Carruthers regards a

number of these properties as incompatible with his thesis.

Developing his intermediate position on modularity, Carruthers argues we should

also expect the properties of many of these modular components to vary

independently of the properties of others. This thesis should also predict that

---

[129] Ibid P.xii
[130] Ibid P.35
[131] Ibid P.60
[132] Ibid P.2
[133] Ibid P.3

individual components can be modified independently of other modules. Finally the thesis would predict that it is possible for some of these components to suffer damage or be missing while leaving the functionality of the whole at least partially intact.[134] Although modules are functionally distinct this does not necessarily mean they function in radically different ways from one another. The reason why can be brought out by examining arguments Carruthers offers in support of massive Modularity.

The main argument Carruthers offers is 'the argument from design for massive modularity'.[135] This argument is based on the design of complex functional systems, particularly as they are found in biology. It suggests we should expect such systems to be constructed hierarchically out of dissociable sub-systems, in turn composed of further sub-systems and so on.[136] They should be constructed in such a way that the whole can be assembled gradually by adding sub-system to sub-system while the properties of individual sub-systems can be varied independently of one another. The system should be constructed in such a way that the functionality of the whole is to some extent shielded from change or damage to individual parts. Because of evolutionary constraints biological systems tend to display a hierarchical modular organization.[137] This is the most effective way of allowing new modules to be added and of allowing present modules to be modified with minimal effect on overall functionality. Attention to evolution further suggests that each module's processing

---

[134] Ibid P.2
[135] Ibid PP. 12-13
[136] Ibid P.13
[137] Ibid P.21

algorithm is unlikely to be totally unique. A more conventional evolutionary course

would be to copy an algorithm from an existing module and then adapt it to a novel

task.[138] In some cases the algorithm will not need modifying at all; all that will be

required is the provision of new input or output connections.[139] This argument

pictures human development as continuous with other species. The minds of other

species from insects to chimpanzees are argued by Carruthers to be massively

modular. In the course of the transition to the human species some modules, such as

the mind-reading system, are likely to be added or enhanced.[140]

The biological evidence is argued to allow for the construction of an argument in

favour of the weak form of the modularity thesis. It suggests that the mind will

consist in a very large number of functionally distinct components.[141] This argument

allows us to predict that cognition will be structured out of systems that are to some

extend dissociable, each of which has a distinct function.[142] As well as widespread

evidence in biology there is also neuro-psychological evidence that the human mind

is massively modular.[143] Damage can affect a particular processing system while

leaving other systems more or less intact. For example autistic subjects lack the

specific ability to reason about mental states while still able to reason about other

tasks. A different channel of evidence comes from artificial intelligence. Human

designers of intelligent systems have also converged on modular organization which

---

[138] Ibid P.31
[139] Ibid P.32
[140] Ibid P.XII
[141] Ibid P.20
[142] Ibid PP.17-18
[143] Ibid P.18

Carruthers suggests is further evidence human minds are themselves modularly organized.[144]

Summary

Carruthers does not dispute the validity of Fodor's criteria of modularity. However, while Fodor thought that most of the criteria must be present for a system to count as modular Carruthers is more lenient. This difference stems from a disagreement between these theorists about how to characterise the mind. Whereas Fodor pictures the mind as a general purpose computer with a limited number of peripheral modular components, Carruthers argues the mind as a whole is a modular system.

It is worth pausing to consider what type of arguments Carruthers is providing for massive modularity. His arguments are motivated by empirical concerns drawn from evolution, biology, neuropsychology and artificial intelligence. Perhaps the central consideration underlying Carruthers' position is continuity. He is concerned with both the continuity between humans and other species and the continuity manifested throughout evolution in general. Carruthers argues that there is empirical evidence that the minds of other species are entirely modular. For Carruthers to claim something is a biological system implies that that system is modular. Here modularity means being composed of a number of functionally distinct processing systems. Carruthers stipulates this as a minimum requirement for modularity whereas he thinks possession of all nine Fodorian criteria would constitute the

---

[144] Ibid P.23

80

strongest form of modularity. Indirect evidence for massive modularity is provided through Carruthers' appeals to human continuity with other species which are understood to be entirely modular and appeal to the character of biological systems in general. Carruthers also offers more direct evidence for massive modularity. Part of this concerns brain damage in autistic subjects which appears to be confined to a specific processing system leaving other functions intact. Carruthers also offers a less convincing argument. He argues that a human propensity to converge on modular systems (exhibited by designers of artificial intelligence systems) reflects the designer's own modular organisation.

Carruthers on Fodorian modularity

Carruthers is appealing to modularity as an answer to question 3; how is attribution of mental states to others achieved? For this reason he has to take issue with certain aspects of Fodor's conception of modularity which is only applicable to peripheral faculties and not to centralised ones such as mind reading abilities.

We saw above Carruthers thinks the strong view of massive modularity is not viable because a number of Fodor's criteria are incompatible with massive modularity. The biggest dispute between these theorists concerns the criterion of informational encapsulation. Carruthers thinks this is unnecessary for modularity. While Fodor argued encapsulation was a core property of modular systems, Carruthers argues massive modularists shouldn't claim that the mind consists exclusively of

encapsulated systems. Even where a system has been designed to process a particular domain of inputs it may need to query other systems for relevant information.[145] For example suppose we accept that the mind-reading system is designed to focus on behaviour together with attributions of mental states and to generate predictions for further behaviour or attributions of other mental states it may still need to query a range of other systems for information relevant to solving the task at hand.

Carruthers does allow for a more limited sense of encapsulation. Recall Fodor argued that encapsulated systems cannot draw on information held outside of the module in other systems. Carruthers suggests what really matters are that modules are frugal in terms of information they require and in terms of complexity of their processing. He does not think this requires encapsulation in a strong sense. Carruthers distinguishes narrow scope and wide scope encapsulation.

Narrow scope encapsulation makes a point about most of the information stored in the mind; it states that a module cannot be influenced by this information in the course of its processing. Encapsulation here envisions a large determinate body of information that cannot penetrate through to the internal operations of the module. This is how encapsulation is normally understood in the modularity debate.[146]

---

[145] Ibid P.10
[146] Ibid P.58

Wide scope encapsulation states that the module cannot be influenced by most of the information held in the mind in the course of its processing. Carruthers argues that this can be the case without a determinate subdivision between the information that can't affect a module and the information that can.[147] What matters is that a module must only consider a small subset of the total information. Whether this is achieved by narrow or wide scope encapsulation is unimportant.

The requirement for frugality still provides support for the requirement that processing systems should have internal operations that are inaccessible elsewhere.[148] To see this consider what would happen if a module's internal operations were accessible to other modules. To make use of these operations other modules would require an internal model of them. This would defeat the purpose of dividing up processing in the first place.

A second criterion that has to go is shallowness of outputs. Fodor thought the outputs of a module are shallow in the sense of being non-conceptual. They generate information of certain sorts but they do not issue in thought or belief.[149] This only makes sense if modularity is confined to fairly basic processing. It is not a coherent criterion if the outputs in question are going to be fully conceptual thoughts or beliefs.[150]

---

[147] Ibid P.58
[148] Ibid P.59
[149] Ibid P.5
[150] Ibid P.7

The third of Fodor's criterion to get dropped is speed of processing. This is also redundant as a criterion of massive modularity, as speed is meant to be relative to the speed of processing in non-modular systems. For Fodor central systems were non-modular so the comparison was meaningful but if all aspects of mind are modular then no such comparisons can be drawn.

The fourth and final criterion to be dropped is a characteristic pace and sequencing of ontogeny which Carruthers refers to as 'innateness'. Carruthers thinks that it may well be the case that many of the mind's modules are innate. Modules likely to be innate include learning systems designed by evolution to acquire and store information about some aspect of the environment. As an example he gives the mind reading system which builds up knowledge of other people's beliefs and goals. However Carruthers argues one should not insist on innateness as a criterion of massive modularity as it is a controversial one.[151]

Although Carruthers retains domain specificity he thinks strictly speaking massive modularity is unable to insist on it. However he argues that many modules will have this feature.[152] Carruthers thinks domain specificity could be dropped without compromising the massive modularity thesis.

Carruthers agrees modules will have instantiation in specific hardwired neural structures though these structures may vary between individuals. Limited access by

---

[151]Ibid P.10

[152] Ibid P.8

other systems to intermediate representations within the module, mandatoriness of operation, and susceptibility to characteristic breakdowns are all preserved from Fodor's original criterion.

In order to claim mind reading abilities are modular we have seen that Carruthers has had to abandon a number of Fodor's criteria.

TT and TOM

It was explained at the beginning that TT was one possible instantiation of TOM (simulation theory being another). TOM is interested in how we ascribe mental states to ourselves as well as how we ascribe them to others. However an issue that divides TT theorists is whether the mechanisms TT postulates for understanding others are also responsible for self understanding. This is tangential to question 1. Carruthers for example thinks they are not:

"must such a theorist be committed to the implausible view that we know of our own mental states just as we know of the mental states of other people — by means of an inference to the best explanation of the (behavioural) data, operated within the framework of a folk-psychological theory? Most theory-theorists have not thought it necessary to travel this route, maintaining, rather, that self-knowledge should be

thought of as analogous to the theory-laden perception of theoretical entities in science".[153]

By theory laden perception Carruthers is referring to the idea that all perception is mediated through a theoretical apparatus. This enables us to focus on salient features of our environment and filter out unwanted information. This means that all perception requires tacit acceptance of a well entrenched theory. On this view perception of one's own psychological states is still a type of observation rather than inference, but it is theory laden observation. One introspects one's own states but makes sense of them through positioning them within the general theory. For some Theory Theorists our grasp of 1st Personal & 3rd Personal states is analogous; we are applying a theory to make sense of our psychological states. These versions utilise a process of inference even in one's own case. Leslie argues a TOM is also responsible for self understanding. His contribution to TT will be explored next.

Leslie's Theory of Mind Mechanism

Leslie offered what is intended to be a translation of TOM into a model of the underlying processes involved and therefore to directly address question 3. His views have been highly influential. He understands TOM as the ability to make sense of behaviour in terms of mental states.[154] In particular this is done in terms of propositional attitudes:

---

[153] Carruthers 1996 P.26
[154] Leslie and Roth 1993 P.105

"In understanding and predicting behaviour the basic work is carried out by the ability to grasp the role of Propositional Attitudes in the causation of behaviour – by understanding agents and attitudes".[155]

TOM is a theory about the specific representational relations, such as hopes, beliefs, wants etc that play a role in the causation of agent's behaviour.[156]

Leslie postulated that we have a modular processing system called a theory of mind mechanism (ToMM). This mechanism is domain specific (rather than a domain general capacity) being tied to our understanding of agents. This allows us to attend to mental states and to their causal role in behaviour.[157] It underlies our ability to conceive of our own and other people's mental states and to reason about behaviour in terms of such states.[158] This mechanism consists of an inferential device which infers states of mind on the basis of behavioural events, along with representational systems called metarepresentations or M-representations.[159]

Metarepresentations depict particular attitudes that agents can take to situations.[160] They are data structures computed by our cognitive systems.[161] Metarepresentations are contrasted with primary representations. Primary representations are literal

---

[155] Ibid P.90
[156] Ibid P.91
[157] Ibid P.105
[158] Ibid P.92
[159] Ibid
[160] Ibid P.88
[161] Ibid P.87

descriptions of the situation perceived. Metarepresentations on the other hand provide 'agent coloured' descriptions of the situation.[162] Representations of the mental states of others are in effect not representations of the world but representations of representations. They are second order representations.

A metarepresentation depicts a special kind of (informational) relation between the agent and a situation. It has a number of components. To begin with it requires a specification of the identity of the agent. This agent is placed in relation to two things: the agent stands in relation to an aspect of reality which will be depicted by a primary representation. The agent is also related to an imaginary situation. This is described by a decoupled representation.[163] The ability to decouple is an essential ability for metarepresentation of others.

Like Gopnik, Leslie's account of TT is influenced by empirical evidence relating to early development in children. He argues the ability to employ TOM is manifested in a number of ways by means of developments that require the operation of metarepresentations. These include acquiring the language of mental state expressions and understanding the consequences of ignorance and false belief.[164] Metarepresentation also links the ability to pretend and understand pretence in others to the employment of ToMM.[165] Leslie distinguishes development of the capacity

---

[162] Ibid
[163] Ibid
[164] Leslie 1987 P.422
[165] Ibid

for pretence from development in understanding objects and events in general. It represents the beginnings of an ability to understand cognition.[166]

"Pretending to oneself is thus a special case of the ability to understand pretence in others (someone else's attitude to information)"[167]

Pretend play is an early demonstration of the ability to characterize and manipulate one's own and others' cognitive relations to information.[168] This ability is one Leslie argues is central to TOM.

To summarise, we make sense of people by attributing psychological predicates. We rely on a module dedicated to monitoring psychological states and their effect on behaviour. This works by constructing inferences based on a combination of two types of representation. First, primary or literal representations based on behavioural evidence. Second, there are theoretical or metarepresentational representations which depict an agent's relation to a situation.

On this picture we receive psychologically characterised data through observation. To get psychological characterisation we rely on a module that automatically interprets these literal representations. This is an example of TT because we transform the brute representations into metarepresentations by the application of the theory. This is facilitated by a module. It involves making a move to the sub-

---

[166] Ibid P.416
[167] Ibid
[168] Ibid P.422

89

personal level. However sub-personal operations are still characterised using terms such as inference which are primarily suited to personal level explanation.

Central empirical evidence for TT

All versions of TT discussed in this chapter support themselves by reference to certain empirical tests. They offer this empirical data in support of their overall framework.

<u>False belief tests</u>

Central empirical evidence for TOM capacities concern a battery of tests knows as "false belief" tests. The original false belief test was introduced by Wimmer and Perner in 1983. Following Plyshyn they argued that someone who has a TOM will be able not only to represent a state of affairs, but also to explicitly meta-represent the relation in which one stands to this state of affairs.[169] One indication that a person possesses these abilities will be that that they employ meta-representational language terms, for example making reference to another person's belief.[170]

Another more complicated way to demonstrate these abilities is to engage in deception. This requires representing the false belief of another as a sub-goal in the

---

[169] Wimmer & Perner 1983 P.104
[170] Ibid P.105

deception.[171] Wimmer and Perner argued that, while evidence suggests even very young children can explicitly represent the relation that they or another person stand in to a state of affairs, it is a more challenging task to meta-represent the difference between one's own relation to this state of affairs and another person's relation to it. Where these relations are different one has to be able to understand that there another person lacks some salient knowledge that one possesses.

In the original experiment a character named Maxi puts chocolate into a cupboard X. When Maxi has gone his mother moves the chocolate to cupboard Y. The subject is then asked where Maxi will look for the chocolate upon his return. The experiment is designed to investigate whether children can represent others as holding particular beliefs when they themselves know the beliefs are false. The experimental subject and another person (maxi) together observe a certain state of affairs (an object being placed in a particular location). The other person then leaves the experimental location. In the other person's absence, but still in the presence of the subject, the state of affairs is altered in a way that could not be reasonably predicted (the object is moved to an alternative location, not visible from the viewpoint of either observer). The subject now knows the new state of affairs is the case and, if in possession of TOM, should also know that the other person still believes the former state of affairs to be the case.

---

[171] Ibid P.104

To answer the question where Maxi will look for the chocolate successfully, the subject must be able to keep their representation of Maxi's belief separate from their knowledge of the true location of the chocolate. This enables the subject to use the representation as a frame of reference with which to predict Maxi's actions. To interpret or predict another person's actions one must be able to confine oneself to the realm of the other person's beliefs, and this is only possible once the subject's representations of these beliefs are firmly established.

The ability to represent false beliefs is therefore understood as a key stage in the development of a ToMM. According to the experimental research this capacity first emerges at four years old. When asked where Maxi will look for the chocolate, children under this age answered he will look in cupboard Y where they know the chocolate to be.

Autism

The experimental evidence for a ToMM is closely tied up with an investigation into the nature of the deficit present in autism. It is hypothesised by proponents of TT that autism is the result of a damaged TOM module.

The term autism represents a cluster of symptoms. The primary symptoms associated with childhood autism are impairments in verbal and non-verbal

communication.[172] In particular the condition is associated with failure to develop normal social relationships. Autistic subjects have a very pronounced difficulty in understanding and coping with the social environment, regardless of IQ. Autism is also associated with a cluster of other symptoms. These include mental retardation, pockets of ability inconsistent with general ability, and 'insistence on sameness'.[173] Another common symptom exhibited in children is that they do not engage in pretend play. However symptoms such as mental retardation are not necessarily present.

A theory of mind deficit is posited to explain why autistic subjects treat people and objects alike.[174] TOM is a mechanism that enables us to conceive of mental states, for example to conceive that another people wants, believes or feels something.[175] This requires the capacity to form 'second-order representations'. These normally begin to develop in the second year of life and mature at about four years. Earlier we saw that, according to Leslie, pretend play is a central ability of TOM. The lack of this behaviour in autistic children provides further evidence for the hypothesis that autism is a TOM deficit. The reason such diverse symptoms as the inability to engage in pretence, and social impairments, can be postulated to be part of the same condition is because, it is being argued, these symptoms stem from the same deficit; a lack of second order representations.[176]

---

[172] Baron-Cohen et al 1985 P.37
[173] Ibid P.38
[174] Ibid
[175] Ibid
[176] Ibid

Although mental retardation is often associated with autism it is significant that it is not always present. Baron-Cohen et al argue that this enables us to see more clearly that there is a specific deficit involved in autism. Even high IQ cases of autism lack TOM whereas other types of mental retardation still display TOM. This means autism cannot be attributed to the general effects of mental retardation. It is a deficit that is independent of general intellectual level.[177]

Sally-Anne test

Baron-Cohen et al therefore devised a new version of Wimmer & Perners False belief task that has become known as the Sally-Anne test. This test is designed to find out whether autistic children possess a TOM. Subjects are confronted by two Dolls named Sally and Anne. Sally places a marble in her basket and then exits the scene. When Sally is no longer present Anne transfers the marble to her box. The subject is then questioned about where Sally will look for the marble upon her return. To pass this version of the false belief task subjects must point to Sally's basket.[178] Subjects must first correctly answer two control questions in order to demonstrate an adequate comprehension of the demands of the test. They must demonstrate that they know where the marble actually is and that they remember where the marble was before Anne moved it.

---

[177] Ibid P.44
[178] Ibid P.41

This test was applied to high functioning autistic children, children with Down's Syndrome, and 'normal' preschool children. Nearly all Downs Syndrome Subjects and normal children succeeded in the false belief task whereas nearly all autistic subjects failed despite possessing higher IQ scores.[179] Baron-Cohen et al argue these results demonstrate, firstly, that normal and Downs Syndrome children predict the doll's behaviour on the basis of the doll's belief. Secondly the results demonstrate that autistic children are not able to appreciate the difference between their own and the doll's knowledge.[180.] Baron-Cohen et al explain this failure as an inability to represent mental states. In other words autistic subjects do not employ a theory of mind.[181] Because they do not employ TOM autistic subjects are unable to impute beliefs to others. This puts them at a significant disadvantage in social interaction because they cannot predict the behaviour of other people.[182]

The ability to represent false beliefs is being understood as a theoretical development or transition from a primitive theory about what motivates others' behaviour to a more sophisticated one. It involves the development of metarepresentational concepts. Psychological states of other agents, particularly propositional attitudes, are understood as metarepresentational constructs posited to explain unobservable inner states of the other. These theoretical concepts are also employed to account for one's own psychological states; hence there is a challenge of resolving the differences between these two theoretical standpoints. It is

---

[179] Ibid P.42
[180] Ibid P.43
[181] Ibid
[182] Ibid

suggested that autistic subjects have difficulty understanding others because they are unable to arrive at psychological/metarepresentational concepts. The sally-Anne test is offered as evidence autistic subjects do not employ psychological concepts to predict others' behaviour.

Does Theory Theory have an answer to the questions raised in this thesis?

<u>Question 1</u>

Question 1 asks what we are doing when we attribute mental states to others. TT makes two main points in response to this question:

Firstly it suggests that we are attempting to explain observable behaviour by constructing theories. We also employ these theories to make predictions about others' future behaviour. These explanations are causal explanations. There are different accounts of exactly how this is done (whether at a conscious or sub-personal level for example) but all versions understand the task the same way. In this sense, theories of how others behave are analogous to all our other folk scientific theories. Morton, for example, argues that attributing psychological states to others involves forming hypothesis about the best explanation for their behaviour in order to explain and predict what they will do in the future.

TT claims that attributing psychological states involves utilising a body of folk psychological knowledge; the central concepts of which are belief and desire. Though TT claims we are employing a theory, many of its supporters distinguish TT from fully fledged scientific theories. Nonetheless, when we are attributing psychological states to others, the theory theorist will claim we are engaged in a part of the wider scientific project of explaining how the universe works (engaging in

what Stich and Nichols call the dominant explanatory strategy). Indeed, Gopnik would suggest an answer to question 1 is equivalent to an answer to how we make sense of anything in the universe. It is an answer that does not treat mental states as distinctive.

The second point TT makes is that an important feature of these theories is that they make reference to unobservable psychological states. All theory theorists agree the psychological states of others are unobservable. However TT's answer to question 1 is often bound up with the question of what we are doing when we attribute psychological states to ourselves. There is disagreement among theory theorists about whether it is also true in the case of self attribution that we are working with unobservable psychological states. Many theory theorists do think these are also unobservable . Others including Carruthers argue the answer to these two questions will be different. While Carruthers claims we know the psychological states of others through an inference to best explanation of the behavioural data made within a folk psychological theoretical framework, in our own case we introspect theory laden states. However although we can observe these states in our own case, what we are observing will already be theory laden. The states have their meaning in virtue of the theory they belong to. Psychological states are envisaged as lying behind the behaviour we actually observe. We encounter neutral data and infer psychological meaning.

The answer TT gives to question 1 is closely connected to a sub-question embedded

in question 1: what is the meaning of our psychological terms? TT's answer to this sub-question is that psychological concepts gain their meaning as theoretical concepts from their location within the explanatory theory in which they are embedded. As Morton, for example, explained TT is suggesting that psychological concepts reference individual members of an interrelated set of beliefs which taken as a whole constitute an implicit theory of the mind. We also saw with Carruthers that the attribution of a psychological state to another agent involves grasping how that psychological state is related to other psychological states and laws belonging to the same theory.

Problems raised by this answer to question 1

A question that should now be raised is whether those assumptions described above are empirical or philosophical ones. The above answer to question 1 is not the result of straightforward empirical deduction. Although any individual attribution of a psychological state to another agent is an empirical matter which is either accurate or inaccurate, the claim psychological states get their meaning from their role in a theory depends on accepting philosophical assumptions about what sort of things psychological states are in the first place. Accepting TT involves more than simply observing what another agent is doing and making a hypothesis about their psychological state. A theory theorist never begins with a neutral observation of a material body and constructs a hypothesis about what animates it. The picture of the practice being described comes laden with assumptions about the meaning of

psychological terms as inner states, and about the practice of attributing inner states to others. This is not a neutral empirical description of the practice; it is informed by a particular philosophical picture of the kinds of things psychological states are. Theory theorists are offering an account of what is happening behind the behaviour we empirically observe. One could agree with their account of the empirical goings on and still disagree with the philosophical picture. As the account is really a philosophical one it raises philosophical issues that need to be evaluated.

The main problem is that all theory theorists assume that psychological states are theoretical. The practice of attributing psychological states is seen as a practice of employing a folk scientific theory where the meaning of psychological states is fixed by their role in this folk scientific theory. TT would claim folk psychological explanations referencing propositional attitudes are causal explanations which can be integrated into a scientific psychology.[183]

Greenwood accepts that folk psychology is theoretical. He distinguishes two uses of theoretical, firstly the term may be used to characterise any description of properties or relations. Secondly it can be used in a more specific sense to characterise descriptions of postulated causal explanatory dimensions with respect to a particular domain.[184] He thinks all classificatory descriptions of human action are theoretical in the former sense. However Greenwood argues folk psychological descriptions of human action are not theoretical in the latter sense. The intentional nature of human

---

[183] Greenwood 1991b P.4
[184] Greenwood 1991c P.72

action cannot be explained in terms of causal explanatory factors. Folk

psychological classifications do not presuppose any causal explanations and are

compatible with competing causal explanations (e.g. an intentional act may have

been the product of psychological motivations or neurological explanations).

Although he thinks they are theoretical Greenwood denies that folk psychology is a

causal explanatory theory. He argues psychological characterisations are neutral

about their causes.[185] Greenwood argues that this is not to deny folk psychological

assertions are subject to empirical evaluation, there is a wealth of evidence that

people act intentionally.

It is puzzling why Greenwood wants to retain the idea our folk psychological terms

are theoretical at all. This use of theoretical does not amount to more than a claim

that concepts are by nature theoretical or they could have no role in a language.

There can be no brute facts unconnected to the rest of our knowledge. However this

is simply a description of what psychological concepts are like but it is not part of

any explanation of how we are able to employ them. It does not tell us anything

about how their meaning becomes fixed in the first place. This stretches the meaning

of theoretical to the point where it loses its usefulness. It does not enable one to

distinguish the role of psychological terms from any accounts of any other

phenomena. Standard TT in contrast takes us beyond the descriptive by postulating

unobservable causally explanatory states.

[185] Ibid P.73

Another problem with TT's answer to question 1 stems from the claim we are always engaged in hypothesis formation during encounters with others. This claim lacks phenomenological plausibility. It suggests that we are always behaving as amateur scientists seeking to explain or predict the things confronting us. However it is arguable that this is not the case. It is not the case that we always take ourselves to be engaged in tasks of prediction and causal explanation. Sometimes, for example, we might simply sympathise with another's situation. It is because of this phenomenological implausibility that many versions of TT retreat to the sub-personal level.

It is also very difficult to accept our everyday psychological vocabulary is the product of explanatory hypothesis formulated to explain behaviour which can be identified independently of such a hypothesis. It implies there is a way of describing the phenomena which is either pre theoretical or employs a different vocabulary to the vocabulary of folk psychology. Such a vocabulary will not reference the regularities our current theory allows us to pick out. On this view, question 1 cannot even be formulated prior to the establishment of TT as the dominant mode of explanation.

TT's answer to question 1 purports to be an empirical move and indeed is centred on empirical data. In addition to the issues raised above there are other philosophical questions tied up in this. For example Carruthers' concerns are also motivated by tacit underlying philosophical assumptions. He is in the grip of a picture of how

102

items in the world can be expected to behave influenced by what we have learned from previous empirical enquiry. It is a philosophical picture in which we are primarily biological organisms behaving in conformance with evolutionary principles. On this picture one should expect empirical discoveries made about other species are likely to apply to humans as well. This picture informs his arguments for modular TT over child scientist versions. Carruthers' arguments stem from the premise that modularity is an extremely pervasive characteristic of biology rather than being motivated by anything distinctive about mind-reading abilities per se. Neither are they based on direct empirical evidence for an overarching modular organisation in humans. They focus on establishing that the mind in its entirety is modular rather than being focused on the modular status of TT capacities. Carruthers sees no reason why we might expect the realm of the psychological to be distinctive. He appears more focused on vindicating the modular model itself rather than on using it to see what light it can shed on our understanding of the psychological. One concern Carruthers has is that he finds it difficult to see how else human accomplishments could be satisfactorily explained if human beings did not possess a modular organisation.

This issue about what the nature of the practices of attributing psychological states to others are is a philosophical one. There is a general assumption these practices must be continuous with natural science but this is just an assumption and not itself an empirically established fact. To a lesser extent this point also applies to a number of other TT positions. The claims made by theory theorists are not driven by an

impartial reading of the data. The interpretation of the data is influenced by general empirical assumptions (such as that our development and design will be continuous with other species).

Question 2

TT has much less to say that could be employed in an answer to question 2. This is the question about what justifies the claims we make about the mental states of others. TT will claim our psychological theories are justified in exactly the same way all scientific theories are justified; by their success in generating prediction and explanation. They are folk psychological predicates.

Justification is not a key concern for TT. Some TT advocates including Gopnik claim the development of our ability to attribute psychological states to others is the outcome of a succession of increasingly sophisticated theories. In this case it would appear to be a matter for evolution. For Gopnik scientific advancement boils down to a capacity to formulate and progress through a sequence of theories and is a capacity we possess innately. Carruthers would claim that the attribution of a psychological state to another would be justified on the basis of a mind reading module alone. It will not make reference to information external to the module itself.

Problems with this answer to question2

An important consequence is that if this is the basis for a particular theories'

justification then that theory will in principle be open to revision, replacement or

elimination. As Greenwood argues the price to pay for granting folk psychological

descriptions a causal explanatory theoretical status is the risk of falsification. They

can no longer be assigned a privileged conceptual position in the explanation of

human action.[186] These explanations can be rejected if they turn out to be inaccurate

or merely impoverished as causal explanations. In such a scenario we should be

obliged to abandon this form of psychological explanation and the ontology of

psychological states that lie behind it.

We saw in the discussion of question 1 that Greenwood takes issue with

Churchland's employment of the term theory.[187] It might seem that his notion of

theory avoids the danger of elimination. However the looser notion of theory

employed by Greenwood above cannot save the authors we have been considering in

this chapter from the threat of eliminative materialism as they clearly have a much

stronger version of theory in mind. One problem for the theoretical view of

psychological states then, is that they must face the threat of elimination.

An important difficulty this answer to question 2 raises is that it is highly

questionable whether it is in fact coherent to treat our psychological descriptions as

in principle eliminable or even substantially replaceable. Greenwood thinks evidence

[186] Greenwood 1991b P.5
[187] Greenwood 1991c P.72

of the inadequacy of causal explanation of human action made in terms of psychological states will not oblige us to abandon using folk psychological description or its underlying ontology.[188] Such states are individuated independently of any postulated causal explanatory dimensions. He takes issue with the eliminativist claim that the meaning of theoretical descriptions are fixed by reference to the causal explanatory propositions in which they occur.[189]

Greenwood argues we could meaningfully ascribe psychological state to others and have evidence for their existence even without a theory of their causal relation to human action. He also points out that experimental evidence suggests children are capable of employing folk psychological concepts from two or three but they do not learn to employ them in causal explanatory descriptions until around four years.[190] The independent evidence for folk psychological ontology includes self-knowledge of psychological states. According to Greenwood, these may be theoretically informed but their accuracy will not be threatened by the inaccuracy of any causal-explanatory relations in which they are employed.[191]

An implication of TT's answer to question 2 is that one is justified in attributing a psychological state to another to the extent that this attribution proves useful in explaining the behaviour in question. One worry is that psychological predicates will not prove maximally useful and will therefore be vulnerable to replacement by more

---

[188] Ibid P.75
[189] Ibid P.76
[190] Ibid P.81
[191] Ibid P.82

106

precise forms of explanation. However, according to Greenwood, many contributors to the debate who defend folk psychology accept that it is a causal-explanatory theory; but deny it is a stagnant one as the eliminative materialist would suggest.[192] They dispute the idea folk psychology displays widespread explanatory failure.

Another worry is that empirical success is not an exhaustive justification for folk psychology even if it is relevant. Greenwood claims explanatory success is not the only relevant justification for accepting the ontology of folk psychological states. He claims that often these phenomena, like the subject of a great deal of theoretical description, can be discriminated more or less directly. The causal or functional role of such properties consequently becomes an object of empirical investigation.[193] However the failure of causal accounts is compatible with retaining the ontology.[194]

This threat of elimination only arises when treating psychological descriptions as a folk theoretical theory that may or not be improved. There are reasons why this may not be coherent. It is hard to take seriously the idea of abandoning our folk psychological predicates. As Greenwood argues we cannot abandon the ontology of folk psychology without abandoning what we are supposed to be trying to explain. We can't give up talk of psychological states without giving up on the attempt to explain human action.[195] One cannot even articulate arguments without referencing

---

[192] Greenwood 1991b P.6
[193] Greenwood 1991c P.79
[194] Ibid P.80
[195] Ibid P.70

107

psychological states. To abandon the ontology of folk psychology would involve

abandoning the subject matter.[196]

A related objection is that there is no coherent successor to folk psychology. The

eliminative materialism argues that talk of psychological states and motives should

be left out of mature science in favour of talk of neurological states which are the

product of scientific theories. However the idea of a more mature scientific theory,

presumably using the language of neurology is just implausible. Even if we were

able to acquire such a vocabulary it would not give us what we seek.

Question 3

Question 3 is where the main disputes among theory theorists occur. It asks how the

attribution of psychological states to others is achieved. Question 3 can also be

understood as addressing a sub-question about how we first acquire this body of

theoretical knowledge which we learn to attribute to others. TT understands question

3 as an empirical question. There are different levels at which an answer to question

3 can be cashed out including functional and physiological explanations. We saw

that, in his discussion of the emergence of TT, Morton identified a converging

interest in constructing answers to question 3, in particular of how we acquire

knowledge of the requisite theoretical states such as beliefs and motives which we

are to subsequently attribute to another agent. When we examined Morton we saw

that TT offers a reply to question 3 in functional terms. We are utilising very general

---

[196] Ibid P.71

psychological concepts. We learn to equate particular instances of these psychological concepts which we observe to general ones. It is the theoretical framework as a whole which facilitates attribution of particular concepts.

Different variations of the theory offer competing answers to question 3. The main dispute is between child scientist and modular versions of the theory. Child scientist versions claim we learn the theory either in a deliberate manner or through passive cultural assimilation. Modular versions claim we rely on genetically bestowed modular structures which equip us for recognizing psychological states. There is a lot of disagreement about what this modular structure consists in. The central disagreement over question 3 seems to concern the origin of the mechanism facilitating our ability to attribute psychological states to others. Gopnik cashes the answer out in terms of manipulation of abstract representations and rules. This operation is the product of an evolutionary capacity. Carruthers conceives this ability as resulting from the development of an innate mechanism rather than through the replacement of a preceding theory. It is achieved through the operation of modular systems. Carruthers would also add that what we are doing is acting in conformance to a preset modular programme. The modular conception of TT also harmonizes with the view of individual psychological states as instances of something very general. New modules tend to duplicate as much as possible and so avoid working with novel information.

There is also disagreement about whether these competing explanations should be

understood as explanations about what is occurring at the personal or sub-personal level. A move to the sub-personal is partly motivated by the obvious objection to TT discussed above that it is phenomenologically implausible to suppose we are relying on a theory in order to attribute psychological states. Theory theorists respond by arguing that this process need not be conscious and many believe the bulk of our theorizing is performed in a non-conscious manner. We do not on this view have conscious access to most of the theorizing that occurs. This kind of theorizing is more akin to the way children make sense of the world than to the explicit theorizing characterising scientific procedure.

Leslie's work, suggests that making sense of others is a matter of processing representations. More specifically we are processing psychological representations (metarepresentations) which are operating alongside literal representations of features of the world. The form these psychological representations take is constrained by the overarching theory. This distinguishes psychological representations from literal representations.

Much of the empirical dispute is focused on autistic subjects who are supposed to lack the ability to attribute psychological states to others. TT addresses question 3 partly through an appeal to autistic subjects. It is hypothesised that they lack a TOM. This means they can be compared with 'normal' subjects in order to find out what normal subjects are doing that autistic subjects are not. They do not employ a theory to attribute states to others. By looking at what faculties they lack the idea is that

Theory Theorists can work out what faculties we employ in order to address question 3.

Problems and Issues with TT's answer to question 3

It is important to note that a question about how something is being done awaits an answer to what is being done. The answers that TT provides to this question only make sense in relation to a particular answer they have already supplied to question 1. One has to accept TT's answer to the former in order to accept their answer to the latter. Only then does it become an empirical question how the process described in the answer to question 1 is achieved. TT interprets question 3 in a different way to the way we will see Wittgenstein and Merleau-Ponty will interpret the question. Gopnik for example interprets the question as asking for a particular kind of account of the structural features enabling explanation and prediction. Whereas Wittgenstein and Merleau-Ponty think the place to begin is with close examination of the practice of attributing psychological states; for Gopnik and other theory theorists, a straight forward description of the practice would be regarded as trivial. They are already in the grip of a picture of the practice in which what people are doing is engaging in tasks of explaining and predicting the empirically observable and neutrally describable behaviour of others. The deeper question they set themselves is how people proceed with this process of explanation and prediction. For example we saw that, according to Gopnik, the answer to question 3 described in functional terms is that we are interpreting neutral data according to a non-theory neutral vocabulary.

We are employing the vocabulary of psychological states to interpret non-theoretical perceptual data. Gopnik would add that we are employing the latest in a succession of theories about the causes of other people's behaviour.

However, even within this response to question 3 there are still philosophical assumptions bound up with this answer: One key philosophical issue concerns the moves TT makes between the personal and sub-personal levels of explanation. This is really two related problems. Firstly that TT assumes it is coherent to move smoothly between these levels. Secondly the use TT makes of sub-personal is particularly problematic. We will look at these problems in order. We noted above that TT lacks plausibility as a personal level account of how we are able to attribute psychological states to others because we have no phenomenological awareness of relying on a theory. We saw that to avoid running onto this objection theory theorists often appeal to sub-personal processing. For example Carruthers assumes continuity between low level biological processes and mind reading. Leslie also applies personal level psychological concepts to the explanation of sub-personal brain activity. He also tries to accommodate the phenomenological implausibility of claiming we experience brute data which we subsequently bestow with a theoretical characterisation by arguing it is already psychologically interpreted behaviour by the time it reaches phenomenological consciousness. We receive neutral data through the sense organs which is subsequently operated on by a ToMM. Sub-personal operations are described in quasi intentional terms by TT.

This takes us to the 2nd problem. It should be noted that it is personal level

theoretical concepts which are doing the work in these accounts. They do so by virtue of occupying functional roles within the larger framework. However it is a highly questionable assumption that personal level psychological concepts can be given this application. It is unclear that TT should be permitted to continue using these same terms to describe sub-personal activity. If at the sub-personal level it does not make sense to describe the processes like this then these versions of TT become vulnerable to philosophical objections.

If Leslie's work is to be understood as providing an account of the sub-personal processing underlying psychological attribution, this is problematic. It involves treating sub-personal states as equivalent to genuine intentional states. The account of metarepresentations enabling an understanding of pretence is an example of a case where the distinction between the personal and sub-personal levels of explanation becomes blurred. Furthermore it is not just the psychological states themselves that are lifted from the personal level. Engaging in deduction is also a personal level phenomenon.

Leslie claims to offer an account of what happens at the sub-personal level when we understand others. It is one which makes use of personal level concepts such as pretence. The idea that we must make sense of behaviour in terms of mental states implies that there are dual levels of explanation confronting the mind reader; a merely behavioural level and a level of intentional action. The intentional level is constituted by theoretical constructs called propositional attitudes.

An implausible implication of the TT strategy is that the real regularities explaining another's behaviour are internal theoretic states whereas the regularities we normally take to drive other people's behaviour are merely superficial. This difficulty is enhanced by the relocation of the theory to the sub-personal level.

Given that TT answers question 1 by claiming that we are making use of an internal and restricted set of concepts and rules specifying the relations between these concepts one might expect this strategy to lend itself to modular classification. However, even Carruthers is reluctant to insist on encapsulated systems being necessary for making sense of others. He thinks such systems would not be capable of addressing question 3 because information outside the mind reading module will be required. Nonetheless his answer would still assert that we rely on a rapid search through a frugal stock of information. He advocates a more moderate notion of wide scope encapsulation in order to avoid a sharp subdivision between information contained in the module and information outside it.

Leslie also recognises that the attitude an agent has to a situation transcends what could causally flow from the literal features of a situation. His notion of metarepresentation may initially seem appealing if it is taken to suggest that these enable agents to directly confront the intentional aspect of another agent's relation to their situation. However, despite the fact he recognises that there is something special about mental states he construes the transition between propositional

attitudes as an entirely causal affair. This is unsurprising as he is working in a TT

framework. One reason this is implausible is because a causal model will struggle to

capture the features of a situation that make a particular agent's attitude towards it

intelligible. Though both Carruthers and Leslie are aware that there are limitations to

what can be captured by a model of causal inner states neither break free from such a

model.

TT offers differing answers to question 3 though it is hard to see what is at stake in

this dispute even taken as an empirical question. It is difficult to tease out the

empirical differences between Gopnik's appeal to an evolutionary bestowed capacity

to manipulate abstract representations and rules conflicts with Carruthers' idea that

theoretical tools are ultimately something we are innately equipped with.

The strategy of identifying particular instances of a psychological state (embodied in

another agent) with abstract theoretical states (posited by a TOM) in order to make

judgements by relying on one's knowledge of the ways in which this theoretical

state is related to other theoretical states leaves open an important question; how are

we able to identify the particular states as instances of a general theoretical state?

What features of a particular state enable us to grasp it as an instance of something

more general?

The false belief test envisaged by Wimmer and Perner also adds a layer of

complexity to TT's answer to question 3. Successful attribution is more difficult if

the epistemic psychological states of another agent conflict with one's own. In such cases, part of the task involved in accurate attribution of a state to another involves comprehending that the other lacks salient knowledge one possesses. Attributing psychological states with contents which do not match reality is also seen as more difficult. TT pictures the psychological states being attributed to others as states of knowledge. The false belief test also suggests that when one attributes a psychological state to another agent one is employing a representation of their psychological state as a starting point in order to draw inferences about their actions. From Wimmer and Perner we can also infer that one first has to form a clear representation of a distinct agent before one can assign psychological states. This is necessary so that we can confine ourselves to their beliefs in order to make predictions about how they will behave. Baron-Cohen's Sally-Anne test also suggests attributing psychological states to others requires isolation of one's own states while constructing judgements about the other agent and their situation. Baron-Cohen would suggest we attribute psychological states to others by grasping theoretical metarepresentational concepts and attributing these concepts to a subject. The Sally-Anne test is used as empirical support for the claim that we are isolating our own internal states when forming judgements about the psychological states of others.

Intertwining of questions 1 and 3

The methodology adopted in this PhD is to treat 1 and 3 as separate questions.

However the answers given by TT make this approach problematic and somewhat artificial as their treatment of these questions is often blurred. A defender of TT might object that this division of TT's answers to 1 and 3 is somewhat arbitrary. This is because there is an ambiguity to question 1. This arises because it is possible to offer an answer to question 1 such as the following: 'what we are doing when we attribute psychological states to others is processing representations'. This type of answer has reductionist implications. It means what we are really doing. It is only at a different level of description that we can be said to be utilising theories about inner states. If one accepts this reduction then an answer to 3 should reference whatever physiological activity is responsible for instantiating this representation processing. However it is easier to keep the contrast with ST in mind if we restrict the answer to question 1 to the one described in this chapter.

There are a number of points in the literature where it is hard to decipher which question is being addressed by TT in regards to 1 and 3 because the answers seem to spill into one another. Indeed TT is not focused on providing an empirically rigorous answer to 3 (e.g. in terms of underlying physiological processes) instead falling back on the personal level language and concepts utilised in its answer to question 1.For example Morton's suggestion that our TOM is learned and then internalised blurs questions 1 and 3. We learn to make personal level generalisations about the way others behave. These generalisations are then internalised and become automatic and sub-personal. In the same spirit Stich and Nichols claim our use of the dominant explanatory strategy may be largely unconscious.

It is also difficult to assess whether TT's claim we rely on an interconnected body of beliefs is being offered in answer to question 1 or 3. This depends on whether it is (in some sense) deliberately deployed or whether it is part of the background facilitating an attribution of a psychological state. Gopnik (like Churchland) claims we are explicitly using theories rather than simply relying on knowledge that can be construed theoretically so her work supports a reading of TT as answering question 1.

Leslie would claim to be offering an answer to question 3 by giving an account of the processes underlying mind reading. However, the claim we are making sense of behaviour in terms of mental states sounds more like an answer to question 1. Similarly Leslie claims that an understanding of agents and attitudes is fundamental to the task of making sense of others. By this he means to imply that psychological states are somehow modelled in the neurological processing.

We could also read Leslie's account as an answer to question1 rather than question 3. On this reading Leslie is claiming that what we are doing when we attribute psychological states to others is interpreting their behaviour in terms of mental states. If we take Leslie's distinction between metarepresentations and literal representations to be involved in an answer to question 1 we are back to the problem that this account will lack phenomenological plausibility as an account of what we do. We do not take ourselves to be processing a mix of representations and

metarepresentations in order to assign a (theory derived) metarepresentation of an inner state to another agent. In any case Leslie's account is presented as an answer to question 3; that we employ metarepresentations in order to attribute psychological states to others. Metarepresentations carry information about relations between the agent and their situation. This situation contains both real and hypothetical components. Leslie's account suggests the essential distinguishing feature of minded beings is that they relate themselves to hypothetical scenarios including future situations. The appeal made to the notion of metarepresentation in false-belief tests also supports an interpretation of them as involved in an answer to question 3 rather than question 1. If the answer to question 1 is that we are employing a TOM, the answer to question 3 will be that we do this by operating metarepresentational capacities.

Question 4

Question 4 was the question of how we could ever perform the seemingly miraculous task of correctly attributing psychological states to others. TT does not really address question 4 at all. It is not vulnerable to the problem facing the traditional argument from analogy. This is the problem how it can even be coherent to attribute psychological states to other people if the meaning of such states is fixed by introspection. This would mean that the meaning one attached to psychological states was exclusively confined to one's own experiences. It then becomes impossible to make sense of the idea of extending the application of these concepts

to other bodies as nothing could possibly ground such an extension. TT avoids this problem by claiming the meaning of psychological states is fixed by a theory whose apparatus (psychological concepts and interrelating laws) is in principle public. Even when introspection is involved it is theory laden. It implies a grasp of the fact that the theory is applicable to others as well as oneself. Indeed it is difficult for TT to even allow space for the distinctiveness of the problem facing the argument from analogy because it treats the task of attributing psychological states as just another branch of the general scientific quest for explanation and prediction.

However there are other problems associated with this manoeuvre. TT fixes the meaning of psychological states in a way that seems to make it difficult to capture the distinctiveness of first personal states. The theory theorist is forced to accept that there is no significant difference between the way one's own psychological states are revealed to one and the way the psychological states of others are revealed. But this fails to capture a distinctive relation we have to our own mental states. First person ascriptions are not the result of an inductive process as no such process is necessary. TT does not allow one to see what is so different about first and third-person cases and this is an important limitation on its usefulness as an explanation of what we are doing. A defender of TT could respond by arguing that the difference being appealed to is more apparent than real. On this view, the insight TT offers is that in fact all ascription is correctly classified as third-personal; it is simply a mistake to think there is anything distinctive about the first personal case. Alternatively a defender of TT could claim that the first-person/third-person asymmetry exists but is not a

difference that runs deep. It is merely an asymmetry of observation, not of description.

Aim C

An aim of this thesis was to evaluate the ways in which TT is related to the classic strategy the argument from analogy embodies for solving the other minds problem and how dependant TT is on the picture supporting this strategy. TT differs from the analogical strategy in that it does not begin from one's own case. As Morton observed there is little that is envisaged by TT as psychologically unique claimed to be involved in the attribution of mental states. This is an impersonal process rather than one based on experience. In fact one even relates to one's own mind as a theoretical entity in some versions of this account.

However, unlike many versions of TT, Leslie does allocate a privileged role to first personal experience in a way we will see is similar to the ST approach described in chapter three. Pretence is employed as a way of breaking away from immediate first personal attitudes to the situation. Pretence is therefore the first step towards adopting another agent's attitude to the situation.

There are other similarities between TT and the analogical strategy. Both treat the mental states of others as unobservable. TT posits laws linking observable behaviour to hidden mental states. Premack and Woodruff explicitly stated that mental states

were not directly observable and theory theorists would all agree. In TT minds are treated as theoretical entities. The view we have to rely on a system of inferences to psychological states presupposes that the states themselves are unobservable. Furthermore, while Theory Theorists see themselves as in the business of polishing their theory, they also have to acknowledge that it is by definition potentially revisable. This suggests that our contact with the phenomena under investigation is always less than fully direct.

The strategy adopted by TT differs from the strategy adopted in the argument from analogy in a significant way. In Morton's discussion of TT it is important to note that the theory user begins by observing patterns of behaviour in other agents before learning to apply these concepts to their own psychological machinations. This inference runs in the opposite way from the analogy user who is extending from his own case –Unlike Mill's arguer from analogy the theory theorist is not usually taking self knowledge for granted. In fact it is in some sense secondary to recognition of others. One identifies one's own states by matching them to observable patterns in the world. However, like the arguer from analogy these patterns are interpreted as signs of unobservable inner states.

One could dispute the idea found in Theory Theorists such as Gopnik that psychological terms really denote abstract entities. This involves disputing the coherence of a sharp contrast between psychological vocabulary and the theory-neutral vocabulary of mere mechanical movement in which the phenomena is

supposedly more naturally described. This would be the vocabulary that gives rise to the formulation of the problem of other minds because it takes the psychological out of the realm of the literal.

Summary

I will briefly summarize the main findings of this chapter before reviewing the philosophical implications. All versions of TT have a number of features in common. The primary purpose of the theory is to explain and predict behaviour. Theoretical reasoning utilizes an inferential process to best explanation. Our everyday understanding of the psychological consists of a rough and ready body of folk psychological knowledge. Psychological concepts only have meaning within this theory. The theory and its components are potentially revisable. A psychological episode is a particular instance of a general pattern. The question of how a theory is acquired is a central concern for many researchers in the field.

There are also a number of topics on which theory theorists disagree. One of these is whether we use a theory to understand our own mental states. They also disagree about the extent to which the theory resembles a robust scientific theory. Though the central issue concerns our ability to ascribe psychological states, researchers differ in their secondary concerns which include how the theory is acquired and how it is to be used to explain deficits in autistic subjects. Different answers have been given to these questions.

This chapter concentrated on the distinction between child scientist and modular versions of the theory. We saw that each of these was capable of further subdivisions. For example we examined a distinction between being innate and learned and, if learned, between being learned individually or through cultural assimilation. Distinctions between the ways a theory could be acquired were discussed in the context of child scientist theories. When examining modular versions of TT we examined distinctions that could be made regarding the extent to which a system could be said to be modular. It also examines Leslie's influential conception of a modular theory- based processing system for mind reading.

The chapter went on to look at empirical evidence for TT in the form of false belief tests. It also considered the implications theory theorists draw for the understanding of autism as a theoretical deficit.

Finally this chapter considered the extent to which TT was able to contribute an answer to the four questions raised in the introduction about how we know other minds and also offer an evaluation of it in terms of some other aims of the thesis.

3 The Simulation theory of Mind

Simulation Theory (ST) is a position formulated in response to TT. It claims that our attributions of psychological states to others are driven by a process rather than by a theory. It disputes the idea that a tacit theory underlies our psychological competence.[197] ST is the theory that we come to know the psychological states of others by simulating them in ourselves. Formulations of simulation theory can broadly be divided into personal level and sub-personal versions. According to personal level versions we consciously simulate being in another's situation in order to work out how they must feel and then attribute the results of this exercise to the other person. Sub-personal ST conceives of simulation as a process executed by neural mechanisms operating below the level of awareness. What is important is that we simulate the psychological states of others in ourselves. This distinguishes ST from its rival, TT.

As we go along we will find that the distinction between personal level and sub-personal level explanation becomes less clear-cut than it initially appears. Simulation theorists appear to treat it as unproblematic to flit back and forth between the two. ST's account of how we understand others has more in common with our everyday surface interpretation than TT's account. Whereas the TT advocate is committed to the view that we may be fundamentally mistaken about what we observe when confronting intentional behaviour due to the theoretical character of

---

[197] Gordon 2009

perceptual observation, ST sees itself as in the business of validating our use of everyday psychological concepts by providing firm empirical grounding for their neural underpinnings. There are similarities here between ST and the 'direct perception' position which chapter 5 will examine. However proponents of ST claim to offer strong empirical support for ST drawn not from the phenomenological level but from neurological data. This chapter will consider how the relation between this data and experience should be understood and whether such data can really be used to support ST over TT.

Early Formulations

The best way to get a grip on ST will be to examine its origins and development. Two philosophers, Gordon and Heal, developed similar theories at the same time and this chapter will begin by examining these.

1. Gordon's concept of simulation

One of the earliest formulations of ST was proposed by Gordon. Interestingly Gordon begins by examining our ability to predict our own immediate behaviour before considering the question of how we understand others.

To see why he begins here it is useful to understand the aspects of TT Gordon was opposing with ST. The main point of disagreement is with the TT view that all predictions of human behaviour are inferences from theoretical premises about

beliefs desires and emotions along with laws connecting them with behaviour[198].

Gordon takes this to be the TT position. His alternative position suggests that the

real work is done by practical simulation supplemented with a capacity for pretend

play.[199]

Gordon argues that, when it comes to making predictions about one's own

immediate intentions; this is obviously not done by making inferences from

theoretical premises. Immediate self-prediction is also of interest because of its

exceptional accuracy. We are rarely wrong in predicting what we are immediately

about to do. If such predictions were derived from the TT model of inference they

would be much less reliable. It would also be very difficult to account for the

confidence we place in these predictions.[200] This is because folk psychology

specifies only the probable or typical effects of psychological states. Atypical

actions would resist prediction.[201]

Although not based on nomological reasoning, declarations of immediate intention

are often products of practical reasoning. This is reasoning which underlies a

decision to act. They form a bridge between practical deliberation and prediction.

This opens up the possibility of using simulated practical reasoning as a predictive

device.[202] It is then easy to see how this capacity for self-prediction could be

---

[198] Gordon 1986 P.61
[199] Ibid P.71
[200] Ibid P.61
[201] Ibid
[202] Ibid P.62

127

extended to cover hypothetical situations.[203] This requires engaging in a kind of pretend play where the hypothetical situation actually obtains while other factors remain (as far as is possible) constant. One imagines oneself in these modified conditions and asks what one will do next. One answers this question with a (hypothetical) declaration of immediate intention. The behavioural output is suppressed.

From this we can progress towards envisioning our ability to predict the behaviour of others by employing simulation. The first step is to imagine that the hypothetical situation represents another person's situation rather than a hypothetical situation for oneself. This allows one to predict what one would oneself do in that situation.[204] Finally one must make adjustments for relevant differences between oneself and the other person.[205]

Gordon emphasised the importance of hypothesis testing and experimentation in practical simulation.[206] One does not only wish to understand another person's behaviour but to predict the form it will take in the course of one's interactions with that person. Steps towards accomplishing this can include making hypothetical shifts in spatiotemporal perspectives and in hypothetical roles.[207]

---

[203] Ibid
[204] Ibid P.63
[205] Ibid
[206] Ibid P.64
[207] Ibid

128

One can have competing hypotheses about another person which require further testing. These can be tested through subsequent interactions. This testing should continue until one is able to predict the other reasonably reliably "forming a fairly stable pretend-world for that person".[208] In forming hypotheses about another person one should obey "the principle of least pretending".[209] The pretend world should resemble the real world as closely as possible. Where people share a common stock of facts and values pretence can be kept to a minimum whereas alien cultures may initially require a higher degree of pretence.

Instead of an ability to understand another's belief in propositional format such as 'X believes that P' one needs only to be able to simulate 'that P'. One makes this assertion within the context of practical simulation rather than making reference to a general folk psychological law.[210]

Gordon claims there is empirical support for this view. He appeals to the same evidence that TT supporters appealed to in the last chapter. Gordon offers an alternative interpretation of the results of Wimmer and Perner's false belief task discussed in the previous chapter.[211] The upshot of this experiment, according to Gordon, is that at four to five years of age the child develops the ability to make

---

[208] Ibid P.65
[209] Ibid
[210] Ibid P.68
[211] See Wimmer and Perner 1983

129

allowances for what the other (Maxi) isn't in a position to know.[212] Prior to this all predictions are made in an egocentric way based on facts the subject knows about.

If, Gordon reasons, 'believes' is a theoretical term whose meaning is fixed by the set of generalizations in which it occurs then mastery of this concept will be a matter of internalizing a sufficient number of these generalisations. If this were the correct account, Gordon argues, we should expect that prior to internalizing these generalisations a child will be unable to predict or explain human action.[213] We would also expect that after internalising these generalisations the child could deal equally well with actions caused by true beliefs and actions caused by false beliefs. Alternatively if at around four years the child acquires the ability to make assertions within the context of a practical simulation this would enable the child to overcome the egocentric confinement to actual facts. We should then expect the change Wimmer and Perner discovered.[214]

Gordon also appropriates Baron-Cohen et al's experimental data discussed in the previous chapter. He argues his account of belief predicts that only children who can engage in pretend play can master the concept of beliefs.[215] It is well established that autistic children lack this capacity. Therefore these experimental findings accord with Gordon's position.

---

[212] Gordon 1986 P.69
[213] Ibid
[214] Ibid P.70
[215] Ibid

Gordon's finishes by speculating that the self-reporting discussed above may only be the tip of the iceberg.[216] Once it is acquired, a capacity for practical simulation may operate primarily at a sub-verbal level without our explicit awareness. Our practical reasoning system may be run 'off line'. This involves disengaging it from its natural inputs and feeding in pretend input. It also involves disengaging the system from its natural output systems so that instead of executing a decision we end up with an anticipation of the other's behaviour. This may be an unconscious motor anticipation of the other's behaviour.[217] These are themes that will be taken up later in this chapter.

It is worth noticing that in this simulationist account, as in the TT account, there is a developmental progression involving discrete stages.[218]

Summary

Gordon's version of simulation suggests that, to understand the minds of others, one begins from one's own case. Like Leslie in the previous chapter, Gordon suggests that first we extend from actual to hypothetical cases. Then we extend from simulating hypothetical cases of one's own experience to simulating hypothetical experiences of another agent. Simulating hypothetical states of one's own experience and simulating states of another agent are treated as somehow parallel and the development of the former ability is somehow meant to facilitate the

---

[216] Ibid
[217] Ibid
[218] See Davies & stone 1995b p.5-6

131

development of the latter ability. Simulation is also intended to down grade the role of inference in understanding others which is central in TT. Simulation is meant to be a practical test rather than a theoretical strategy.

Gordon's suggestion that hypothesis testing plays a central role in understanding others is implausible as a phenomenological level characterisation of the process although it has some plausibility in marginal cases such as his example of understanding agents from alien cultures. This raises a question about what level ST is supposed to be offering a characterisation of. After all, ST is supposed to be an account of what enables everyday understanding of others and not just novel cases. Continuity between personal and sub-personal levels of explanation is not treated as problematic on this account.

Gordon appeals to Wimmer and Perner's false belief task in favour of ST, evidence previously utilised as evidence for TT. Gordon argues the documented results signal the development of an ability rather than of a new theory. He also offers arguments against the TT interpretation of the evidence.

2.Heal's concept of Replication

At the same time as Gordon introduced the term simulation into the mind-reading debate, Heal introduced a very similar notion of replication. Heal contrasts what she calls the functional strategy with the replicative strategy. To begin with it will be interesting to examine her characterisation of the functional strategy as she has TT

in mind. The functionalist strategy rests on the assumption that explanation of

psychological states or actions via reference to beliefs, desires or emotions is a

causal process. Heal argues this approach is resolutely third-personal.[219] People are

viewed as "complex objects in our environment whose behaviour we wish to

anticipate but whose causal innards we cannot perceive".[220] We are supposed to

causally anticipate this behaviour on the basis of a composite body of folk

psychological theory. Heal notes that such a theory will necessarily be enormously

complex.[221] She argues we should be reluctant to credit ourselves with implicit

knowledge of such a theory unless no alternative account is available.

Heal next presents her alternative. She argues her replicationist account is the more

economical of the two accounts.[222] Like Gordon's, her account also begins by

describing what happens in our own case. One thinks about the world in the context

of making decisions and forming opinions. The future is complex and opaque. This

gives rise to the need to conceive of possible but non-actual states of affairs. One is

able to imagine how one's aims and opinions may change. Heal argues that this

ability also makes it possible to gain a certain sort of understanding of other

people.[223] One can bring to bear this theoretical knowledge of the world together

with these imaginative abilities to gain insight into other people "without any further

elaborate theorizing about them".[224] This does require making one assumption

---

[219] Heal (1986) P.45
[220] Ibid
[221] Ibid P.47
[222] Ibid P.51
[223] Ibid P.47
[224] Ibid

however; other people are like oneself in the sense of being thinkers in possession of the same fundamental cognitive capacities and propensities.[225]

By way of example, Heal explains what happens when one wants to predict another person's action. She claims that this requires that one must attempt to replicate his thinking. To do this, she further suggests, one first attempts to place oneself in the other's initial state by imagining the world as it appears from their point of view. Next one deliberates to see what decision emerges.[226] Heal claims that similar methods would also apply to working out another person's thoughts feelings or past intentions.

Heal continues to develop her account of replication by considering potential objections to this account. The first objection is that the ability to replicate another person already requires possession of a theory about the interrelations of psychological states and behaviour. One will make this objection if one reasons that replication demands that, on the basis of observation of another, one is able to recognise what psychological state that person is in and then put oneself in that state. Heal responds that such an objection makes a mistake about the direction of gaze of the replicator.[227] The replicator is not focused on the target subject but on the world around that subject. It is a matter of what this world makes the subject think which forms the basis for the beliefs one attributes to the subject.[228]

---

[225] Ibid
[226] Ibid
[227] Ibid P.48
[228] Ibid

To replicate one has to re-centre the world in imagination. For example this will involve appeal to theoretical principles such as visual occlusion in order to know what it is possible for the target to perceive.[229] However this theoretical knowledge is not to be equated with theoretical knowledge about how psychological states relate to one another. Modifications to the core replication process will need to allow for different personalities, styles of thinking and so on.[230]

An objection to simulation/replication is that for a simulation to work the simulator must deploy a body of psychological knowledge.[231] For example Dennett argued that the state one simulates (or replicates) is not a real belief but a make-believe belief.[232] However Heal denies that make-believe belief should be understood as a theoretical state. It involves imagining, which is something we already do on our own behalf.[233] The sequence of thoughts connecting an imagined state of affairs to an imagined decision parallels the sequence of thought connecting a real belief to a real decision. All that is required is an ability to distinguish real belief from entertaining a possibility and the ability to attribute a belief one generated in oneself to another person.[234]

Heal argues replicationist and functionalist strategies must be understood as mutually exclusive.[235] In particular she argues against the view that replication could

[229] Ibid
[230] Ibid
[231] Davies & Stone 1995b P.18
[232] Dennet 1987 P.100
[233] Heal 1986 P.49
[234] Ibid P.50
[235] Ibid P.51

be understood as a primitive strategy that will give way to more sophisticated functionalist theory. Heal argues a requirement of this view is the assumption that rationality can be given a complete formal definition in terms of syntactically specifiable inference rules. She explains that the view replication could be succeeded by functionalist theory might seem plausible if we think that we are initially unable to anticipate another person's action without replicating their thoughts because we do not have access to the thought itself or to the respects in which oneself and others are similar.[236] On this view reflection subsequently persuades us that the thoughts themselves possess some intrinsic character. We learn to deduce some non-demonstrative specification of relevant similarity. Hereafter when we use psychological terminology it refers to these features. However Heal does not accept this account. Heal argues closer analysis of the notion of relevant similarity will show that it is resistant to the functionalist paradigm. She also argues that our access to another person's thought continues to be indirect and demonstrative.

Heal argues that the difference between functional explanation and replication is that the latter makes psychological explanation central. Psychological explanation is concerned with rendering the thoughts or behaviour of another person intelligible. This involves showing them to be exercises of cognitive competence or rationality motivated by purposes and reasons. When one comes to reflect on the replication strategy one employs, the theoretical context one places it in will be that of cognitive

---

[236] Ibid P.52

competence. This involves "trying to get things right"[237] or achieve "success in judgment".[238] This feature will present itself as the respect in which self and others are similar. Cognitive competence, Heal claims, cannot be cashed out in terms of inference rules or judgment-forming procedures. This is thought to mark a crucial incompatibility between replication and functionalism. Inference procedures or judgments cannot be arranged in a clear hierarchy with some procedures identified as fundamental.[239]

Summary

One motivation that Heal had for proposing replication theory was to oppose the resolutely third-personal nature of TT. She argued the theoretical process postulated by TT is implausibly complex. It is not clear if the point is just that science should value simplicity in its theories (a valid point in itself) or also a stronger point that it is implausible that we must undertake a theoretical process of this nature (be it at a personal or sub-personal level) in order to make sense of others.

Like Gordon, Heal suggests the process of understanding others begins with our own case. Again it involves a move from actual to hypothetical experience and a further move from this to the experiences of others. Again simulating hypothetical states of one's own experience and simulating states of another agent are in some sense

---

[237] Ibid
[238] Ibid P.53
[239] Ibid

analogous and the development of the former ability is again meant to facilitate the development of the latter ability.

Heal suggests replication requires a theoretical assumption of similarity between oneself and other agents. However she argues that replicationist (or simulationist) and functionalist (or theory based) strategies must be defined in a way that makes them mutually exclusive. She argues that the notion of intelligibility required in making sense of others is resistant to capture in theoretical terminology.

General characteristics of ST

1.Mind-reading

Like TT, ST is usually interested in explaining our mind-reading abilities.[240] Goldman defines mind reading as "the attribution of a mental state to self or other".[241] To mind-read one must "form a judgment, belief, or representation that a designated person occupies or undergoes (in the past, present, or future) a specified mental state or experience".[242] To count as an act of mind-reading this belief or judgment must have as its object a psychological state rather than a piece of behaviour: "The state attributed must be mental rather than merely behavioural; otherwise the attribution doesn't qualify as *mind*reading".[243]

---

[240] It is worth noting that some simulationists such as Gordon (2009) argue that simulation processes may facilitate processes other than concept dependant mind-reading.
[241] Goldman 2009 P.312
[242] Ibid
[243] Goldman 2009b P.2

Mind-reading is usually distinguished from behaviour-reading, a form of explanation that does not make reference to psychological concepts. For example, Susan Hurley explains the contrast as follows: "After all, it might be said, all we ever 'really observe' is behaviour in environments; we infer mental states from this. However, mind-readers do not merely keep track of the behaviour of other agents, but also understand other agents in terms of their mental states".[244] In the case of ST; mind-reading begins with "the mind-reader taking someone else's perspective and generating pretend mental or behavioural states that match the other person's. These are ... used as inputs to the simulator's own psychological processes, including decision-making processes, while these are held 'off-line', producing simulated mental states and behaviour as output. The simulated outputs are then assigned to the other person".[245]

2. Representation

A key concept operating in the mind-reading schema is representation. Gordon suggests that the resources the brain uses to guide one's own behaviour are "modified to operate as representations of other people".[246] Mind-reading is defined by Gallese and Goldman (in a paper we will consider shortly) as the activity of representing specific psychological states of others.[247] These states include perceptions, beliefs, goals and expectations. Gallese and Goldman argue that this requires a system of representation known as folk psychology. They claim that MNs

---

[244] Hurley 2006 P.18. See also Gallese 2007a
[245] ibid P.20
[246] Gordon 2009
[247] Gallese & Goldman 1998 P.495

are a part of our folk psychologising mechanism.[248] More recently Goldman has argued that "to attribute a mental state to an individual is to represent that individual as being in that state".[249] In the case of simulation theory Gallese & Goldman suggest that "other people's mental states are represented by adopting their perspective: by tracking or matching their states with resonant states of one's own".[250]

3. Simulation is off –line

Representing the psychological states of others is different to directly perceiving these states. Hurley argues that "the simulator copies the states of the other and uses the copies in her own decision-making equipment".[251] The simulation producing the representation of another person's psychological state is something that takes place 'off line': "simulation can be regarded as off-line copying. This enables you to regard yourself and others as similar, to identify with others, and to understand the motivation of others' actions in a means/ends structured way".[252]

Simulation is described as offline because it involves the use of one's own decision making processes or mechanisms. When these processes/mechanisms are applied to one's own situation they can be thought of as 'online'. However, when the purpose is to figure out how someone else will feel or behave, rather than to feel or behave that way oneself, these abilities must be taken offline. For example, Hurley

---

[248] ibid P.495
[249] Goldman 2009b P.2
[250] Gallese & Goldman 1998 P.493
[251] Hurley 2006 P.20
[252] ibid P.19

characterises simulation as "the off-line use of practical abilities....When I use practical reason off-line in mind-reading...I activate my own normative and deliberative dispositions".[253]

Currie and Ravenscroft hold a similar view: We possess a decision-maker for choosing a plan of action on the basis of our beliefs and desires. ST posits a mechanism for manufacturing 'pretend' beliefs and desires and feeding them to the decision-maker. ST also posits a mechanism for taking the decision-maker 'off-line' thereby "directing its output away from the systems responsible for behaviour".[254]

The term offline comes from computer science. It usually signifies that a computer is not connected to other computational devices or networks. In the present context the point of calling simulation offline is that all processing occurs in the individual brain without interacting with anything external. Because the simulation has been taken offline it does not result in the production of overt behaviour in the simulator (for example, if the output that results from simulating an event as if it were happening to the agent is anger, this simulation does not produce an angry reaction in the simulator. Instead an angry reaction is predicted of the simulated target). Because the process is offline it requires we first make *copies* of the states of the target to use in the simulation. Simulation then transforms these copies of actual states of the target to produce hypothetical future states of the target.

---

[253] Ibid P.22
[254] Currie and Ravenscroft 1997 P.162

Simulation and cognitive science

In the above accounts from Gordon and Heal simulation takes place at the personal

level (although Gordon suggested such accounts may only be the tip of the iceberg).

In the cognitive sciences the term simulation usually signifies an automatic and

unconscious process. This process may underlie personal level simulation but may

also take place in the absence of any personal level awareness. Recall that according

to Gordon and Heal we employ a personal level ability to think about our own

situation while making modifications to this process in order to account for relevant

differences between ourselves and the person we intend to simulate. One way to

describe this would be to say we use our own situation as a model which we

manipulate in order to use it as a model of another person's situation. In cognitive

scientific accounts this model is understood in terms of the neurological processing

and equipment that underlies our ability. The view that if ST is correct we should

expect to find empirical evidence of its neurological underpinnings precedes the

search for actual empirical data to be cited as evidence. However the development of

Implicit ST as a theory is in fact closely intertwined with the particular

neuroscientific data that supporters of the theory have cited as evidence of the truth

of their theory.

The neural machinery of Simulation

The neural machinery can roughly be divided into two key components:

The first component is the Mirror Neuron System. Mirror neurons (MNs) fire in one

of two conditions: when performing a particular action/undergoing a particular emotion and when observing the equivalent performance in another person. The second component of the neural machinery underlying implicit simulation has been termed the Who System. This system has the role of attributing a psychological state to the self or to a target.

The MN system and sub-personal ST

The link between MNs and simulation based mindreading was first jointly proposed in a paper by Gallese and Goldman in 1998. MNs had recently been discovered in area F5 in the premotor cortex of macaque monkeys.[255] This area controls hand and mouth movements.[256] Mirror neurons of Macaque monkeys were shown to discharge not only when actively engaged in performing actions such as grasping objects, but also when observing other individuals performing similar actions.[257] Gallese and Goldman were not arguing that MNs constitute a full-scale realisation of the simulation heuristic but that they may represent a primitive version, or a precursor of a simulation heuristic underlying mind-reading.[258]

Evolution

Gallese and Goldman argued that the best way to get to grips with our ability to understand others is to adopt an evolutionary frame of reference capable of building

---

[255] Gallese & Goldman 1998 P.493
[256] Umiltà, et al 2001 P.155
[257] Gallese & Goldman 1998 P.493
[258] Ibid P.498

on the results from different fields of investigation including neurophysiology and developmental psychology.[259] Mind-reading could serve an evolutionary advantage by enhancing our overall fitness for coping with the environment. Detecting another agent's goals or psychological states could help an observer understand and anticipate the agent's forthcoming actions and enable the observer to make appropriate adjustments to his own responses.[260] Gallese and Goldman suggest that the capacity to mind-read others involves the ability to employ a simulation routine. They further argue that our ability to simulate is likely to have evolved from an action execution/observation matching system instantiated by MNs.[261] Gallese and Goldman speculate there is cognitive continuity in the domain of intentional-state attribution from non-human primates to humans, and that the neural underpinnings of this continuity are provided by MNs.[262] This continuity is grounded in the shared ability of humans and primates to detect goals in the behaviour of members of their fellow species. This capacity to understand the goal of an action relies on a process matching the observed behaviour to the observer's action plans.[263]

Goal related

MNs discharge during specific goal-related motor acts.[264] Their activity is correlated with specific hand and mouth acts. Activation is not correlated with the execution of individual movements, such as the contractions of a particular muscle group.   A

---

[259] Ibid P.493
[260] Ibid PP.495-496
[261] Ibid P.493
[262] Ibid P.500
[263] Ibid
[264] Ibid P.495

144

motor act is distinguished from mere movement by the presence of a goal.[265] Gallese

and Goldman argue this distinction allows the motor system to be interpreted as

instantiating psychological states such as purpose or intention rather than just being

involved in the control of movement. The most effective visual triggers of MN

activity were actions where the experimenter or another monkey interacts with an

object using either their hand or their mouth. Grasping, holding and manipulating

motor acts were shown to be the most effective at triggering MN responses.[266] It was

also established that neither the object alone, nor the agent alone, evoke a MN

response. Furthermore mimicking the action without an object and performing the

action using tools also fail to evoke a MN response.[267] There was a tight correlation

between observed action and motor response not only in terms of the general goal of

the action, such as grasping an object but also in terms of the style of execution. For

example MNs are also selective in terms of the type of grip that is being

employed.[268]

MNs in Human beings

Gallese and Goldman also appeal to a growing body of evidence for the existence of

equivalent systems in humans. This evidence divides into two strands. The first

strand appeals to studies using transcranic magnetic stimulation (TMS).

The observation of actions was shown to activate the premotor cortex in monkeys. If

this also happens in humans, it is reasoned, this activation should elicit enhanced

---

[265] Ibid P.493
[266] Ibid P.495
[267] Ibid
[268] Ibid

motor evoked potentials induced by the application of TMS to the motor cortex, because this has strong anatomical links to premotor areas.[269] This was shown to be the case.[270] This TMS evidence also provides a further link between MNs & ST

The experimental findings were that whenever we observed someone performing an action, the same motor circuits became active that are active when we ourselves performed that action.[271]

The second strand of evidence comes from brain imaging experiments. These have been used to establish the anatomical location of MNs in human beings. In these studies the experimental subject observed another person grasping objects.[272] The results of these studies implicated the cortex of the left superior temporal sulcus, the left inferior parietal lobule and the anterior part of Broca's region in this type of observation.[273]

Empirical evidence and The TOM debate

MNs are hypothesised to underlie the process of mind reading, or serve as precursors to such a process (Gallese favours the idea MN activity directly constitutes an act of mind reading whereas Goldman thinks MN activity causally contributes to mind reading but does not fully explain it).[274] The discovery of MNs is thought to score points for ST over TT. Gallese and Goldman lay out what they understand to be the

---

[269] Ibid
[270] See also Fadiga et al 1995
[271] Gallese & Goldman 1998 P.495
[272] Ibid
[273] Ibid
[274] Goldman 2009b PP.5-6

key characteristics of each theory. They argue that it is central to the TT philosophy that mindreading is depicted as a 'detached' theoretical activity.[275] Mindreading is accomplished by appeal to a TOM which is analogous to a scientific theory. The psychological states one attributes to others are understood as unobservable theoretical posits. These are appealed to in order to explain and predict behaviour in the same fashion a physicist appeals to electrons and quarks to explain the behaviour of entities in her chosen field.[276] TOM is a functionalist model comprised of a set of causal/explanatory laws relating external stimuli to inner states and behaviour.[277] The attribution of a psychological state to another person results from theoretical reasoning that utilises tacitly known causal laws.

ST depicts mind reading as involving the attempt to mimic psychological states of the other. This theory is partly based on doubts that mind readers even tacitly represent the type of causal/explanatory laws posited by TT. Instead people use their own psychological mechanisms to calculate and predict the psychological processes of others. First one creates *pretend* desires, preferences, and beliefs of the type one takes the other person to have in oneself.[278] Next these pretend preferences and beliefs are fed into one's own decision-making mechanism, which outputs a pretend decision. Finally this decision is taken 'off-line' and used to predict the other person's behaviour.[279] Simulation can also be employed to retrospectively determine what psychological states have already occurred in another agent by working

[275] Gallese & Goldman 1998 P.497
[276] Ibid P.496
[277] Ibid
[278] Ibid
[279] Ibid

147

backwards from the observed action.[280] Gallese and Goldman point out that, for simulation to work, the pretend beliefs and desires must be sufficiently similar to the genuine states.[281] In the case of pretending to see and pretending to act, Gallese and Goldman claim, a correspondence has already been established between pretend and natural psychological states.[282] They also argue there is informal evidence that other pretend states including beliefs and desires functionally resemble their genuine counterparts. Gallese and Goldman also claim simulation involves matching the psychological activity of the simulator with that of the observed agent.[283]

Gallese and Goldman argue that the predictions these rival theories will make about mind reading allow for the possibility of empirical discrimination. We saw that TT does not predict we will employ pretend states that mimic those of the agent we are observing. Furthermore TT does not predict we will make use of our own decision-making system to arrive at a prediction. On the other hand, ST hypothesizes that mind reading involves an attempt to mimic the psychological activity of the observed agent. Empirical evidence of psychological mimicry during mindreading would therefore offer a way of discriminating between ST and TT. This evidence would adhere well with ST but would not be predicted by TT. Gallese and Goldman claim MN activity creates in the mind reader a state that matches that of the observed agent. TT, unlike ST, has no reason to predict the mimicking activity of

---

[280] Ibid P.497
[281] Ibid
[282] Ibid
[283] Ibid

MNs. This is more than just theoretical inference.[284] It is possible to interpret MN activity as constituting a plan to execute a particular action.[285] When activated by observation of another agent MNs execute the same action. In these cases the action plan is tagged as belonging to the other agent and subsequently inhibited.[286] Gallese and Goldman argue this is nature's way of stepping into the other's shoes as ST predicts we should.[287] Gallese and Goldman were not claiming MNs constitute a full-scale instantiation of the simulation routine, especially not in monkeys. Their argument was that MNs carry out a primitive version, or perhaps a precursor of a simulation routine that might underlie mind reading.

The TMS experiments described above provide another link between MN activity and ST. These experiments showed that mirroring systems in humans facilitate activation in the same muscle groups that are being employed by the observed agent. This supports the idea that when one observes another agent performing some action, a neural event occurs in oneself which is 'qualitatively the same as an event that triggers actual movement in the observed agent'.[288] TT would not predict that the same muscle groups would be facilitated in the observer and in the observed agent; a mind reader represents another agent's behaviour in purely theoretical terms. ST, on the other hand, predicts psychological occurrences in the mind reader will be analogous to psychological occurrences in the observed agent. Gallese and Goldman claim that although simulation is said to be taken off line, evidence of

---

[284] Ibid P.498
[285] Ibid P.497
[286] Ibid
[287] Ibid
[288] Ibid P.498

149

muscular activity at this level is still compatible with ST and even lends support to the theory. The behaviour of patients with prefrontal lesions who suffer from a clinical phenomenon known as 'imitation behaviour' can be accounted for within the ST framework as an impairment of the inhibitory mechanisms. Gallese and Goldman also suggest that we can infer from this that normal humans, observing another agent performing an action will generate a plan to perform the same action, but that this plan will be inhibited.[289]

Since the publication of Gallese and Goldman's paper there have been two major experimental attempts to link MN activity in monkeys with the coding of others intentions.[290] The first was carried out by Umiltà et al in 2001. Their experiment set out to show that MNs are not simply matching performed and observed actions. A subset of MNs becomes active even when part of the action is occluded and must therefore be inferred.[291] In their experiment a macaque monkey was placed in two experimental conditions. In the first 'full vision condition' the monkey observes an experimenter grasping an object. In this condition MNs have been shown to fire. In the second 'hidden condition' the monkey is first shown an object being placed behind an opaque screen. While the object is hidden by the screen the monkey observes a human experimenter reaching behind the screen towards the object. Despite the fact the object is hidden MNs still fire.[292] The importance of the presence of the object was affirmed in parallel full vision conditions and hidden

---

[289] Ibid P.499
[290] Iacoboni 2009 P.124
[291] Umiltà et al 2001 P.155
[292] Ibid P.156

conditions where the monkey first has a chance to observe that there is no object behind the screen. MN activation does not occur in either set of conditions.

Umiltà et al suggest action understanding could be based on a mechanism that triggers a motor representation of the action. For this to be successful in cases where the object is occluded the monkey must know there is an object behind the screen and must observe the experimenter's hand disappearing behind the screen.[293] They argue MNs are able to generate a motor representation of an observed action not only when the monkey views the action but also when it knows the outcome without seeing the most crucial part.[294] To selectively respond in hidden conditions MNS must infer and represent both the occluded action and the occluded object.[295] Umiltà et al conclude that even where visual information is limited MNs can "place the observer in the same internal state as when actively executing the same action".[296]

Iacoboni argues that this experiment does not yet fully establish that MNs code the intentions of others. Although MNs do not respond to miming actions this only establishes that they can distinguish between grasping and non-grasping movements. However a grasping movement can form a part of very different intentions.[297] The experiment does not address the question of whether MNs distinguish between grasping, for example, to eat and grasping to place the object elsewhere. This issue was subsequently addressed in an experimental study by Fogassi et al in 2005. The

[293] Ibid P.160
[294] Ibid
[295] Ibid P.161
[296] Ibid
[297] Iacoboni 2009 P.124

study by Fogassi et al examined parietal MNs located in the inferior parietal lobule in macaque monkeys. These neurons discharge both when the monkey performs a motor act and when it observes another performing a similar motor act.[298] The study by Fogassi et al demonstrated that neurons coding a specific grasping act displayed different activations when this act formed part of different actions.[299] Similarly when observing actions performed by others neurons responded differentially when the act was embedded in a different action.

Fogassi et al studied parietal neurons that were active in association with grasping movements of the hand in two main conditions: In condition one a food item was brought to the mouth. In condition two an identical item was placed in a container. A third condition which was a variant of condition two was also added in which the item was placed in a container near to the monkey's mouth. This was to control for kinematic differences between the main conditions. This condition required the same arm flexion, as condition one. A small percentage of the neurons discharged with equal strength regardless of the motor act that followed the grasping. However the majority (about two thirds of the neurons studied) were influenced by the subsequent motor act.[300] The neurons that were most responsive in condition one were less responsive in conditions two and three.[301] All neurons that were most responsive in grasping to place actions displayed the same selectivity regardless of the kinematic differences between condition two and three. Fogassi et al argued that this finding

[298] Fogassi et al 2005 P.662
[299] Ibid P.664
[300] Ibid P.662
[301] Ibid P.663

152

demonstrates that the crucial factor determining intensity of discharge is goal rather than kinematics.

Fogassi et al next introduced a visual task in which the monkey observed an experimenter perform the same actions that were involved in the motor task. Once again it was found that the majority of neurons were differentially responsive in regard to the overall action.[302] The majority of the neurons Fogassi et al tested exhibited the same specificity during grasping observation and grasping execution.[303]

Fogassi et al note that, from an engineering perspective, it may look uneconomical for MNs to code identical motor acts differently depending on the final outcome. It might seem more desirable to have multipurpose neurons for grasping that can be employed whenever needed. However this objection would fail to take into account the fluidity with which different motor acts follow one another.[304] Fogassi et al argue a 'kinetic melody' depends on the different motor acts that form an action being tightly linked so that the execution can occur without gaps. Motor acts, though not related to one another independently of the global action, form prewired intentional chains.[305] Every individual motor act in these chains is facilitated by the previous act. The experimental evidence also favours a chain linking neurons coding subsequent motor acts. For example neurons activated by passive flexion of the

[302] Ibid PP.663-664
[303] Ibid P.664
[304] Ibid P.665
[305] Ibid

forearm also have tactile receptive fields located on the mouth, some of which also respond during grasping actions made with the mouth. When the monkey touches or grasps an object using its hand these neurons appear to facilitate the opening of the mouth.[306] This is what we should expect as when someone initiates a motor act they usually have in mind the overall goal of the action to which it belongs. The intention is set before the start of the movement and is already reflected in the initial motor act.[307]

This study is seen as providing further support for the belief that the fundamental role of MNs is to allow an observer to understand the goal of a motor act.[308] Fogassi et al argue it is because the monkey knows the goal of a motor act it executes that it is able to recognise the goal of a motor act it observes another individual performing. The observation triggers activation of the same neurons active during execution of the action.[309] Fogassi et al's study builds on this by showing MNs are able to discriminate between identical motor acts by appeal to the actions in which they are embedded. The individual motor acts form part of a chain leading to the actions final goal. This enables the monkey "to predict the goal of the observed action and, thus, to ''read'' the intention of the acting individual".[310] The observer will have an internal representation of what the overall action is most likely to be dependent on which motor chain has been activated. Fogassi et al admit that it is more difficult to specify how the correct motor chain becomes activated as this is

---

[306] Ibid
[307] Ibid
[308] Ibid
[309] Ibid
[310] Ibid P.666

154

likely to be influenced by a number of factors. These include the context in which the motor act is initiated and the type of object grasped. Repetition of experience can either strengthen or weaken particular motor chains.

The who system

The second component of the neural machinery underlying implicit simulation has been termed the 'who system'. This system executes the final step in the implicit simulation process; that of attributing a psychological state to the self or to a target. MNs are claimed to manufacture representations of inner psychological states. These representations subsequently have to be attributed to an agent. Before examining this component it will be useful to consider some philosophical considerations informing the research which supports it.

The problem of self-other discrimination

Jeannerod and Pacherie are also interested in knowledge of one's own psychological states. Self knowledge has a different status in their version of ST. Jeannerod and Pacherie think simulation is required not only to recognise the actions and intentions of others but also to recognise our own actions and intentions. They claim that, traditionally, philosophers have not taken the idea there could be a problem of self-other discrimination seriously.[311] This is mainly because the problem of self identification has not been taken seriously.[312] The main reason for not taking the

---

[311] Jeannerod and Pacherie 2004 P.114
[312] Ibid PP.112-113

problem seriously is that some judgments are thought to be identification free. While

it is obvious that when making judgments such as 'I think X', one may be mistaken

about X it is often thought, following Descartes, that one cannot be mistaken about

who the subject of the thought is.[313]

The Cartesian outlook envisions an asymmetry between the way we know our own

psychological states and the way we know the psychological states of other people.

An acceptance of this asymmetry prevents the self-other distinction from arising as a

philosophical issue.[314] While philosophical doubts can be raised about the existence

of an external world, and about the existence of other minds, the asymmetry of

knowledge ensures we can be certain about our own psychological states.[315] While it

is possible to make mistakes when identifying the self as an object, such errors are

not possible for the self as a subject.[316] Unfortunately this picture makes any claim

to be able to observe the psychological states of others philosophically problematic.

The Cartesian position entails a particular picture of the relation between body and

mind. Bodies are understood as mere material objects. The behaviour of bodies is by

definition non-intentional.[317] No properties intrinsic to a body allow distinctions to

be drawn between mere bodily happenings and intentional behaviour. It is only the

inner psychological accompaniments and causes that differentiate the two.[318]

Psychological states on the other hand are internal states with properties which are

---

[313] Ibid P.113
[314] Ibid P.114
[315] Ibid P.113
[316] Ibid P.116
[317] Ibid P.137
[318] Ibid

independent of external properties.[319] Their existence however must always be a matter of inference, as these properties can never be directly observed.[320] The difference between behaviour and bodily movement is a matter of the inner psychological accompaniments and precede a piece of behaviour.[321] However as these accompaniments are unobservable it is always possible that any inference to their existence will be incorrect.[322] In the Cartesian picture, the gap between psychological states and physical behaviour is both ontological and logical.[323]

Jeannerod and Pacherie argue that this Cartesian picture raises the problem of other minds. How can we claim to reliably know about the psychological states of others if these are something we never observe? We have no non-inferential evidence to make us think other people could even be the subject of psychological states or processes. This theme will be taken up in more detail later in this chapter. Jeannerod and Pacherie argue that this framework even makes the meaning of psychological state terms problematic; how can these terms and concepts be univocal if totally separate criteria are involved in the application of these terms to other people that the ones we use to apply them to ourselves?[324]

Jeannerod and Pacherie suggest reasons why the problem of self identification should be taken seriously. They claim allegiance with an alternative philosophiance

---

[319] Ibid P.115
[320] Ibid
[321] Ibid
[322] Ibid
[323] Ibid P.138
[324] Ibid P.114

picture associated with philosophers including Wittgenstein and McDowell. This

position rejects the dualist framework underlying the Cartesian position.

Wittgenstein argued that our criteria for the application of psychological terms must

be public.[325] McDowell recommends we should take the concept of a 'human being'

as a basic unity.[326] This allows us to think of behaviour and mentality as much more

integrated. This has the consequence that the psychological states of other people

can be objects of experience in suitable conditions.[327] Jeannerod and Pacherie argue,

contrary to the Cartesian picture, that intentional bodily behaviour is not intrinsically

indistinguishable from non-intentional bodily behaviour.[328] However it is worth

stressing an important difference between the account being advocated here and the

one offered by Wittgenstein. What we are sensitive to is psychological

representations. However we are about to see that these are not psychological

representations as ordinarily understood. Though we have access to psychological

states these are not yet psychological states of agents. These representations are still

understood as representations of inner psychological states which are hidden and

require simulation. This differs from a position like Wittgenstein's where it is

psychological states of people themselves which inform our perception thus negating

the need for causal inferences to inner states.

The focus is on actions and intentions. Georgieff and Jeannerod argue that action is a

central means of communication between individuals and that determining the agent

---

[325] Ibid
[326] McDowell, 1982, p.470
[327] Jeannerod and Pacherie 2004 P.115
[328] Ibid P.137

158

of an action facilitates in differentiating between the self and others.[329] Observing

actions and their effects provides clues for both understanding their meaning and

attributing them to an agent.[330] Jeannerod and Pacherie argue that intentions, at least

in some circumstances, are transparent in others' behaviour. Their existence may be

directly observed and does not require inference from behavioural cues.[331] They

examine the links between perception and action, and present empirical evidence

that intentions are perceived. As Georgieff and Jeannerod explain, action here is

understood not just in terms of the overt appearance of observable muscular

movements but also in terms of covert internal representations of a goal and of the

means to achieve it.[332] Georgieff and Jeannerod postulate a close relationship

between the overt and covert aspects of an action. For example mentally simulating

movement of the right hand, increases brain activity in several areas directly

involved in motor behaviour.[333] These two aspects of an action stand in an

asymmetrical relationship as overt actions necessarily involve covert components

whereas covert actions do not necessarily involve overt components.[334]

Against the Cartesian picture, Jeannerod and Pacherie argue that self-ascription of

agency is not identification-free.[335] They claim that the problem of self-identification

does arise. There is a problem that requires solving, both for the self as object and

for the self as subject. Jeannerod and Pacherie take the issue of self-other

[329] Georgieff and Jeannerod 1998 P.466
[330] Ibid P.467
[331] Jeannerod and Pacherie 2004 P.115
[332] Georgieff and Jeannerod 1998 P.465
[333] Ibid P.467
[334] Ibid P.465
[335] Jeannerod and Pacherie 2004 P.113

discrimination and, by extension the issue of self identification, for the self as object to be uncontroversial.[336] They argue that issues involving identification of self as object and self as subject are often interdependent and misidentification is possible in both cases.[337] For the problem to arise for the self (or other) as subject, three conditions must obtain. The first condition is that the intentions of others must be directly perceivable.[338] If we could only be non-inferentially aware of our own intentions, there would not be a problem of self-other discrimination. For this problem to arise we must further be aware of both our intentions and those of others in the same way.[339] This might make it seem like there is no need for a simulation process. However they are not directly perceivable as the intentions of others and simulation is required for this. In both our own experience and observation we simulate a psychological process. Then we assign the psychological output. But in our own case it seems odd to call it simulation.

This leads us to the second condition that must obtain. It is that we should be primarily aware of intentions, be they our intentions or other peoples, as unattributed or naked intentions.[340] There are two ways in which an intention might be naked: an intention could have an impersonal representation which implicitly specifies only an action plus a goal or it could be personal coding an agent plus an action plus a goal but leaving the agent's identity unspecified.[341] For example when Mary observes

---

[336] Ibid P.116
[337] Ibid
[338] Ibid P.115
[339] Ibid P.138
[340] Ibid P.115
[341] Ibid P.138

John opening the door she would be primarily aware of an intention to open the door rather than being aware of John's intention to open the door. The same should also be true when it is Mary herself who intends to open the door. Jeannerod and Pacherie claim it is possible to be aware of an intention, without knowing whose intention we are aware of.[342]

The third condition that must obtain for there to be a problem of self-ascription is that the channels providing the collateral information used to determine the owner of an intention are not infallible or a form of identification-free first-person knowledge.[343] The source of this collateral information will depend on whether the intention yields an overt action or whether it merely yields a covert response. In the overt case the information is to be found in the outside world. To discover if an intention is one's own one looks to see if it is one's body that is performing the action. There is empirical evidence that this methodology is not totally infallible.[344] In the case of covert actions evidence concerning schizophrenic subjects shows that the attribution mechanisms employed are not immune to error.[345]

Shared representation

Jeannerod and Pacherie interpret MN activation as giving rise to 'shared' representations of psychological states. Our shared representations of actions allow

---

[342] Ibid P.139
[343] Ibid P.116
[344] Ibid P.139
[345] Ibid

us to perceive the intentions of others. The actions of others are represented to the same extent as one's own actions.[346] Georgieff and Jeannerod argue that the fact we find partly overlapping cortical activation when representing one's own action and when observing an action performed by another agent suggests that the same representation can be shared between different agents.[347] The notion of a shared representation carries implications for questions both about how one can be aware of one's own actions and about one's ability to distinguish these actions from the actions of others.[348]

Jeannerod and Pacherie claim that the existence of shared representations provides evidence for the existence of naked intentions.[349] The appeal to evidence that brain areas active when representing a self-produced action and brain areas active when observing another person performing the same action partially overlap.[350] This includes the motor brain. By shared representations they mean that psychological states involved in actions, including not yet executed actions, and psychological states involved in observed actions appear to share neural representations. Evidence from monkeys demonstrates that MNs are activated in the performance of a specific motor action, and also when witnessing the same action performed by another agent.[351] Jeannerod and Pacherie interpret these findings as evidence that MNs represent a particular action type irrespective of agent. Evidence for shared

---

[346] Ibid P.130
[347] Georgieff and Jeannerod 1998 P.468
[348] Ibid P.468
[349] Jeannerod and Pacherie 2004 P.139
[350] Ibid P.130
[351] Ibid

representations in humans comes from neuroimaging experiments demonstrating the existence of a cortical network comprised of sections of the inferior parietal lobule, the ventral premotor area, and the SMA which show activation both during intending and imagining actions on the one hand and observing them performed by other people on the other.[352]

An interesting feature of the shared representation capacity is that the representations of others can influence the observer's action system.[353] Georgieff and Jeannerod speculate that this phenomenon may underlie our ability for empathy as well as explaining delusions of alien control (either of being able to control others or being controlled by them) experienced by schizophrenic subjects. These delusions could be due to improper monitoring of shared representations.

Disentangling shared representations (the who system)

MN activation must be understood as one component of a larger system. The burden of attribution of psychological states does not fall squarely on to the MN system. The purpose of the who system becomes clear when we frame MN activation within the wider neural context. Although MNs afford us shared representations of the psychological and emotional states of ourselves and others this sharing is only partial. It is the non-shared component of a representation that facilitates its attribution to self or other.

---

[352] Ibid P.131
[353] Georgieff and Jeannerod 1998 P.474

163

According to Jeannerod and Pacherie, because both self and other representations

exist at the same brain location, they need to be disentangled. Georgieff and

Jeannerod note that understanding action implies the existence of representations

that are common to several people.[354] They also argue that a mechanism is therefore

required to assign these representations to the correct agent. They argue that this

involves the processing of specific signals at the level of these representations.[355]

Georgieff and Jeannerod point out that, as well as revealing shared activation of

MNs, studies of premotor neurons in monkeys also revealed a difference in cortical

activation during performing and observing conditions.[356] Whereas all premotor

neurons coding self-produced movement are activated during performance of action,

activation is confined to MNs during observation of action.[357] Alongside the

common network responsible for shared representation described earlier, there are

separate non-overlapping cortical zones whose activation is restricted to a specific

condition; self or other produced action.[358] These comprise a network involving

other premotor neurons including canonical neurons. This network only discharges

when a monkey performs an action and not when it observes another agent

performing it. Working in conjunction these networks allow attribution of shared

representations to other agents. Jeannerod and Pacherie argue it is a critical feature

of shared representations that these representations only partially overlap. 'The non-

overlapping part of a shared representation serves as a cue for attributing the action

---

[354] Ibid P.468
[355] Ibid
[356] Ibid P.472
[357] Ibid PP.472-473
[358] Jeannerod and Pacherie 2004 PP.130-131

164

either to the self or to another. As Georgieff and Jeannerod explain, activation of those areas overlapping self-produced and observed action (therefore common to all individuals) provides the conditions that allow this system to interpret the MN activation as representing an observed action. Simultaneous activation of non-overlapping areas (areas activated only by self-produced activity), by contrast, facilitate the system's interpretation of the MN activation as representing a self-produced action.[359] It is the role of the Who System to disentangle these overlapping representations.

This picture has philosophical ramifications for the Cartesian picture discussed above which opened a gulf between the direct way in which we access our own psychological states and the indirect way in which we infer the psychological states of others. The mechanisms involved in self identification must take into account the existence of other agents.[360] This is the case even with covert actions. Attribution mechanisms cannot operate in a solipsistic fashion.[361]

Intentional action

Jeannerod and Pacherie argue a simulation process underlies action recognition and attribution.[362] Both overt and covert actions are centrally simulated by our neural

---

[359] Georgieff and Jeannerod 1998 P.475
[360] Jeannerod and Pacherie 2004 P.130
[361] Ibid P.131
[362] ibid P.113

networks.[363] When it comes to understanding action they argue simulation is the default procedure though it may be supplemented or overridden by theoretical processes.[364] They also argue it is at the root of our mind reading abilities which they speculate are likely to involve a fusion of explicit and implicit simulation processes, with sub-personal neural simulation forming the basis of explicit psychological simulation.[365]

Jeannerod and Pacherie argue that adopting a simulation account of action involves three commitments[366]: Firstly intentional bodily behaviour must possess intrinsic characteristics which distinguish this behaviour from other types of physical movement. We must be able to distinguish intentional and non-intentional behaviour in order to identify plausible targets for simulation. The second condition for simulation is that we must be perceptually sensitive to these characteristics. Thirdly Jeannerod and Pacherie suggest simulation requires that, when undertaking an action and when observing an action performed by another agent, the same processes and mechanisms should be involved. They agree with Gallese and Goldman that observation of another's action should activate the same mechanisms in oneself that would be active if the action was intended by oneself.[367] They claim that there is now robust empirical evidence that this is the case. This includes evidence that implicit knowledge of the motor rules involved in producing an action influence

---

[363] Ibid P.124
[364] Ibid P.128
[365] Ibid
[366] Ibid
[367] Ibid

166

visual perception of action.[368] The visual system exploits this knowledge not only to recognise actions but also to recognise intentions. They also argue that a reason why motor images and observed actions are best understood as covert simulated actions is because these events both activate brain areas concerned with preparation of action.[369]

Embodied Simulation

Gallese formulated the thesis of embodied simulation (E.S.). This is one of the clearest examples of sub-personal ST. E.S. is a "modelling mechanism" postulated to allow us to model the behaviour of others as intentional experiences. This functional mechanism is automatic and unconscious. It is a mandatory process and it is prereflexive. Together with the mechanisms underpinning it, E.S. is claimed to allow us to understand the minds of others and to provide a firm biological grounding for this understanding.

This mechanism is triggered by the observation of another person. It is a simulation of the other's action, intention, emotion or sensation. Most importantly E.S. is argued to be a fundamental biological mechanism facilitating empathy as well as for facilitating a more general understanding of another person's mind.[370] In postulating this mechanism Gallese does not wish to deny that more explicit cognitive evaluation of social stimuli may also have a role to play in facilitating an

---

[368] Ibid P.129
[369] Ibid
[370] Gallese et al (2007) P.132

167

understanding of the minds of others.[371] However he is claiming that, at least in relatively simple cases, the ascription of intentions occurs by default.[372] Consequently Gallese et al claim that the other's emotion is "constituted experienced and therefore directly understood by means of an embodied simulation".[373] This embodied simulation results in a shared bodily state.[374] It is an intersubjective process. E.S. is the utilisation of implicit models of others' behaviour and experience.[375] The actions, emotions and sensations of others activate our own internal representations of the body states associated with these stimuli.

Though E.S. is a brand of Simulation Theory, Gallese is careful to distinguish his position from the standard version. It is worth spelling out the contrasts as a way of getting a grip on what E.S. is and what it is not. 'Standard' simulation, according to Gallese et al, is a deliberate and therefore explicit process in which the observer imaginatively generates the appropriate psychological states, were he or she to find themself in the situation of the person under observation.[376] The psychological states of the observed person are then inferred from this simulation. Gallese's alternative model is intended to call into question the need for such an explicit deliberation or interpretation (in denying a central role to introspection E.S. is also of course opposed to Theory Theory). E.S. must be understood as prior to any explicit theorising that may occur. E.S. also differs from mainstream cognitive science in

---

[371] Freedberg and Gallese 2007 P.198
[372] Gallese et al (2007) P.136
[373] Ibid P.144
[374] Ibid
[375] Freedberg and Gallese 2007 P.198
[376] Gallese et al (2007) P.143

168

that it treats action prediction and intentional ascription as much more closely related; in that they share a functional mechanism.[377]

Embodied Simulation is the externally triggered activation of the same neural networks underpinning our first-personal experiences.[378] The discovery of MNs, along with other neurons possessing similar properties, has a key role in the formulation of E.S. MNs discharge not only when performing goal-related hand actions such as grasping objects, but also when observing other individuals performing similar actions. The activations of these dual-action neurons, when observing others, are interpreted as literal (embodied) simulations of the observer's motor system. Actions/expressions belonging to, or closely related to, the observer's repertoire are mapped onto the observer's motor system.[379] This means that mere observation can trigger excitation of the motor system. The mechanisms of E.S. are involved in predicting, as well as comprehending behaviour. Depending on which motor chain is activated by an observation, the observer will activate the motor schema of what the agent is most likely to do next.[380] It is hypothesised that this mechanism is formed using statistical detection of those actions which frequently follow other actions, this knowledge being acquired through observation and habitual performance. Populations of MNs can be chained to code the motor act along with those actions that would normally follow from it in a given context.

---

[377] Ibid P.137
[378] Ibid P.142
[379] Ibid P.135
[380] Ibid P.137

Gallese et al claim that there is a wide variety of evidence for the existence of these mechanisms. This includes evidence for a process parallel to the observation of motor actions that occurs when observing other peoples' facial expressions.[381] For example, in experiments conducted by Dimberg, subjects shown pictures of emotional facial expressions, displayed spontaneous and rapid electromyographic responses in the facial muscles that correspond to the facial muscles involved in the pictured person's facial expressions (happy and angry faces were used). In studies of disgust, the same neural structure (the anterior insula) and overlapping location is also found to be active in the experience and observation of disgust.[382] Though this simulation is usually unconscious it can have a small conscious effect, for example the simulation of another's emotional facial expression can be accompanied by the experience of a small dose of the emotion simulated.[383]

Gallese et al also consider a potential difficulty for the E.S. hypothesis: why if we all have a MN system and possess the capacity for E.S., do we find such a range of individual differences in our capacity to understand others?[384] They suggest that severe deficits in intersubjective understanding may be due to faults in the individual's mirror neuron system or to a disruptive emotional-affective regulation of the system.[385] For example, subjects with autism have difficulty with imitation including difficulty imitating and recognising facial gestures.[386] Gallese et al argue

---

[381] Ibid P.141
[382] Ibid
[383] Ibid P.148
[384] Ibid P.153
[385] Ibid P.155
[386] Ibid P.154

that these difficulties result from the incapacity to establish a motor equivalence between the demonstrator and the imitator.[387] Autistic subjects do not show activation of the M.N. system. The imitation deficits can be characterised as examples of defective embodied simulation. Subjects with ASD have been shown not to show automatic mimicry of the facial expression of basic emotions. Gallese et al also suggest an individual's particular psychological attitude could be a key variable determining the degree and quality of the activation of shared neural circuits (They use the example of a sensation; pain).[388] A variety of other factors including culture age and gender differences are also said to influence our capacity to understand others.[389]

Attunement

These considerations about the deficits that result from flaws in the MN system lead us to consider the benefits that result from an intact MN system capable of modelling E.S. In particular it brings us to consider the role E.S. plays in intersubjective understanding. E.S. is at the foundation of our ability to reach what Gallese terms attunement with others. The term 'attunement' is put forward by Gallese as a replacement to 'mirroring'.[390] The understanding of others involved in attunement takes us beyond straightforward simulation of their psychological states. From what has been said so far it might seem that the MN system's task is simply to duplicate the behaviour we observe. However Gallese et al warn the reader that the

---

[387] Ibid P.155
[388] Ibid P.142
[389] Ibid P.153
[390] Ibid P.152

term mirroring can be misleading if it is understood too literally. What is needed for intersubjective communication is not the ability to copy the other but to produce responses that are congruent with her expressive behaviour. A response can be attuned in a number of different ways; for example it can serve to complement or to modulate the other's behaviour.[391] In some cases literal duplication might be the most congruent response to the other, for example returning their smile, but this is only one type of attuned response and certainly not the most prevalent one. Furthermore what we have in these cases is not literal mirroring, insofar as such responses are deliberate. In fact Gallese et al suggest that automatic simulations are often accompanied by inhibitory mechanisms which enable one to observe behaviour without imitating it.[392] The simulation is suppressed, keeping it below the action threshold. Imitation therefore is not the primary function of E.S. It is not even clear that the idea of literal mirroring would be coherent. An E.S. is filtered through one's past experiences, capacities and attitudes so is never a literal replication. What is important in the context of empathic attunement is that the E.S. is an accurate enough replication of the others psychological state to allow one to produce a response that is helpful or appropriate.[393]

Gallese et al illustrate these points with the example of infant-mother interactions. If the mother were to cry when the baby cried this would not really be an attuned response at all. It would reflect contagion rather than attunement.[394] It would not

---

[391] Ibid
[392] Ibid P.151
[393] Ibid
[394] Ibid

contribute anything useful to an intersubjective infant-mother relationship. After all, much more is demanded in such a relationship. It is a very important relationship in that it is within the intersubjective boundaries such a relationship affords the infant that he or she is able to form an understanding both of their own individual mind and of the minds of others.[395] Attuned responses will facilitate and regulate this understanding in a way that imitative ones are unlikely to.

It will be useful to be as clear as possible on the relationship between E.S. and attunement. Gallese et al suggest that the observation of another elicits an automatic simulation, and that this mechanism enables empathic understanding which can eventually lead to attuned responses.[396] They also claim the MN system and E.S. constitute the biological basis for attunement.[397] However Gallese et al also argue that the mere existence of a mirror system, while necessary for attunement is not sufficient. One reason they give for this is because of the range of individual differences we have been discussing.

We are now in a position to see that the deficits in empathic understanding discussed above in relation to autism can be characterised as attunement deficits. In taking this line Gallese et al distinguish their position from conventional thinking on this topic. These deficits are widely held to be deficits in a TOM module.[398] Gallese et al argue their alternative position fits the available data much better. (The orthodox view is

---

[395] Ibid PP.151-152
[396] Ibid P.151
[397] Ibid P.152
[398] Ibid P.155

that the basic deficit is in the capacity to theorize but this, Gallese et al suggest, sits

uneasily with high functioning ASD individuals who rely on theorising as a

compensatory strategy. What autistic subjects appear to be compensating for is the

loss of a more direct experientially given understanding of others).[399]

It can be seen that the basic concepts of E.S. and attunement are argued to contribute

to our knowledge of how we understand others, and also have implications for our

understanding of autism and of child development. They have also been argued to

have implications for two further areas; aesthetic appreciation and psycho analysis.

Gallese's development of the concepts of E.S. and attunement are partially

explicated through discussion of their application to these fields.

Aesthetic appreciation

E.S. is here extended to our appreciation of artefacts as well as expressions, gestures

and actions. Once again the theory of E.S. is seen as presenting a challenge to

mainstream ideology, which has tended to privilege cognitive responses to works of

art to the exclusion of embodied responses.[400] Most recent art history and criticism

has neglected the role of emotional responses. The discovery of MNs has been

argued to enable us to understand the basis of our physical and emotional responses

to works of art and to images in general. Various forms of E.S. can be seen to play a

role in aesthetic appreciation. Freedberg and Gallese argue that E.S. supports the

ideas of thinkers such as Merleau-Ponty who drew our attention to our sense of

---

[399] Ibid P.156
[400] Freedberg & Gallese 2007 P.198

physical involvement in paintings and sculptures and to the possibilities of "felt

bodily imitations of the implied actions of the artist".[401] This physical involvement is

manifest in at least two ways: firstly in a sense of imitating the motion seen or

implied in the work and secondly in the enhanced emotional responses to the

artwork.

The mechanisms of E.S. are claimed to facilitate a bodily empathy. Often this takes

the form of a felt activation of the muscles that appear to be activated within the

artwork. For example these felt responses correspond to the parts of the body that

are "threatened, pressured, constrained or destabilised".[402] Freedberg and Gallese

use the example of empathy for pain. Viewing images of damaged body parts

activates part of the same brain regions activated in our own sensations of pain. This

can account not only for our own physical sensation but for the corresponding

feeling of shock. Freedberg and Gallese suggest that this physical empathy easily

converts into emotional empathy for the way the body is damaged. The same holds

for emotional response where we might automatically simulate the depicted

emotional expression.

The actions, intentions, objects, emotions and sensations represented in a work of art

can all initiate an E.S. In particular there is a lot of evidence for the activation of

emotional states in response to depictions of both facial and bodily expression.[403]

---

[401] Ibid P.198
[402] Ibid P.197
[403] Ibid P.201

E.S also allows the reconstruction of action from observation of static images.[404]

This can result in a simulation of an action or gesture depicted in the work or even of

an action indirectly implied by the painting where for example the final stage of the

action is hidden.[405] There seems to be for Freedberg and Gallese another sense in

which an artwork can elicit an E.S. which is that it is recognised as something man

made. The marks on the canvas for example are the visible traces of goal directed

movements.[406]

Freedberg and Gallese stress that E.S. is not offered in order to exclude historical,

cultural or contextual factors from playing a significant role in explaining aesthetic

appreciation. Neither are they claiming that the activation of MNs is sufficient for

aesthetic appreciation.[407] They explain their methodology: "we 'bracket' the artistic

dimension of visual works of art and focus on the embodied phenomena that are

induced in the course of contemplating such works by virtue of their visual

content".[408] E.S. is only being considered as the neural mechanism underpinning

certain aspects of this process. Freedberg and Gallese suggest that these mechanisms

may be modified by a variety of contextual factors be they historical social, cultural

or personal.[409] Nonetheless we should not understate the importance of this

discovery. For the first time, they argue, it allows us to understand the, frequently

noted but imperfectly understood, feelings of physical reaction we have to

---

[404] Ibid P.202
[405] Ibid P.198
[406] Ibid P.202
[407] Gallese and Freedberg 2007 P.411
[408] Freedberg and Gallese 2007 P.197
[409] Ibid P.199

artworks.[410] These mechanisms are crucial insofar as they constitute a basic level of automatic empathic response to artworks.[411] Aesthetic appreciation cannot be fully envisaged without considering MNs and their role in embodied and empathetic responses.[412] They allow an automatic and direct experiential understanding of the intentional and emotional contents of images.[413] This understanding is precognitive.

<u>Contextual sensitivity</u>

In the above discussion of intentional understanding there appears to be an attempt to show how a contextual sensitivity can be built into our automatic processing. The context of the discussion is how the activation of MNs can facilitate an appreciation of an intentional action rather than a series of movements. Some MNs are context sensitive in the sense that they are only activated in response to a movement, for example a grasping action, if it occurs in a certain context. If an agent grasps an object to bring it to his mouth this will cause different neurons to discharge in the observer than if he goes on to place that object in a cup.[414] Identical motor acts can be discriminated according to the context in which they are embedded. Depending on which of these parietal mirror neurons discharge, a different motor schema is activated.[415] In this way the observer is able to predict the next action. These motor pathways may be built up by past experience. The ability to ascribe intentions then

---

[410] Ibid
[411] Ibid P.202
[412] Gallese and Freedberg 2007 P.411
[413] Freedberg and Gallese 2007 P.202
[414] Gallese et al 2007 P.136
[415] Ibid P.137

becomes the ability to predict their goal. All the situation-specific knowledge the subject needs is built in.

E.S. and theoretical activity

Gallese et al claim E.S. works in conjunction with more theoretical strategies.[416] It is only meant to account for simple intentions. E.S. is not being offered as a monolithic account of how we understand others. Gallese is often at pains to point out that E.S. is not being offered as the full explanation of our capacities for understanding others or aesthetic appreciation. However he is claiming that it can account for much more than we realise. He also claims it is prior to any theorising processing that may also be involved.

Empathy

Gallese claims that E.S. is a fundamental biological mechanism that facilitates empathy as well as a more general understanding of another person's mind.[417] Gallese understands empathy as "the capacity to experience what others do and yet to attribute these shared experiences to others and not to the self".[418] He seeks to broaden the use of the term beyond the psychology of emotions to incorporate sensations and inner imitation of others.[419] In his view the concept of empathy should be extended to accommodate and account for all different aspects of expressive behaviour enabling us to establish a meaningful link between others and

---

[416] Ibid P.144
[417] Gallese 2007 P.148
[418] Gallese 2008 P.773
[419] Ibid P.774

ourselves.[420] This is because Gallese thinks that both an overview of the philosophical development of the term empathy and an examination of the neurological mechanisms underlying empathy suggest the restriction of the term to the domain of the emotions is arbitrary and reductive. Gallese explains that the concept of empathy was originally created as a way to capture the affinity we have with the objects of aesthetic experience and only later extended into the domain of intersubjectivity.[421]

A fundamental connotation of empathy is that one appreciates a similarity between oneself and the other person.[422] This similarity stems partly from the common experience of action we can share.[423] Gallese recognises that empathy is primarily a concept belonging to the phenomenological level.[424] However he believes that the neural level has a key role in bringing about this phenomenological experience.

We-centric space

According to Gallese "interpersonal relations are established when a full-blown self-conscious subject of experience is not yet constituted".[425] Gallese employs a notion of "we-centric space" which helps to establish interpersonal relations prior to the development of a fully self-conscious subject. This common space facilitates our capacity to share the meaning of actions, intentions, feelings, and emotions with

---

[420] Gallese 2003 PP.176-177
[421] Ibid P.175
[422] Gallese 2007 P.147
[423] Gallese 2003 P.176
[424] Gallese et al 2007 P.158
[425] Gallese 2006 P.16

others. This capacity grounds the sense of familiarity we experience towards

others.[426] Gallese argues this we-centric common space is sub-personally

instantiated.[427] Neural mechanisms enable the shareable character of actions,

emotions, and sensations.[428] The sensations and emotions that others display can be

'empathized', and therefore *implicitly* understood, through a mirror matching

mechanism.[429] This intersubjective we-centric space provides the individual with a

powerful tool to overcome the epistemic gap between him/herself and the other,

produced by the establishment of a self-centred perspective.[430] We have an

attenuated intersubjectivity before developing our subjectivity.

The problem of undifferentiated representations

According to Gallese the original epistemic problem we must all solve is one of

conceptually separating ourselves off from others rather than establishing that other

minds exist. The 'epistemic gap' between inner experiential knowledge and outward

observation confronting Mill does not emerge until after we have the tools to cross

it. Before Mill's problem can be formulated we already have an understanding of

others as more than mere objects: "the 'selfness' we readily attribute to others, the

inner feeling of 'being-like-you' triggered by our encounter with others, are the

result of the shared we-centric space".[431] ES is proposed to be the foundation for this

we-centric space. ES therefore is best understood as an attempt to narrow the

---

[426] Gallese 2007 P.147
[427] Ibid P.148
[428] Ibid
[429] Gallese 2003 P.176
[430] Gallese 2006 P.16
[431] Ibid P.16

epistemological gap between self and other rather than to vault it. Representations of psychological states begin as undifferentiated with regard to self or other and have then to be subsequently assigned. The key problem that the sub-personal simulationist tries to tackle is how we disentangle these undifferentiated representations in order to assign them either to simulator or target rather than the problem of how to access information about the targets psychological states.

How does Simulation theory address the key questions identified in the introduction to this thesis?

Question 1

Question 1 in the introduction asked what we are doing when we attribute psychological states to others. ST agrees with TT that we are engaging in an act of mind reading; exercising an ability to attribute inner psychological states to the self or to another agent. The purpose is to explain and make predictions concerning their behaviour. Mind reading involves making a judgement that an agent is in a particular psychological state. The point that simulation theorists wish to stress is that we are engaging in a process rather than an intellectual task.

ST's answer is that we are simulating other agent's psychological processes using our own psychological apparatus then attributing the psychological output of this

processing to another agent rather than to ourselves. The psychological output is used to represent another person's inner psychological state. We are exercising an ability to place ourselves in another agent's situation. This is claimed to be the same ability we employ to work out how we would feel in hypothetical scenarios. In the case of simulation, this process is understood as a kind of pretence. Our psychological experiences are accorded a representational status. We are also making allowances for differences in the others knowledge and beliefs. We are simulating or replicating the psychological processes of another agent which have resulted in their current psychological state. ST views the process of understanding another agent's psychological state as a component in a wider process of predicting how one's intersubjective encounter with this agent will develop. Explanation and prediction are interwoven in ST. They would claim that all understanding of psychological states occurs within the context of prediction of how social interaction will develop. It is an experimental strategy. For some Simulation theorists we are employing the same strategy we employ to grasp our own psychological states.

In offering an alternative account of what we are doing when we attribute psychological states, ST is also implying an alternative account of what these inner psychological states amount to. ST understands the notion of psychological states differently to TT. Though they share the view that they are inner states causally producing behaviour, for ST psychological states are revealed by introspection. Psychological states are not theoretical states which gain their meaning from the position they occupy within a theoretical or functional framework. At least from a first person point of view they are observable. They are therefore not thought of as

182

in principle revisable. Their meaning is not exhausted by their role within a scientific theory.

Problems with this answer

There is a parallel objection to the one made to TT to be made here. If taken as a depiction of a personal level process ST lacks phenomenological plausibility. It only seems to be a strategy we employ in unusual cases of ambiguity. It is certainly not a default methodology we are aware of using. This explains why it is usually not posited as a personal level process.

Another problem is that the meaning of our psychological concepts is anchored in first personal experience. This is also a key problem facing the argument from analogy. ST manages to avoid the problem facing TT that it ignores the $1^{st}/3^{rd}$ person asymmetry of our psychological states only to raise another problem; how properties whose meaning is given by first personal experience can coherently be attributed to others.

Some sub-personal formulations of ST avoid this difficulty by stating that the psychological representations employed in a simulation routine are neither first nor third-personal. However the idea of unattached representations which are subsequently assigned either to oneself or to another agent is controversial. Again it seems highly counter intuitive to suppose we could grasp a psychological state with

no intrinsic ties to a psychological agent. At the personal level our psychological states are reflexive. When one experiences them one experiences them as one's own. The self is implicated in the experience. It is incoherent in such cases to question who to assign the psychological state to. This picture of unattached psychological representations also differs significantly from the standard simulation process this chapter began with and could not be correctly described as its sub-personal analogue.

Question 2

Question 2 asked what justifies claims we make about the psychological states of others. ST has something to offer as to how the theory as a whole is justified. It is justified, according to ST, by similarity in the psychological makeup of ourselves and other agents. For ST to be regarded as a coherent method for correctly assigning psychological states to others there must be an assumption that one's own psychological processes and those of other agents function in a sufficiently similar manner. Simulated states must be sufficiently like the states they simulate. In the case of sub-personal versions of ST there is also some empirical justification for assuming this similarity in the form of neurological evidence for MNs. However the more pertinent question is how particular attributions of psychological states to others are justified. Here it seems they are justified through similarity between the agent one seeks to simulate and oneself.

As we saw with TT, justification is not a prime concern for many theorists working in TOM. An exception is the work of Jane Heal. Heal's view was that simulation is justified to the extent that it allows us to see another agent's behaviour as the rationally intelligible outcome of psychological processing. It is difficult to be sure how representative this view is of modern simulation theory however. Heal's view seems to be in tension with the idea that simulation is a process which takes place without recourse to anything external to the simulation itself. This would suggest that justification is internal to the simulation process whereas Heal is suggesting we need to take account of the process as an event in the wider world. We are justified in attributing psychological states to the extent they make intelligible a subjects response to their environment. It is agent plus environment that is the focus of the simulator's attention.

Problems with this answer

ST faces a problem also facing the original argument from analogy; it has to assume similarity between ourselves and others. However it is very difficult to support this assumption when working with an understanding of psychological states as something we grasp from inner experience. If ST is to be understood as offering a justification it will be a justification based on inference from a single case and therefore highly unpersuasive. However this criticism may not be totally fair as ST does not present itself as being in the business of offering philosophical justifications for our psychological attributions. They could be understood as

instead making a more modest claim that these attributions are justified to the extent they enable us to make reliable explanations and predictions about others. If simulation accurately describes what we are doing then it appears to be reliable. We are able to interact with others in a meaningful way.

Question 3

Question 3 in this context asks how attribution of psychological states to other agents is achieved. ST addresses this question by offering empirical data that will support the answer they have given to question 1. Different versions of ST offer different answers to this question. There are two broad types of answer:

The versions of ST this chapter began by examining suggested attribution is achieved through employment of a personal level ability for psychological processing; combined with a capacity to make appropriate adjustments for relevant differences between oneself and another agent. Methodologically the confidence we have in our ability to make accurate self prediction is extended to making predictions about others. We are relying on a personal level capacity we have for predicting how we would feel in hypothetical situations and then using this capacity to work out how others are likely to feel. This is a capacity belonging to the imagination. It involves pretending that we inhabit another person's situation while making allowances for relevant differences between their psychological makeup and our own. We imaginatively adopt the position of another agent by attempting to

view the world from their perspective. The focus of the simulator is on the environment as it appears from the agent's perspective rather than on the agent. Adopting the others perspective involves imaginative re-centering of the self in order to accommodate such factors as differences in perspective which may affect which relevant facts about a situation the simulation process will utilise (e.g. if a subject is unaware of some relevant feature of the situation we cannot include this feature in our simulation of her psychological states). The adjustments we make include hypothetical adjustments to the perspective one has on the hypothetical situation and to the role one occupies within this situation. We then employ these manufactured states, which can include desires, preferences, and beliefs, as inputs for psychological processing including decision making. The next step is to suppress the behavioural output that this simulation inclines us to make. Instead we attribute the output to the other person. Successful simulation can be a gradual process involving trial and error. One may begin with competing hypothesises about how another is likely to feel or behave and further simulation and observation of the other may be required. When engaged in personal level simulation the hypothetical scenarios we construct should be as plausible as possible while accounting for the resulting behaviour.

Advocates of personal level simulation claim that it may be a personal level manifestation of a capacity that usually operates at a sub-personal level. This brings us to the second type of approach ST offers to question 3. Most contemporary formulations of ST claim attribution of psychological states to others is achieved

via a sub-personal analogue of this simulation process operating inside the brain. This process requires no conscious awareness. It is an automatic process triggered by observation of others. Our capacity for psychological experience is taken 'offline' by disengaging it from our personal experience and substituting hypothetical input depicting another agent's situation. This means that all processing takes place internally within the simulator without a need for continued interaction with external factors. The psychological processing must also be disengaged from its natural behavioural output so that we output a belief about the others current psychological state, or immanent behaviour, rather than an action. One's actual responses to the hypothetical scenario must be inhibited. Our own psychological processing capacities are instead employed as a model which we manipulate to simulate the situation and experiences of another agent. The behaviour of others is modeled as intentional experience. Simulation is able to get off the ground through a process of matching observed psychological states of the other to one's own states. The simulation process employs copies of these states created by one's own psychological processing faculties rather than the directly perceived states themselves. These copies are then transformed in to hypothetical states of another agent. Simulation employs practical rather than theoretical capacities of the simulator.

The model sub-personal simulation employs is cashed out in neurological terms. Simulation employs the neural networks which normally underlie first personal psychological experience. The simulation process involves the MN system acting in

conjunction with the 'who' system. MNs fire when observing other agents

psychological manifestations, and in doing so produce shared representations. MNs

represent psychological phenomena independently of any representation of the agent

experiencing these phenomena. These psychological manifestations include

expressions and intentionally coloured (for example goal-orientated) behaviour.

These automatically trigger MN activation. The Who System then assigns these

representations to other agents. It does this by relying on the non-shared aspect of

the psychological representation which, in the case of simulation, acts as a signal to

attribute the representation to another agent. This neurological activity is understood

to instantiate mind reading or as a precursor to the process. In the latter view, the

suggestion appears to be that familiar goal orientated action provides the easiest way

of matching the psychological states of others to one's own. Simulation may also

work retroactively to work out which psychological states resulted in observed

behaviour.

Problems with this answer

We have already made the point that simulation lacks plausibility as a default

personal level process enabling us to grasp the psychological states of others. The

idea that the process can be reallocated to the sub-personal level and still remain the

same sort of task is also problematic. On this picture sub-personal accounts assume

the answer to question 1 given by personal level simulation is correct. We use

ourselves to model the psychological processes of other agents. However they claim

the simulation process initially executed by the imagination is reassigned to off line sub-personal mechanisms.

Another problem with sub-personal accounts is that it is highly questionable whether they really provide evidence in support of a standard simulation process. Closer inspection of the postulated mechanisms of simulation reveals that they are not instantiating the simulation routine described in the answer to question 1. We are not employing first person capacities and then attributing them to another agent, Instead these accounts offer us a modified explanation of what we are doing when we assign psychological states; we are working with impersonal representations and solving a problem of allocation. It is not so clear that, in this sub-personal account, a simulator can take their own psychological states for granted, at least not as states of oneself. The asymmetry between first and third-personal states which motivates ST is undermined. In fact the idea that psychological representations overlap makes it difficult to explain the asymmetry between first and third-personal states revealed in experience

Furthermore it is not clear the idea of sub-personal simulation really makes sense in the first place. It is not clear what justifies our characterising firing MNs as psychological representations. It is unclear whether the neurological data is coherently described as a simulation routine at all. In fact, as well as the problem that the story being offered does not look like simulation, there is also a more general problem. It is not just that the data does not allow us to claim that we are

putting oneself in another agents place and making psychological attributions. This data does not provide any account of what is being done at the level of psychological description. While the data is informative about the contribution the neurological level makes to seeing others as psychological agents it does not tell us what we are doing. At the very least it seems unlikely that the neurological processes we have been discussing can take the whole burden of psychological understanding.

Question 4

Whereas question 4 was not a real issue for TT it is an issue for ST. This is because it is raised by the answer they give to question 1. Whereas in TT psychological states get their meaning from their positioning inside a theory, for ST the meaning is tied in with a first personal perspective. While this means ST is not vulnerable to the objection made against TT, that it ignores the distinctive asymmetry between first and third-personal psychological states, question 4 becomes very salient. If the psychological states ST is working with are inner states of oneself revealed through introspection then the question of how they could coherently be applied to others must be raised. The sub-personal account of ST complicates the issue somewhat because it can claim to be employing a third-personal account of psychological representations and offering an explanation of how these representations become assigned to agents. The coherence of such a claim has been challenged above.

<u>Aim C</u>

An aim of this thesis is to evaluate the ways in which TT and ST are related to the classic strategy the argument from analogy embodies for solving the other minds problem and how dependant both TT and ST are on the picture supporting this strategy. They are both dependent on the picture of the mind as inner and (at least in the case of other minds) unobservable. The case of ST is more problematic than TT. ST is often more explicit than TT about the way the theory relates to the traditional strategy for solving the other minds problem which is the argument by analogy. ST is often claimed by supporters to offer a solution to the classical philosophical problem of other minds. For example, Iacoboni sketches the problem of other minds in the following way:

"if I only have access to my own mind, which is a very private entity I can only access, how can I possibly understand the minds of other people? How can I possibly share the world with others, how can people possibly share their own mental states?"[432]

However proponents of ST seem to disagree about whether the problem of other minds is a genuine problem. On the one hand, standard ST matches the starting point of the argument from analogy; it begins from one's own case. In these formulations one's own mind is treated as observable but the minds of others are not. The traditional formulations of simulation that this chapter began with treat the problem

as genuine. Their methodology strongly resembles the strategy adopted in the argument from analogy. Both Gordon and Heal begin by taking for granted access to one's own psychological experience before attempting to extend access to other agent's psychological experience (via hypothetical cases of first personal experience). First personal access is privileged in both accounts. Our knowledge of the psychological states of others in contrast is claimed to be indirect. Other minds are thought of as hypothetical entities.

Sub-personal versions of ST, on the other hand, often explicitly reject the framework of the problem of other minds. It is claimed by Iacoboni, for example, that a virtue of ST (or at least of the evidence supporting it) is that it does not allow this framework to get off the ground. Iacoboni thinks empirical evidence supporting implicit ST shows that the problem does not really arise. He also suggests that the problem of other minds stems from a Cartesian conception of the mind as private and solitary.[433] He expresses sympathy with Wittgenstein's view that it is a pseudo-problem. He also notes that we tend to predict other people's behaviour both accurately and effortlessly.[434] Iacoboni argues implicit ST mechanisms enable 'the sharing of mental states between individuals'. He also argues MN activity should be equated with the coding of intentions in other individuals. We employ simple mirror mechanisms to interpret the intentions associated with everyday actions we observe others performing. Our neural mechanisms do not process 'mere behaviour' but directly respond to the emotions and intentions of other agents. An understanding of

---

[433] Ibid P.122
[434] Ibid

the intentions associated with those actions allows us to anticipate what another person will do next, allowing the prediction of not yet observed actions.

Similarly, Gallese also often explicitly stresses that we do not have to rely on analogical arguments to the existence of other minds because we possess ES mechanisms which save us the trouble of having to draw any inferences. He also stresses that the process of empathy is not the outcome of explicit analogical inference.[435] It is the product of a direct matching between another's expressive behaviour and one's own affective experience. In putting forward his case Gallese makes use of the phenomenological notion of the lived body developed by Husserl and others. The bodies of others are not experienced as mere corporeal objects but as something alive and similar to our own experiencing bodies.[436] When we 'directly recognise others' it is as 'selves like us, rather than as bodies we must then satisfy ourselves are 'endowed with a mind'. This idea will be examined in more detail when we examine the "direct perception" answer to the question of how we grasp the psychological states of others.

Proponents even claim this gives ST an advantage over TT. Iacoboni draws a comparison between the TT approach to understanding others and the argument from analogy.[437] TT, though, models itself more closely on a scientific attempt to understand a physical phenomenon on the basis of data and well known physical laws. It posits causal laws linking human behaviour and internal psychological

---

[435] Gallese 2007 P.147
[436] Gallese 2003 P.176
[437] Iacoboni 2009 P.123

states. Iacoboni notes that arguments from analogy have been heavily criticised by past thinkers on the grounds that the form of reasoning is far too complex for something we apparently accomplish so quickly and effortlessly, and he indicates that TT suffers from the same failing. The starting point for the reasoning process utilised in the argument from analogy is that "I only have access to one mind, which is my own mind".[438] One therefore has to draw inferences from the behaviour of others to their inaccessible psychological states. One first works out that there is a connection between one's own mind and the behaviour of one's own body. For example if one is in pain one may scream.[439] One next draws an analogy between the body of another person and one's own body. This suggests there may also be an analogy between the other person's body and the other persons mind.[440] So if one observes another person screaming one concludes that person is in pain. Iacoboni argues that this methodology does not allow complete certainty about the psychological states of others or to share their feelings and emotions. Iacoboni argues ST, unlike TT, requires no knowledge of causal laws linking human behaviour and internal psychological states and very little inferential processing.[441] It is safe from the accusations of implausible complexity. ST is also in a better position to account for the recent neuroscientific evidence.

In sub-personal accounts of ST it is argued that both our own psychological states and those of others can be objects of direct experience in suitable conditions.

---

[438] Ibid PP.122-123
[439] Ibid PP.122
[440] Ibid PP.122-123
[441] Ibid P.123

Psychological states of other agents do not require an inferential process based on non-intentionally characterised cues. For example we saw that ES mechanisms are argued to avoid the need for inference. They are also argued to instantiate an intersubjective realm prior to the formulation of a first personal perspective. Indeed Gallese thinks the real epistemic problem each individual faces is to establish an individualised perspective which requires distinguishing oneself from other minds by subsequently coming to realise one's own individuality within an intentional community.

Jeannerod et al also resist the picture in which the intentional behaviour we observe in others is on a par with non-intentional movement. They claim that ST provides evidence that intentional qualities are directly perceivable. Jeannerod et al argue there is indeed a problem to be solved when attributing psychological states to an agent; but it is a problem that applies equally to the first-person as to the third-person case. There is no special problem about the psychological states of others. Jeannerod and Pacherie overcome the problem by introducing a new problem into philosophy, the problem of self-identification.[442] We have already seen that they make a case for thinking action and intention are not immune to error though misidentification. However they argue that it is just such immunity to error, hand in hand with the accompanying inferential character of our knowledge of other minds that creates the illusion of an asymmetry between our knowledge of our own psychological states and of the psychological states of other agents.[443] The

---

[442] Jeannerod and Pacherie 2004 P.140
[443] Ibid

asymmetrical relationship results in radically different criteria being employed in the attribution of psychological states to ourselves and others. This makes it difficult to believe we are applying the same concepts in each case. The flip side of the claim that self-attribution of psychological states is not immune to error is that our knowledge of other minds need not always be inferential.

Georgieff and Jeannerod's claim that representing intentional actions already presupposes other agents suggests that there is no deep problem of other minds. On this view the analogical argument's premise that we have incorrigible access to our own psychological states is as flawed as the premise that we only have indirect access to others psychological states. Jeannerod et al seek to undermine the sharp contrast between these two ways of apprehending psychological states.

In discussions of intentional understanding there appears to be an attempt by sub-personal ST to show that a sensitivity to the intentional qualities of movement and expression is built into our automatic processing of others. This makes sub-personal versions of ST appear close to 'direct perception' theories that will be examined later. The context of the discussion is how the activation of MNs can facilitate an appreciation of an intentional action rather than a series of movements. Some MNs are context sensitive in the sense that they are only activated in response to a movement, for example a grasping action, if it occurs in a certain context. If an agent grasps an object to bring it to his mouth this will cause different neurons to

discharge in the observer than if he goes on to place that object in a cup.[444] Identical

motor acts can be discriminated according to the context in which they are

embedded. Depending on which of these parietal mirror neurons discharge, a

different motor schema is activated.[445] In this way the observer is able to predict the

next action. These motor pathways may be built up by past experience. The ability to

ascribe intentions then becomes the ability to predict their goal. All the situation-

specific knowledge the subject needs is built in.

Proponents of sub-personal ST both buy into the picture that gives rise to the other

minds problem and try to answer the other minds problem without falling into

problems associated the argument from analogy. However their account of what is

involved in directly accessing the psychological states of others remains

unsatisfactory. Proponents of sub-personal ST attempt to align themselves with a

philosophical position in which our perception of other might be termed 'direct'.

This is a position which will be explored in the later chapters of this PhD. However

any claims proponents of ST wish to make regarding the direct nature of

intersubjective perception are incompatible with the core structure of simulation. ST

proposes a particular answer to question 1. They take this answer to be driven by

empirical evidence they provide in answer to question 3. However the empirical data

does not dictate their answer to these questions. One can't proceed from the

empirical data to any answer to question 1. Attempting to do so will either involve

---

[444] Ibid P.136
[445] Ibid P.137

postulating an implausible personal level description of what we are doing or an illegitimate relocation of personal level characterisations at a sub-personal level.

The use Gallese makes of the phenomenological notion of the lived body is problematic in the context of ST. Because ST treats psychological states as inner it must assumes an inner/outer distinction in the case of others which parallel one's relation to one's psychological states. This marks a crucial diversion from direct perception which does not begin from a first person perspective. Similarly Iacoboni's claim that ST can facilitate a non-inferential grasp of the psychological states of others which does not have to utilise laws linking inner states with behaviour is in conflict with a core principle of ST which is that we are inferring another's psychological states by means of a simulation. The reason this problem is often overlooked is because sub-personal formulations of ST are not depicted as utilising non-intentionally characterised data in their inferences. Instead they make inferences on the basis of goal-orientated acts such as grasping motions. Nonetheless inferences are still being made on the basis of assumptions about the significance of the behaviour one observes.

Gallese's suggestion that the central epistemic problem ES mechanisms enable us to solve is not the problem of how to assure ourselves that other agents are also the subject of psychological states but rather the problem of how we can conceptually separate ourselves off from other minds to achieve a conception of oneself as an individual entity takes us outside the framework of ST. There is no justification for

calling this a simulation process. Similarly the claim by Jeannerod et al that there is

no special problem of other minds undermines the need for simulation. Their claim

that there is a problem of self-identification conflicts with the ST model which

simply takes one's own psychological states for granted. Furthermore Jeannerod et

al dispute the idea we find an asymmetry in the way one knows one's own

psychological states and the states of others. Georgieff and Jeannerod's denial that

our access to other agent's psychological states is indirect would sit more coherently

in a direct perception theory. All these attempts to sophisticate ST in fact mark a

departure from the theory.

Aim B

ST usually considers its opposition to TT to be empirically driven. A significant

proportion of this empirical data is taken from the same studies that TT appealed to

in the last chapter including the false belief tests and data on autism. Simulation

theorists argue that their theory offers a more plausible explanation of the empirical

findings.

Sub-personal versions of ST are largely driven by the discovery of MNs which they

regard as a major empirical breakthrough. MNs are either interpreted as directly

establishing that we perform simulation routines or taken as evidence of a biological

precursor of these routines. The key point for ST is that the same mechanisms are at

play during personal psychological experience and observation of others. ST claims

to provide for the possibility of empirical discrimination between itself and TT. The key point is that there is no reason TT should predict that the same mechanisms should be operative during first and third-personal experience whereas ST hypothesizes this will be the case because the simulator is attempting to copy the perceived agent rather than representing their psychological state theoretically. MN activity is argued to constitute exactly the kind of data ST predicts. First-person familiarity with intentional actions (underwritten by MNs) is a condition for recognizing these actions performed by other agents. MNs are also argued by Jeannerod et al to provide empirical evidence that psychological states come unassigned to an agent. ST also appeals to evidence that we display activation of the same muscle groups that the observed agent is using. We are claimed to literally simulate the actions of others just below the threshold of action.

Sub-personal versions of ST also appeal to data on autistic subjects. For example difficulties autistic subjects are known to have recognising others psychological states can be explained as resulting from an incapacity to establish a motor equivalence due to a defective MN system. As well as exploring intersubjective understanding, sub-personal formulations of ST also draw support from unrelated fields such as aesthetic appreciation. MNs are clamed to enable our empathetic responses to art works.

As we saw with TT in chapter two, ST is also motivated by wider empirical considerations as well as by direct empirical evidence. In particular they appeal to

201

the plausibility of simulation within an evolutionary framework. They also argue

that the cognitive continuity it posits between humans and primates is an advantage.

In fact most evidence for the role of MNs is drawn from primates and the evidence

for a MN system in humans is more tentative. ST is also driven by scientific

concerns for simplicity of explanation. An objection ST has for TT concerns the

plausibility of claims that we represent tacit theoretical laws in order to understand

others. TT is thought to offer an explanation that is too complicated to be usefully

scientific. Although these concerns are empirical they are not theoretically neutral.

They buy into a particular picture of how science proceeds.

MNs are also employed by sub-personal ST as empirical persuasion for claiming

that the framework underlying the problem of other minds is mistaken. MNs appear

to be sensitive to the intentional qualities of movement and the wider intention of an

act is what drives MN activation. However if this is the correct way to interpret this

evidence than it actually undermines the framework motivating ST.

It is not clear that MNs really provide strong empirical support for ST. While it is

true there is no reason to expect the same mechanisms to be active during

psychological episodes and observation of those episodes in others it does not follow

that, in the observation cases, these activations are accurately characterised as

instantiating simulations of psychological states. ST have been unable to offer a

convincing case that the MN data really does underlie a simulation process. Indeed

even proponents of ST do not consistently interpret the data in this way with some

interpreting the findings as evidence for unassigned representations rather than simulations.

Crossing personal/sub-personal boundaries

An illegitimate movement between personal and sub-personal levels of description featured in a number of objections made to ST in this chapter and it is worth emphasising the problem. We have seen that ST tries to avoid meeting philosophical difficulties such as its phenomenological implausibility as a personal level account by relocating simulation at the sub-personal level. However there are problems associated with a move to sup personal being thought to legitimate personal level descriptions. One difficulty is that the theory makes an unacceptable leap from a modest piece of neurological data to a robust psychological claim. The fact that we find matching neurological activation does not entail a personal level description in which we observe a psychological intention and then solve a problem of who to assign it to. Indeed it does not entail any personal level account of what is going on. ST thinks the empirical data supports a particular account of what we are doing but it only does so because they describe the data in a particular way. They are not entitled to this description.

Summary

This chapter began by looking at the earliest and clearest formulations of the simulation process postulated by Gordon and Heal. They oppose the TT position in which we employ theoretical inferences to understand and predict others. Instead, they claim, we extend a capacity to place ourselves in hypothetical situations by making further hypothetical adjustments to encompass features of another agent's situation. This is understood as a more practical venture. Though ST is described as a personal level process Gordon and Heal suggest simulation may also occur at a sub-personal level. This chapter then outlined general features of ST: that it is a species of mindreading, it employs representations of psychological states and it takes place offline. This chapter proceeds to examine sub-personal versions of ST which have come to dominate the field. It explores the neurological structures that are supposed to instantiate a simulation routine. It looks at the empirical data which is claimed to give ST the edge over TT. In these accounts the problem requiring simulation is reframed as a problem of assigning disembodied psychological representations to the self or to another agent.

The chapter went on to consider ST's responses to the questions raised in the introduction. These responses raised the problem that ST lacks phenomenological plausibility as a personal level process. It is this feature which motivates a move to the sub-personal level which turns out to be problematic. This move raises a difficulty about the coherence of employing personal level descriptions at a sup

personal level. The answer to question 1 also threatens to confine the meaning of psychological terms to first-personal experience. Any claim in answer to question 3 that the neurological processes described are really executing an analogue of the simulation routine offered in personal level accounts is also unconvincing.

ST is meant to be an empirically driven theory. However we saw that their answer to question 1 influences their answer to question 3 by shaping their reading of the empirical evidence. MNs are understood as representations of psychological states. ST also fails to fully appreciate the difficulties inherent in their answer to question 1. Matters are further complicated by the fact that sub-personal versions of ST do not want to claim that the meaning of psychological terms is anchored in the first-person. In fact they do not offer an account of how they get their meaning.

Though some simulation theorists suggest the empirical data offers a way out of certain philosophical problems such as the problem of other minds, this PhD is arguing that their answer to question one in fact raises these very problems. Their answer to question one raises question 4. Psychological states are being understood as inner states. Their meaning is revealed, in our own case, by introspection rather than belonging to a public theory. A gap is opened up between our knowledge of our own personal states and those of others which is philosophically problematic. Some simulation theorists attempt to avoid this difficulty by treating psychological representations as agent neutral. However this move cannot accommodate the reflexive character of psychological states. The empirical data is neutral as to what

personal level description we might give of it. This chapter next considered how ST

positions itself in relation to the argument from analogy. We saw that their attempts

to make intersubjective perception of others direct were blocked by their answer to

question 1. Finally the chapter considered the extent to which ST is an empirical

rather than a philosophical position. Though there is a considerable amount of

empirical data informing ST the position is still driven by a philosophical picture

which determines how the data is read.

4 Hybrid Theories of Mind

This chapter will examine some exchanges that have been made between TT and ST positions. These will concern questions about the extent to which ST and TT are mutually autonomous, and about what role experimental evidence can play in deciding between these two theories. It will also explore some hybrid positions in the TOM debate. Most of the advocates of hybrid positions will already be familiar to the reader of the previous chapters on TT and ST. This is because most advocates of TT and ST now allow some space for a limited role being played by the rival theory. Hybrid properties of a theory can be either intentional or unintentional. We will begin with unintentional cases where one theory has not done enough to distinguish itself from its rival counterpart. This will require further spelling out the different forms each theory can take. Before turning to TT-ST hybrid positions, this chapter will explore some refinements that have been proposed for both parent theories.

Similarities between ST and TT

Although TT and ST are often presented as diametrically opposed theories, closer examination reveals some significant overlaps between them. In particular closer inspection of ST shows that it does not offer a total breakaway from TT. Davies and Stone argued that ST relies on theoretical assumptions about similarity or

rationality.[446] According to them, the contrast between the TT and ST strategies is between knowledge-rich and the knowledge-poor, rather than between knowledge-rich and knowledge-free, approaches to mindreading.[447] In support of this claim they appeal to Goldman's concession that ST employs a theoretical assumption that other people are psychologically similar to oneself, and Heal's parallel concession that ST needs the assumption that others are like oneself.[448] Davies and Stone note that Goldman goes further allowing that more widespread theoretical induction plays a role in mind reading. Goldman concedes that regularities concerning behaviour and even individual idiosyncrasies can be learned inductively.[449] Davies and Stone suggest such concessions mark a move away from pure ST towards a hybrid position. Later in this chapter we will see that Goldman has subsequently taken the same view; he now advocates a Hybrid ST-TT position.

Other examples of ST acknowledging a role for theoretical processes are easy to find in the literature. For example, Gordon is happy to concede that explanations about other people's behaviour are couched in terms of propositional attitudes including beliefs and desires and even that predictions concerning others behaviour exploit this framework of causal and nomological relations.[450] He also concedes that there can be formal regularities that can be cast as causal laws. For example a belief p and a belief 'if p then q' typically cause a belief q. However he thinks there are limitations to this strategy. There is an indefinite number of possible circumstances where these

---

[446] Davies and Stone 2001 P.18
[447] Ibid
[448] Ibid P.16
[449] Ibid P.18
[450] Gordon 1986 P.165

regularities fail to hold.[451] In such cases, Gordon argues, we must employ these generalizations 'in the context of practical simulation'.[452] He suggests that non-specifiable constraints on one's own practical reasoning system aid in defining the application of folk psychological rules. They provide a starting point which can be modified as one learns more about others. Gallese et al also claim E.S. works in conjunction with more theoretical strategies with E.S. facilitating an automatic and non-inferential understanding of other's intentional states and actions.[453] Even the clearest advocates of ST then seem amenable to some kind of theoretical processing.

Both TT and ST come in a variety of formulations. This makes it difficult to formulate decisive arguments or evidence in favour of one position over the competitor. Modifications to each theory have been made with the rival competitor in mind. In order to evaluate the strength of the boundaries between TT and ST it will be useful to first review the key ways in which particular instantiations of each theory can vary.

Boundaries of TT

1. Innate versus learned

We have already seen that a theory may be innate or learned. If theory is understood as the product of a learning process then this process can involve either scientific theorizing or cultural assimilation.

---

[451] Ibid P.166
[452] Ibid P.165
[453] Gallese Eagle and Migone 2007 P.144

2. <u>Modular versus non-modular</u>

We also saw that theories may also be envisioned as modular (e.g. Carruthers) or non-modular (e.g. Gopnik).

3. <u>Broad verssus narrow</u>

The term 'theory' can be employed in a narrow sense which stipulates that the body of knowledge should consist of generalisations with the status of psychological laws. A closely related idea is that FP makes use of theoretical entities or processes distinct from ordinary observable processes that it seeks to predict and explain.[454] The term 'theory' can also be employed in a more inclusive sense incorporating any stock of psychological information.[455] For example Stich and Nichols use theory to include any body of psychological information whether or not this information has the structure of psychological laws.[456] Whereas early versions of TT typically posited internally represented knowledge structures with explicit rules, the idea of sentence like or rule based knowledge structures has since come under pressure from alternative models including connectionist models. In these models knowledge involved in prediction is located in connection strengths between the nodes of a network. It is thought to be implausible to interpret such networks as encoding sets of rules.[457]

---

[454] Stich and Nichols 1995 P.88
[455] Davies and Stone 2001 P.1
[456] Ibid PP.3-4
[457] Stich and Nichols 1992 P.46

Stich and Nichols are key advocates of a broad TT position. Their recommendation

is motivated by a concern to preserve the integrity of the TOM debate in the face of

multiple formulations of both TT and ST. Later in this chapter we will examine

disputes about which hypothesis is actually supported by a given piece of

experimental data. Stich and Nichols suggest such disputes can sometimes have a

terminological origin as both TT and ST are used in a variety of ways which are not

always clearly specified by their protagonists.[458] They argue in favour of a definition

of each theory that preserves mutual exclusivity.[459] Their suggestion is that the

debate should be between 'broadly construed' TT and 'off line simulation'.

(Goldman's work is a central example of off line simulation. This definition of ST

is opposed to Gordon's rival notion of 'imaginative identification').[460] An advantage

of this is that successful arguments against broadly construed TT will also be

arguments in favour of off line ST whereas arguments against particular

instantiations of TT do not necessarily lend any support to ST. For example, an

argument against a version of TT that insists on law-like generalisations will not

automatically lend support for ST as there are rival versions of TT which do not

insist on these generalisations.[461] They also think these are the only plausible

theories about the mechanisms behind our ability to predict the behaviour of others.

Furthermore Stich and Nichols think that there are domains of knowledge where

narrowly construed TT accounts are not plausible. For example it is implausible to

think a law-like structure could underpin our judgement about what counts as polite

---

[458] Stich and Nichols 1995 P.87
[459] Ibid P.90
[460] ibid P.91
[461] Ibid P.90

211

or impolite behaviour.[462] They argue it is possible that FP will turn out to resemble these kinds of judgements rather than scientific judgements.

4. Personal level versus tacit

This distinction is not absolute as no TT supporter makes the implausible claim that all mind reading involves conscious theorising about what another agent believes or intends. Rather theory theorists can differ about the degree to which our knowledge of FP is tacit. Davies argues one may employ a theory without being aware one is doing so. One need not be able to articulate the theory explicitly. In such cases the theoretical knowledge is tacit or implicit.[463] When using a theory one may be unaware of the theoretical knowledge or of the conceptual resources involved.

The notion of tacit theory raises again the question of what counts as a theory. Heal offers a useful account of theory. Heal allows that theories may be tacit but claims explicit theories provide us with paradigmatic examples of what counts as a theory.[464] For Heal a theory is an articulated structure composed of elements. Both the structure and the elements entail claims that will either be expressible in a public language or by rules of inference. A theory must be more than the conjunction of a number of propositions. The propositions must be systematically related to one another in a way that allows one to express the ways in which they interact and the

---

[462] Ibid P.88
[463] Davies 1994 P.192
[464] Heal 1996 P.76

context in which each is important .[465] It is not enough to be able to predict what will happen in a particular case, one must be able to locate the case amongst other possible cases; one must be able to express what would happen in different circumstances and why.

Stich and Nichols offer an explanation of what is meant by tacit theory. The rules or principles of an internally represented knowledge structure are understood as constituting the agents theory of the domain in question. Whereas for some domains of knowledge an agent may have partial conscious access to these rules or principles, in the majority of cases we have no conscious access and in such cases the theory is tacit.[466] Stich and Nichols claim that in this type of case we need not be conscious of employing a theory.[467] Stich and Nichols think we engage a 'largely tacit psychological theory' when involved in prediction, interpretation and explanation of FP practices.[468] TT advocates differ about the extent to which we are able to explicitly grasp the rules of FP.[469] While some TT theorists have understood the rules as platitudes, it is also possible for advocates of broad TT to claim the rules are no more accessible than the rules of grammar.

A related dimension along which TT positions can vary concerns the explicitness of a theory. Explicitness here refers to the form in which theoretical information is

---

[465] Ibid P.77
[466] Stich and Nichols 1992 PP.35-36
[467] Ibid P.39
[468] Davies and Stone 2001 P.4, Stich and Nichols 1992 P.36
[469] Stich and Nichols 1995 P.90

stored rather than to its availability for conscious articulation.[470] Davies and Stone

explain that explicit representation has a language-like format. This information has

to be accessed before it can be used. There is disagreement about whether explicit

representation is involved in tacit theory with Heal arguing for its involvement and

Stich and Nichols arguing it is not required.[471] Stich and Nichols claim that a

connectionist network could embody tacit theoretical knowledge even though the

activity of these networks does not involve the operation of explicit language-like

representations. They think it is unimportant whether theory turns out to be

represented in rule like structures or in connectionist architecture. However, they

acknowledge that they are entitled to make this claim only because they construe

theory broadly. As we just saw, they recommend the broad construal of theory so

that TT and off line ST exhaust the possibilities in the TOM debate. If one starts

with a narrow construal of theory then a discovery that we employ connectionist

architecture for mindreading would undermine TT (though it would not be evidence

for ST).[472] This discovery would still be compatible with a broader notion of TT

however.

---

[470] Davies and Stone 2001 P.3
[471] Ibid
[472] Stich and Nichols 1992 P.49

Boundaries of ST

1. Frequency

Goldman suggests three dimensions of strength along which versions of ST can vary.[473] The first concerns the frequency with which simulation is employed. The strongest formulations of ST claim it is always used for mind reading. A more moderate theory will claim simulation is the default process which can be overridden in special circumstances.[474] Weakest formulations of the theory will claim simulation is only employed occasionally.

2. Status

A second way in which ST theories can differ from one another concerns the 'source' or status of simulation. The strongest versions claim it is the primary mode of mind reading. Weaker versions claim simulation is merely a shortcut employed during mindreading, this being conceived as ultimately derived from theorizing.

3. Range

Goldman's third dimension of strength over which ST positions can differ concerns the range of mental states thought to be amenable to simulation-based mind reading. The strongest versions of ST will claim all mental state types are amenable. Weaker versions will restrict the role of simulation based mindreading to a set of subtypes such as propositional attitudes or emotions.

---

[473] Goldman 2006 P.42

[474] This is the position Jeannerod and Pacherie take with regard to action understanding. Simulation may occasionally be supplemented or overridden by theory (Jeannerod and Pacherie 2004 P.128)

4. Degree of explicitness

Jeannerod and Pacherie suggest two further dimensions across which simulation theories can vary.[475] Simulation may be understood as involving the explicit acting out of another's mental states and processes. Alternatively it can be understood as an implicit, automatic and unconscious process. One can also find a blend of these two approaches.[476]

5. Content

Jeannerod and Pacherie's second dimension of variability concerns whether the content of ST is restricted to propositional states or whether it also includes mental imagery.

6. General vs. task-specific

Another way Goldman suggests simulation theories may differ is over whether we possess a general simulation system responsible for all instances of simulation based mindreading or whether we have a number of task-specific simulation mechanisms.

Undermining theoretical opposition between ST and TT

1. Problems with tacit theory

In this context the notion of tacit knowledge is vulnerable to philosophical objections. On its broadest construal it runs into danger of trivialisation. On the

---

[475] Jeannerod and Pacherie 2004 P.128

[476] The terminology of explicit and implicit is equivalent to the distinction between personal and sub-personal levels of explanation.

broadest construal, it appears that any activity could be argued to be making use of a tacit set of principles. As Davies and Stone explain, any belief is the outcome of some mechanical process and this can always be construed as a tacit theoretical one.[477] Any skill an agent might possess can potentially be given a step by step explanation appealing to mediating theoretical principles. The agent who performs these skills might be argued to possess a tacit understanding of the principles employed even though these principles cannot be articulated. Davies and Stone give the example of an agent predicting the pressure in gas cylinders. The agent achieves this by carrying around a cylinder of gas. The agent heats this cylinder in order to simulate the pressure changes in other gas cylinders when they are heated. On this very loose notion of theory, the agent counts as having theoretical knowledge of general principles relating temperature and pressure in gas cylinders.

If we stretch the notion of tacit theory even more thinly this problem comes up in an even more pernicious form. Even inanimate objects have been shown to conform to general laws. The worry this raises is that a notion of tacit theory could even be inclusive enough to incorporate the behaviour of planets and rocks. Davies and Stone point out that this reading of theory would entail it is not simply the agent who might turn out to embody tacit knowledge of the principles relating temperature and pressure, but the gas cylinder itself.[478] However they point out that the gas cylinder is not working with representations of temperature and pressure so argue we are not actually dealing with the possibility of tacit theoretical knowledge in such cases.

---

[477] Davies and Stone 2001 PP.21-22
[478] Ibid

2. Compatibility

The example of an agent using a gas cylinder in order to predict the pressure in other gas cylinders is construed by the authors as a simulation process. The example therefore suggests that the two positions may turn out to be compatible, at least on this very loose construal of tacit theory.

Davies was an early advocate of TT who argued that simulation couldn't provide a philosophically fundamental account of how we employ mental concepts.[479] Like Stich and Nichols, Davies was explicit that he wished TT to be understood in a fairly broad and inclusive sense of theory. He stressed that theoretical knowledge may be tacit or implicit.[480] While both Goldman (who at this time argued that the outcome of simulation is not meant to be determined by any theory possessed by the simulator)[481] and Gordon were firm that their positions stood in direct opposition to TT, Davies argued that the idea of simulation does not by itself entail a strict opposition to theorizing. One could accept simulation is involved in prediction and explanation without thereby being committed to philosophical claims about the nature of mental states, or the conditions required in order to possess mental concepts.[482] Davies argued that hybrid positions are at least coherent. For example, one may hold that simulation is a useful heuristic for predicting another's future behaviour, or explaining their antecedent behaviour while maintaining that mastery of the mental concepts involved is constituted by tacit theoretical knowledge of a

---

[479] Davies 1994 P.191
[480] Ibid P.192
[481] Ibid P.195
[482] Ibid P.196

218

theory, and that this knowledge enables normative evaluation of the simulated output**.**

3. The threat of collapse

Davies was one of the first philosophers to introduce the idea that ST may collapse in to TT. In making this claim he had in mind a particular construal of what is involved in simulation. He began by distinguishing two possible ways to interpret the simulation routine. On the first interpretation the simulator employs their imaginative faculties to entertain hypotheses about mental states, for example 'I believe that p',' I desire that q.[483] They then advance to a conclusion about a further mental state, action or intention. For example 'I believe that r'. On the second interpretation, which Davies termed 'imaginative identification' the simulator imaginatively adopts these actual mental states. The simulator imagines believing 'that p' and desiring 'that q'.[484]

Davies suggests that the first construal opens up a threat of collapse between ST and TT. This is because the processing involved could match the shape of a conclusion (about, for example, an intention) based on premises (about the agents intentions) embedded in a psychological theory. He claims that for a cognitive processing system to embody tacit theoretical knowledge of the relevant domain it must involve a derivational process that matches the structure of a causal process. It must involve causal mediation between representational states in a fashion analogous to the way

---

[483] Ibid P.204
[484] Ibid P.207

in which a rule mediates between premise and conclusion.[485] Davies point was not

that TT and ST must collapse on this understanding of simulation, only that collapse

was possible unless further theoretical stipulations were made to distinguish the two

positions. As Heal explains there is a risk when both input and output to the ST

process are explicit representations of psychological states. If simulation is

understood this way it looks probable that patterns of causal dependence will be

discovered holding between these representations. It is also likely that these patterns

will possess a structure that implies internal mediating processes characteristic of a

tacit theory are at work.[486]

If, Davies suggested, instead of generating hypotheses displaying the form 'x

believes that p', simulation is instead envisaged as imaginatively adopting a mental

state such as 'that p' then such states can no longer be thought of as appropriate

inputs to a mechanism embodying tacit knowledge of a psychological theory.[487] In

their place we simply have 'pretend belief' and 'pretend desire' states.[488] The

mechanisms mediating transitions between these cannot be understood as

manifestations of tacit knowledge of the principles of a psychological theory.

Heal took up the concerns about a collapse raised by Davies and argued these

concerns actually have a wider application than Davies had envisioned. She argued

that the danger of collapse between ST and TT was harder to avoid than Davies

---

[485] Ibid P.205
[486] Heal 1994 P.131
[487] Davies 1994 P.207
[488] Ibid

realized.[489] Heal agreed with Davies that the first conception of simulation does not do enough to establish that there is a non-theoretical approach. Heal along with Goldman acknowledged that a good explicit theory can produce a sequence of representations paralleling the unfolding developments in the phenomena to be understood (just as one's own mechanisms are meant to run parallel to the other's decision making processes during simulation).[490] On this conception of ST, simulationists may share with theory theorists the assumption that structural relations of dependence hold between explicit input and explicit output. Heal argues that this assumption weighs against ST and in favour of TT.[491]

Heal disagreed with Davies that the threat of collapse could be avoided by switching to an alternative conception of ST that construes simulation as imaginative identification with another agent. This is because she holds that the risk of collapse is not attributable to a particular conception of simulation but to an unarticulated conception of the nature of the question as a whole. Heal claims that there are two ways of conceiving the TT vs. ST debate: firstly as an empirical question about how our ability to predict others future thoughts feelings and actions on the basis of their current psychological states is implemented at a sub-personal level.[492] Secondly as a primarily a-priori question about how abilities at a personal level are interrelated. In particular as a question about what happens when we think about the thoughts of

---

[489] Heal 1994 P.129
[490] Ibid P.131
[491] Ibid P.132
[492] Ibid

others.[493] Heal argues the threat of collapse is motivated by an adherence to the first way of conceiving the debate.

Heal argues the picture behind TT posits a division between objects on the one hand and thoughts about them on the other. The former are out in the world while the latter are in people's minds or heads. Thinking about the two involves separate parcels of knowledge.[494] She argues that the simulatonist who adheres to Davies first conception of simulation will not dispute this picture. This type of simulation theorist will agree we have a specialised body of knowledge about thought separate from our body of knowledge about objects. However this simulation theorist will suggest that in cases of predicting others we simply use our own minds off line to generate predictions without having to appeal to this body of specialised knowledge.

This type of off line strategy is not unique to prediction of psychological items. It could also be employed with non-psychological items.[495] Heal gives the example of using a heart connected to pretend inputs in order to simulate cardiac behaviour. The heart has a structure and mode of operation that could be fully specified by an adequate theory of its workings. Such a theory's predictions will be systematically dependent on the heart's various features.[496] In the heart simulation case, if one component of the heart differed then all predictions exploiting this feature would differ while all predictions in which it did not feature would remain constant. Heal

---

[493] Ibid P.138
[494] Ibid PP.132-133
[495] Ibid P.133
[496] Ibid P.134

suggests this is the same pattern we should expect to find if the predictions were delivered by a theory. The causal role played by the heart-simulator in mediating prediction is analogous to the logical role played by statements in a theory. Various features such as size and structure determine how the heart functions. Its mode of operation is fixed by these features which could be specified within a theory.

If we conceive simulation this way, the collapse comes when one asks what could carry or encode more information about a type of thing than such a thing itself.[497] Heal argues that this picture of mental simulation does not differ significantly from TT, provided we are prepared to accept the definition of tacit knowledge we have been working with.[498] This information is available to a subject if they are able to extract it easily. Tacit theories entail no commitment to explicit knowledge or to the form in which information is carried.[499] Anything that fills the right logico-causal role will count as a vehicle of tacit knowledge. The portable heart will count if we carry it around and interrogate it effectively.[500]

Heal next suggests that this threat of collapse does not depend on what conception of simulation we adopt.[501] She disagrees with Davies suggestion that if advocates of ST stipulate that the simulator represents 'that P' rather than 'I believe that P' they can avoid the risk of collapsing into tacit theory. Heal points out a difficulty with this suggestion, one must begin with a representation of the form 'so and so believes that

---

[497] Ibid P.135
[498] Ibid P.134
[499] Ibid P.135
[500] Ibid
[501] Ibid

P' if one is to maintain a grip on the distinction between one self and others.[502]

Subsequently imagining 'that p' is only one component of this total thought. This is still a thought about another person's mental event.

If we stick with conceiving the debate as an empirical one, Heal argues, the notion of tacit theory must be redefined for there to be an issue between TT and ST.[503] For example the issue could concern whether the tacit theory is located within or outside of one's practical reasoning system (this is how Stich and Nichols suggested we should conceive the debate). Alternatively one could insist theoretical information must have an explicit representational structure. Heal thinks this would be the best option but is still pessimistic about resolving the ST-TT debate for two reasons. Firstly, there is vast disagreement on what could count as a theoretical representation. Secondly there is no reason to assume that investigation of our cognitive architecture will provide a decisive confirmation to this question.[504]

Heal argues that the best way to keep a genuine distinction between ST and TT is for ST to take issue with the assumption motivating TT that there is a division between objects and thoughts. This observation has its roots in the obvious fact that objects are one thing and thoughts about them are something else.[505] This makes it look plausible that thinking about them will involve appeal to separate bodies of knowledge. It will be through accessing a body of knowledge of psychological

---

[502] Ibid P.136
[503] Ibid P.135
[504] Ibid P.137
[505] Ibid P.133

224

phenomena that one will be able to make predictions about another's thoughts and behaviour.[506] Heal resists the view that one can separate thinking about thoughts from thinking about their objects.[507] If one holds instead that the capacity to think about thoughts is necessarily an extension of the ability to think about their objects then we cannot make this division.[508] In thinking about another person's thought 'that p' one is exercising the same cognitive skills that are exercised when thinking 'that p'. Acquiring the ability to think about other agent's thoughts involves applying in a particular way one's ability to think about the contents of these thoughts.[509] These capacities are interrelated. Heal argues that understanding what it means to claim another is thinking 'that P' is to be able to appreciate what impact this may have in their other thoughts and behaviour.[510] Thinking about another's thoughts is a special application of an ability to think about the subject matter of those thoughts.

Davies and Stone argue that whether there is a collapse depends on the dimensions of strength of each theory. If TT is understood in a minimal sense, and if we also accept that ST requires some theoretical background then there will be a collapse. However if TT is understood, as Davies and Stone do, in a stronger sense as involving a stock of self-sufficient FP laws then there is no collapse.[511]

---

[506] Ibid
[507] Ibid P.137
[508] Ibid PP.138-139
[509] Ibid P.138
[510] Ibid
[511] Davies and Stone 2001 P.19

<u>Goldman on the threat of collapse</u>

Goldman now advocates a hybrid TT-ST position. A motivation for this change of position is a response to threats of collapse. Goldman argues that if an argument for collapse is successful then what it really establishes is not that ST collapses into TT but that we have a simulation-theory hybrid.[512] If it turns out that the final part of the simulation process utilises a theoretical inference we still can't dismiss the simulation-like character of the earlier stages.

Goldman tackles the worry that simulation must ultimately be the result of some kind of theorizing.[513] He explains that the threat of collapse can be construed in several ways. He first formulates it as follows. To move from simulating a pretend decision making process to an attribution of the output state to another agent involves relying on a theoretical assumption that the psychology of the other agent is relevantly like one's own.[514] As mind reading is applicable to a large number of other people this assumption must be quite general. If this is a psychological generalisation then ST is making use of a tacit psychological theory. However Goldman argues an appeal to the resemblance of others to oneself does not invoke any law of FP.[515] TT laws describe mental state transitions that take place in an individual or between an individual and her environment. A law linking oneself to

---

Goldman 2006 P.30
[513] Ibid P.32
[514] Ibid P.30
[515] Ibid P.31

others describes similarities across individuals. Furthermore Goldman thinks it is not obvious a priori that we actually employ a resemblance to self premise. Mind reading can be applied to targets that bear little resemblance to a simulator, including cartoons and even moving geometric shapes.[516]

Goldman next considers a construal of the threat suggested by the work of Dennett. This is the idea that simulation amounts to using oneself as an analog computer. Simulators initially put themselves in other people's initial states and then act on these states in order to output a result to attribute to the other people.[517] It is hard to see how this can be anything other than a kind of theorizing. Knowledge of the simulation subject is required to drive the simulation. Goldman's response is to draw a distinction between theory driven simulation and process driven simulation.[518] Dennett's suggestion would belong to the former category. Theory driven simulations are most useful when trying to simulate a system that is radically different from oneself. This requires a good theory of the thing we are attempting to simulate. Process driven simulations can be used when there is a strong resemblance between the simulator and what it is trying to simulate. All that is then required of the simulator is that she employs the same processes that the target is employing. In cases of mindreading the simulation process must match the process going on in the other person. The simulator's initial states must also match those of the other person.

---

[516] Ibid
[517] Ibid
[518] Interestingly it is Goldman's idea of process driven simulation that put ST in danger of collapse according to Davies.

Goldman introduces a third formulation of the threat of collapse through consideration of an objection to the above solution. Tacit theories might not be represented in the system but could still underlie the system's operations. If we accept that a theory can be tacit and sub-personal, it is clear that any simulation, including process-simulations, can be accounted for in terms of tacit grasp of a theory. This raises the question of how we define a tacit theory. Goldman considers Davies's suggestion that each separate proposition in a tacit psychological theory must correspond to a separate element inside the person and that these elements causally mediate between premises and conclusions. The causal pattern must duplicate the logical pattern of the relations within the tacit theory.[519] Goldman concedes that on this definition of tacit theory, simulation does indeed incorporate tacit theorizing. However even the originator of the theory now admits that the definition of tacit theory is too inclusive. Goldman argues that on a less inclusive definition there is a clear demarcation between TT and ST and no threat of collapse.

TT and ST Hybrids

Hybrid TT-ST theories claim that some specified components of our TOM abilities are underpinned by TT while and others are underpinned by ST. They come in a variety of formulations. A hybrid position may combine the possible processes of TOM acquisition in any way it sees fit.[520] Some hybrid positions will now be explored.

---

[519] Ibid P.32
[520] Carruthers and Smith 1996a PP. 4-5

1. Heal's Hybrid Position

Returning to the debate between Heal and Davies regarding the threat of collapse, having argued for an ST position that avoids collapse into TT, Heal then proposes the possibility of a genuine hybrid of both theories. We saw that, for her, simulating another's thoughts involves the capacity to think about the object of those thoughts together with some extra component. It is the character of this extra component that raises the possibility of a hybrid. On Gordon's conception of ST the extra is minimal and does not include extra concepts or a theoretical framework. However Heal suggests the extra could include a grasp of general concepts such as belief, perception, feeling, desire and action along with principles governing their interaction.[521] Although this gives ground to TT, Heal argues that there is no collapse. Predictions are reached on the basis of information about particular thoughts. Simulation is essential for this and it could not be achieved by a theory. It is not sufficient to know generalities such as beliefs and desires characteristically lead to actions. Heal is sceptical that a theory could specify how each particular belief p, q, r could lead to reliable predictions given desire S. Reliable predictions can only be achieved by entertaining the actual thoughts and thinking through their implications.[522] Earlier, for the sake of argument, Heal had allowed that complex intellectual abilities can plausibly be viewed as manifestations of a tacit theoretical understanding. She now argues that this assumption is in fact highly questionable.[523] It is open to the ST supporter to deny this assumption. The content of thought is

---

[521] Heal 1994 P.141
[522] Ibid P.142
[523] Ibid

229

immensely complex and variable and experimental work in TT has so far confined itself to fairly trivial cases of understanding others.[524]

Heal offers an account of our mind reading abilities that still favours ST while suggesting there are limitations to the simulation approach.[525] She suggests philosophers and psychologists should not oppose simulation to theory. They should instead ask what is the appropriate realm of each and about how they interact. Her view is that simulation must be central in dealing with the contents of others mental states. It is less relevant for dealing with non-content; which Heal thinks belongs in the domain of TT. By content Heal means the representational aspect of a mental state in virtue of which it specifies what the world is like or could be like, and in virtue of which its match with the world can be assessed.[526] Examples of content are specified in that-clauses when we attribute beliefs, desires, intentions, emotions to others.[527] Non-content refers to an aspect of a mental state that differentiates it from other mental states where this difference is not explained in terms of the different content of representations.[528]

Although content is supposed to be the domain of ST, Heal does not think that fully specifying the content of a mental state exhausts what can be said about it. This would imply content has a single mode of appearance.[529] She thinks an adequate

---

[524] Ibid
[525] Heal 1996 P.75
[526] Ibid P.77
[527] Ibid
[528] Ibid P.85
[529] Ibid P.77

explanation of our ability to predict other agents' psychological states will need to appeal to non-content to account for differences in propositional attitudes and to account for perception and sensation. Heal allows that we have theoretical knowledge about the mental states of other people. This theoretical knowledge is composed of generalisations about beliefs, perceptions emotions etc where these mental states are construed as broad categories. However she argues that these categories are insufficient for making predictions about others. These generalizations do not make direct predictions about particular cases.[530] Heal suggests non-content can be found both in the differences between propositional attitudes and in the nature of sensation and perception.

Heal offers arguments against the view that a pure TT could account for content. TT is unable to provide systematic insight into our responses in both usual and unusual circumstances.[531] We may have a general idea about the kinds of response that are likely to emerge from combining particular sorts of propositional attitudes. For example if we believe an agent desires some water, and that he also believes there is a glass of water in front of him, we can predict he will take a drink. This would count as usual circumstances. However if we add the unusual circumstance that the agent believes the water source may be contaminated this would not be the correct prediction. We could still gain the correct prediction by including the information about the unusual circumstance as a theoretical premise. Again however there may be further unusual circumstances of relevance. Perhaps the agent has already been

---

[530] Ibid PP.77-78
[531] Ibid P.83

vaccinated against the danger so it is safe to drink. Possible information about unusual circumstances is potentially limitless. TT would require a systematic theory of relevance. Heal thinks the magnitude and intricacy that such a theory would involve means we are unlikely to possess such theory even tacitly.[532]

2. Botterill and Carruthers' Hybrid Position

Botterill and Carruthers were originally influential advocates of TT. They now offer a restricted role for simulation in the enrichment of theory.[533] They concede that using one's own cognitive resources is sometimes the only way to mind read. As an example they suggest that if someone is asked what the president will answer when asked to name the capitol of Nebraska, their ability to predict his answer depends in part on the person knowing that the answer is Lincoln. The authors accept Heal's proposal for a hybrid but stress that the role of theory is fundamental in providing us with a conception of mental state types.[534] Botterill and Carruthers argue ST is utilised by TT for what they call 'informational enrichment'. By this they mean one can input another person's pretend beliefs into our inferential processing system. One can then attribute the outputs of this process to the other person.[535] Equally one can begin with a goal and work out, as if for oneself, the steps required to achieve this goal before attributing the other with those sub-goals.

---

[532] Ibid P.84
[533] Botterill and Carruthers 1999 P.89
[534] Ibid
[535] Ibid PP.89-90

According to Botterill and Carruthers there are three reasons an advocate of TT should permit simulation to play this role. Firstly processing a fully comprehensive theory of thinking is too big a task for a TOM subsystem to accomplish.[536] As Heal noted, inferences are often holistic. They will depend on which of their other beliefs they take to be relevant. For TT to be able to predict another's inferences would not only require knowledge about the agents other beliefs but also knowledge about which of these beliefs the agent considers to be relevant. Secondly the TT advocate must already allow that people posses a capacity to process inferences on the basis of suppositions because this is what hypothetical reasoning consists in. One supposes some state of affairs to be the case and then reasons from there. This allows one to see the consequences of a new belief or plan before deciding to adopt it. Thirdly if TOM really embodies a complete theory of thinking one should be capable of predicting thoughts others might have even though one's own reasoning processes might not enable us to reach these conclusions. In fact, however, one's internal capacities constrain the range of inferences one may assign to other agents.

Botterill and Carruthers draw the line at informational enrichment. They stress that TT should not concede any other role to ST. Simulation may proceed from already attributed beliefs to further beliefs or from already attributed goals to sub-goals. To remain intact TT must oppose any suggestion that ST has any role in the initial attribution of thoughts to others.[537] It must also deny that ST could have a role in providing conceptions of mental states. Finally TT must also resist any suggestion

---

[536] Ibid P.90
[537] Ibid P.91

simulation may be involved in prediction of action on the basis of intention or in prediction of intention on the basis of desire.

Slightly earlier work by Carruthers also paved the way for this hybrid position. Carruthers had already recommended limiting the scope of ST. He allowed ST may be useful in explaining our ability to attribute beliefs and desires to others, and perhaps to ourselves, in some cases. Nonetheless he argued simulation cannot provide the fundamental basis of our understanding of agency.[538] He also acknowledged that theoretical processing may be supplemented by simulation. This is because TT posits general theoretical knowledge of a kind that could plausibly be bestowed innately. This knowledge will not be content specific. Therefore, Carruthers argues, it clearly requires supplementation to produce fine-grained intentional descriptions and predictions.[539] He suggests there are only two options how this can be achieved. Firstly we could supplement the basic FP theory with a lot of more specific theoretical knowledge about how people with specific beliefs and desires will be disposed to think or act. The second option is to use simulation to generate predictions and explanations about the other. Carruthers suggests that we probably use a mix of both strategies and that simulation is at least sometimes employed. The first option would require a very extensive theory capable of capturing the inferential role of each particular concept. Carruthers argues that it is implausible either to suggest that we possess such a large volume of theory and even

---

[538] Carruthers and Smith 1996b P.22
[539] Ibid P.24

more implausible that it is innately given.[540] In contrast, Carruthers suggests, ST provides a simple alternative. One can rely on their own grasp of the concept in order to simulate the role it will play in the others mental transactions. Carruthers argued that the TT advocate should have no objections to assigning this role to ST as it leaves the central claim that we rely on a fundamental theoretical framework intact.[541] Simulation can be allowed a role in generating fine-grained intentional explanations and predictions, providing it does not attempt to replace the fundamental role of theory.

By way of showing that the role of TT is still fundamental, Carruthers advocates a fundamental role for theory in providing knowledge of one's own mental states. However he thinks it is better to understand the role of theory here by analogy with theory-laden perceptual knowledge of theoretical entities that we find in the sciences.[542] This is because Carruthers thinks it is implausible that self-knowledge is structurally equivalent to our theoretical knowledge of other people (achieved by making an inference to best explanation of behavioural data within the guidelines of a FP theory). At the conscious level beliefs about one's own mental state are non-inferential.[543] Carruthers argues that the alternative accounts of self-knowledge offered by ST are not plausible. Accounts like Goldman's take self knowledge for granted. Simulators are supposed to be able to access their own mental states. Carruthers argues such access is problematic on this type of account. There is a

---

[540] Ibid P.25
[541] Ibid
[542] Ibid P.26
[543] Ibid

235

question what it is exactly that the simulator is supposed to access. The state accessed cannot be construed as a state that normally fulfils a certain causal role or satisfies some theoretical description on pain of collapsing into TT.[544] The only possibility Carruthers thinks this leaves is a sort of quale or distinctive feeling. As well as noting that simulationists would not be happy with this conclusion,[545] Carruthers also argues that, in fact, not every type of mental state is accompanied by a distinctive feeling.[546]

Carruthers next tackles Gordon's alternative construal of ST. This version does not require introspective access to one's own mental states. Instead all that is required is an 'assent routine'. One uses one's practical reasoning system in pretend mode. The outputted representations made within this pretence refer to the other person. For Gordon a child begins by acquiring the ability to make assertions that express, rather than describe, its own intentions and desires. It next learns to prefix them with 'I believe' that P or 'I desire' that P. From these beginnings it learns to make more complex attributions of belief and intention to others such as 'A believes that P'. It also comes to understand the difference between standing and occurent beliefs (beliefs that are currently instrumental in the formation of behaviour). Carruthers objection is that none of this can help provide the child with access to his own occurent beliefs other than through drawing inferences based on its own recent behaviour.[547] He points out that thoughts often occur a while ahead of the actions

---

[544] Ibid P.29
[545] Ibid
[546] Ibid P.30
[547] Ibid P.35

which they rationalise. Simulation will not be capable of making attributions of thought to oneself that are even close to simultaneous with the thoughts themselves. Furthermore those thoughts that do not result in actions are permanently beyond the reach of simulation.[548] Because ST cannot plausibly account for self-knowledge the only defensible form of the theory will be one guided by TT.[549]

3. Stich and Nichols Hybrid position

Two more key advocates of TT have since moved to a hybrid position. Stich and Nichols now advocate a hybrid theory. Though originally advocates of TT they now claim that neither modular versions not versions that appeal to a theory-like information base can adequately account for our mindreading abilities because simulation-style processing is also essential.[550] They argue that mind reading involves a variety of processes. Some of these make use of information-rich theory, others involve simulation, and some processes do not fit either classification.[551] No monolithic theory is capable of explaining the full range of facts.[552]

Stich and Nichols are critical about the original setup of the TOM debate. In this debate both ST and TT were conceived as polarised monolithic theories (the authors include their own earlier work).[553] They suggest that the main contrast between the positions has been between information-rich and information-poor processes. They identify two unfortunate consequences of construing the TOM debate in this

---

[548] Ibid P.36
[549] Ibid P.37
[550] Stich and Nichols 2003 P.154
[551] Ibid P.212
[552] Ibid P.211
[553] Ibid P.212

manner.[554] Firstly this construal ignored the possibility a hybrid of the two positions might be correct. Secondly simulation was understood in extremely broad terms. It included a variety of disparate processes. Stich and Nichols suggest this is partly because ST was construed as the adversary of TT. Strictly speaking they think the term simulation ought to be retired from the debate because it is used in such a diverse number of ways that it cannot be said to pick out any natural or theoretically interesting category.[555] In advocating a hybrid, they limit their own usage of simulation by taking pretence driven off line simulation as prototypical.[556]

There are two key areas in which simulation figures in Stich and Nichol's modified account of mind reading. Firstly it underlies inference prediction.[557] They claim that a good theory of mind reading needs to explain our remarkable ability to predict the inferences other people will make.[558] To understand Stich and Nichol's explanation of this requires a brief diversion in order to familiarise ourselves with some new terminology they introduce into human cognitive architecture.

Standard cognitive architecture maintains that the mind contains two types of representational states; beliefs and desires.[559] These two types of state are functionally different from one another. They have different causes and different pattern of interaction with other components of the mind. This architecture also

---

[554] Ibid P.132
[555] Ibid P.134
[556] Ibid
[557] Ibid P. 135
[558] Ibid P.104
[559] Ibid P.13

posits a practical reasoning system with the functional role of monitoring the

evolving stock of desires and goals in order to notify a 'planner' when to construct a

plan.[560] The practical reasoning system then consults the desire system for desires

that are incompatible with the execution of this plan. If any are found the plan is

rejected and the planner is asked to construct another plan. If no incompatible

desires are found then instrumental desires are generated for each intermediate step

that will be involved in the plan's execution.[561] Stich and Nichols also claim that

propositional attitudes are relational states. Having a belief or desire with a

particular content involves having a representation token with that content. For

example having the belief that Socrates was an Athenian consists in having a

representation token with this content stored in one's 'belief box'.[562] Similarly

desiring state x involves having a representation token state x in one's 'desire box'.

The authors also posit a 'possible world box'. This contains representation-tokens

which in this case have the function of representing what the world would be like

given certain assumptions. We do not have to believe these assumptions to be true or

desire that they be the case.[563] The possible world box is an area used by our

cognitive systems to construct and store representations of possible worlds.

Stich and Nichols posit a subcomponent of the inference mechanism called the

UpDater.[564] Often an agent will acquire new information or adopt a new belief that

is either directly incompatible with his current beliefs or with an entailment of these

---

[560] Ibid P.14
[561] Ibid
[562] Ibid PP.14-15
[563] Ibid P.28
[564] Ibid P.30

239

beliefs.[565] This new information may be acquired either through perception, through inference or from the reports of others. Stich and Nichols hypothesize that a cognitive mechanism, or cluster of mechanisms, which they call' the UpDater' underlie this process. The inference mechanism involved in the formation of real beliefs, including the UpDater subcomponent, is also involved with representations in the possible world box.[566] When a pretence premise is introduced into the possible world box, the UpDater treats the contents of this box in the same way it treats the contents of the Belief Box when a new belief is introduced. The UpDater scans the representations and eliminates or modifies any that are not compatible with the pretence premises.[567] This explains how cognitive systems identify which beliefs require modification in the light of new beliefs.

To return to the issue of how ST underlies inference prediction, Stich and Nichols claim this ability is explained by the fact that our predictions about other people's inferences are underwritten by the same mechanisms that underwrite our own inferences. Inference mechanisms operate on representations in a possible world box in the same manner they operate on representations in the belief box.[568] Stich and Nichols argue that advocates of what they refer to as scientific TT have not been able to offer an explanation of our success in this area. An appeal to a strong theory about how other people reason would be extravagant given that the task could easily

---

[565] Ibid P.30
[566] Ibid
[567] Ibid P.32
[568] Ibid P.104

be accomplished using inference mechanisms themselves.[569] As well as considerations about simplicity, Stich and Nichols think considerations about the accuracy of our ability for inference prediction also weigh in favour of ST.[570] We demonstrate an ability to make successful predictions about a vast range of cases including cases that bear little resemblance to those we have previous experience of. Our success in other types of mindreading is comparably poor. This becomes less puzzling if one accepts that people use their own inference mechanisms in the case of inference prediction. The inference mechanism*s* involved in the construction of beliefs in the belief box operate in a similar manner on representations in the possible world box.

We now come to the second key area where ST features in Stich and Nichols account of mind reading. This concerns the role played by the Planner in generating predictions about how others will attempt to satisfy their desires, and about which instrumental desires are generated.[571] According to some off line accounts of simulation, the practical reasoning mechanism is provided with inputs by a pretend belief and desire generator. Although Stich and Nichols do not posit anything that would correspond to a pretend desire generator, the information in the possible world box which the planner uses could be seen as an elaboration of the idea of a belief generator.

---

[569] Ibid P.104
[570] Ibid P. 105
[571] Ibid P. 135

Default belief attribution is another area where Stich and Nichols think scientific TT cannot accommodate the facts and must make room for ST. Strictly speaking, default belief attribution is not compatible with their own strict criteria for off line ST. This process is not off line and it does not employ non-standard inputs.[572] However Stich and Nichols acknowledge the influence of Harris in showing that this could be a kind of simulation.[573] They argue that this ability explains how adults can quickly and correctly attribute large numbers of beliefs to others without apparent evidence. A mind reader constructs a model of another person in their possible world box. In doing this she initially includes many of her own beliefs.[574] Stich and Nichols argue that there is no plausible explanation of default belief attribution in terms of theoretical processes and that none have been put forward by scientific TT. However they also argue that default belief attribution is not sufficient to account for the accuracy of belief attribution in adults. It works alongside a theoretical strategy and a second simulation strategy. The theoretical contribution consists in a cluster of strategies for attributing discrepant beliefs to the model of another person and a mechanism for clamping these beliefs. This clamping clears the path for the second simulationist contribution; it ensures these beliefs will not be removed by the UpDater when this removes all default attributed beliefs that are incompatible with the discrepant beliefs. The role of the UpDater should be interpreted as a simulationist strategy, even on Stich and Nichols restricted definition because it also plays a role in updating the mind reader's own beliefs and in hypothetical reasoning.

---

[572] Ibid P.140
[573] Ibid P.133
[574] Ibid P.106

Stich and Nichols claim other versions of ST also require theory to play a role. For example Goldman suggests that a simulator can utilise another's perceptual situation in order to infer the other has certain experiences or beliefs which match the ones the simulator would have in that situation. Stich and Nichols claim this would have to be an information-rich process rather than an information-poor one.[575]

4. Goldman's hybrid position

Goldman also argues that while traditionally TT and ST have been taken to be in opposition with one another, they need not be understood to be in competition.[576] He suggests that there are a number of ways in which the two strategies might be combined. The most obvious possibility, given our previous discussion is that theorizing could implement simulation. Goldman distinguishes two ways this could occur. Simulation could be identified with a particular process and it can also be conceived as a higher-order control process overseeing a number or lower-order processes. The higher-order process, the lower-order processes or both might be implemented by a tacit theory.[577] For example the control process might rely on a theory that specifies how to select pretend states, which psychological mechanism will act on these states, and how the pretend states are to be inputted into this mechanism.[578] We saw above that it was suggested that if simulation is implemented by theory there is a threat of collapse. Goldman argues that even in this type of case one can make a case that simulation is still valid providing all third-person

---

[575] Ibid PP.133-4
[576] Goldman 2006 P.43
[577] Ibid
[578] Ibid P.44

243

mindreading is argued to involve simulation. Here it would still be the case that no mindreading was possible without simulation.

A second possibility Goldman considers is that simulation and theory processes co-operate with one another. He gives two examples of how this could occur. Firstly, when attempting to predict another person's decision, theoretical reasoning may be employed to infer their initial states in order to select pretend inputs for simulation. A simulation routine is subsequently performed using these pretend states. Secondly a 'generate and test' strategy can be used to infer the prior psychological state of an individual based on observation of a later psychological state. In this case theory is used to generate hypotheses about the individual's prior state or states. Subsequently one simulates being in these hypothesized states to test whether the output matches the individual's current state.[579] Though Goldman's sympathies are with ST he suggests this kind of case provides compelling reasons to adopt a hybrid position. In this type of case we have to work backwards to mind read whereas our psychological mechanisms (which underlie simulation as well as for our own psychological experiences) only work forwards. This makes pure simulation unfeasible for this type of task.[580]

Goldman also offers a final way in which we can conceive a combination of TT and ST. Theoretical processing and simulation could occur independently of one another. Some types of mind reading would be executed entirely by simulation mechanisms

---

[579] Ibid P.45
[580] Ibid

while others would be executed entirely by theoretical processes.[581] Neither process

would implement or cooperate with the other. On this conception it could be that

certain types of mental states are dealt with entirely by either TT or ST.

Alternatively it could be that certain types of evidence are always exploited by one

of the two theories.

Having advocated a hybrid approach in place of the traditional opposition between

TT and ST, Goldman suggests we need to reframe the TOM debate. He suggests it

should be reframed as a debate between ST and 'simulation-neglecting TT'.[582] A

consequence of adopting this new framework would be that it would no longer be

sufficient for TT to establish that theorizing is involved in mind reading; TT must

also establish this is not implementing simulation. Another consequence is that ST

does not have to establish the total absence of theorizing in mind reading. However

it needs to do more than just establish that theorizing is not the main method of mind

reading. It must provide evidence of simulation as defined by what Goldman refers

to as its positive aspects.

Goldman distinguishes between positive and negative approaches to defining ST.[583]

Negative approaches construe ST in terms of what it denies; the mind reader's

supposed belief in a set of folk psychological laws. Positive approaches construe ST

in terms of the assertions it makes. There are two central assertions. Firstly ST

asserts that pretend states are employed in mindreading. Secondly it asserts that

---

[581] Ibid
[582] Ibid P.46
[583] Ibid P34

mind reading involves the operation of the same processes or mechanisms that are employed by the target.[584] Whereas negative approaches to defining ST construe any evidence of the involvement of theory as a threat, positive approaches see evidence of theoretical operations as compatible with ST. Positive approaches leave open the possibility that these theoretical mechanisms are responsible for the implementation of simulation routines.[585]

Craruthers has objected to Goldman's way of reframing the TOM debate. He argues it is biased in ST's favour. ST wins if it can be shown to have any role in mind reading. Conversely TT and Modular approaches both lose if mind reading turns out not to be entirely theory driven or module driven. Carruthers argues that this is unreasonable given that many TT advocates now acknowledge some role for simulation. Goldman has a very loose notion of simulation and a very strict notion of theory.[586] However much of the data Goldman appeals to in support of his position could equally support a moderated version of TT.

Carruthers suggests the TOM debate should be understood as a debate about which position is more central and whether either is more fundamental than the other.[587] By way of example, he considers cases of mindreading where the operation of the mirror systems is shown to be involved in recognizing others emotional states on the basis of facial expressions. On Goldman's modified construal of the TOM debate

---

[584] Ibid
[585] Ibid
[586] Carruthers 2006b
[587] Ibid

this kind of data provides support for ST and evidence against TT. Carruthers argues

this type of data is actually compatible with a form of TT that allows a role for

simulation while maintaining that core concepts involved in mindreading are

information-rich and either innate or a product of theorizing. He suggests mirror

systems might have evolved prior to mind reading facilitating either emotional or

imitative learning.[588] Mirror systems would then be available to be appointed to

mind reading tasks as mind reading capacities evolve. Identifying others emotions is

a demanding task. It would therefore be unsurprising to discover conceptual

representations of emotion activated either by experience, or by MN induced

experiences of the same emotion. Such a finding does not entail that the recognition

of one's own emotional experiences are primary. It is compatible with the idea that

their identification relies on a sizeable stock of learned or innate information about

emotions and their causal relations. Carruthers also notes that Goldman's own

account of ST has to concede a crucial role to theory in places.[589] This account

conceded that the selection of inputs for simulation must be guided by theory. It also

conceded theory is used to generate hypotheses about the individual's prior state or

states when executing a 'generate and test' simulation strategy.

Carruthers picks out what he interprets as the distinguishing mark of Goldman's

Hybrid position. He claims that first-person knowledge of mental states occupies a

central place in Goldman's hybrid account. The simulator's awareness of his own

mental states is prior to awareness of the mental states of others and plays a

---

[588] Ibid
[589] Ibid

foundational role in mindreading.[590] Carruthers objects that an awareness of one's

own states can't play this foundational role unless this awareness by itself can

provide the simulator with a substantial body of theoretical knowledge concerning

the causes and interactions of these mental states. After all, Goldman has conceded

that such knowledge is indispensible for prediction and retrodictive explanation of

others. One just has to allow that this theoretical knowledge is either innate or

acquired by theorizing and the distinction between Goldman's ST and versions of

TT that include a role for simulation collapses. Carruthers next argues the fact

people unknowingly fabricate explanations of their own behaviour counts as direct

evidence against the possibility that first-person awareness could be sufficient to

provide this body of evidence. These fabrications possess the same immediacy and

introspective obviousness as normal explanations of one's own behaviour. A second

objection Carruthers makes is that if everyone initially learned about the causal roles

of mental concepts from their own case, there would be a lot of variation concerning

the time it takes people to master the theoretical knowledge required for mind

reading whereas everyone acquires this skill at about the same age.[591] We would also

expect children to acquire the ability to attribute mental states to themselves before

they learn to attribute them to others. Again, Carruthers argues, this contradicts

empirical findings.[592]

---

[590] Ibid
[591] Ibid
[592] Ibid

5. Perner's Hybrid Position

Perner also advanced a proposal to distinguish simulation from tacit theory. This proposal was motivated by a desire to avoid the threat of collapse.[593] He thought that although the distinction between TT and ST appears intuitively obvious it is surprisingly difficult to list their differential traits. He proposed simulation should be defined as predication-implicit knowledge about the mind. He argued that every use of theory exploits this kind of knowledge. Perner contrasts what he sees as true simulation with reasoning by analogy. Reasoning by analogy deals with mental states in full propositional form.[594] Simulation is only concerned with mental states in a minimal sense of concern. It concerns itself with mental states in the sense of taking them off line. The mental states involved in simulation belong to the simulator. This fact is not represented. However it is implicit in the fact that it is the simulator who has them.

Perner argues any use of TOM that makes use of predication-implicit knowledge about the domain necessarily includes an element of simulation.[595] The theory makes tacit use of the fact that one's mechanism for judging content is relevantly similar to the other persons. A foundation of simulation underlies theoretical tasks. This keeps the theory manageable because it contains prediction-implicit knowledge about the other persons mind.

---

[593] Perner 1996 P.90
[594] Ibid P.94
[595] Ibid P.96

249

However Perner argues that this concession to ST is not enough to establish a pure simulation.[596] Perner argues that empirical evidence suggests that the use of predication-implicit knowledge cannot do all the work in explaining our mind reading abilities. Evidence points to a consistent theoretical bias in mental state attributions. One example Perner appeals to concerns data collected by Wimmer et al for a phenomenon they characterised as 'inference neglect'. Young children do not understand that others can know something by inference alone in the absence of perceptual accesses to the information. For example, in the company of an observer, children were shown that a box contained only one kind of object. Next an item is taken from the box and placed in an opaque bag. This is done while shielding the object from both the child and the observers perspectives. The child is asked both whether he knows what is inside the bag and whether the observer knows what's inside. The child correctly answers that he knows what is inside the bag (an easy inference as the box only contained one kind of object) but incorrectly answers that the observer does not know. Perner argues that if simulation is the sole strategy we have for mind reading we would expect the child to produce the same answer to the question about the observers knowledge that they produced about their own knowledge.[597] He concludes that the only viable position is a hybrid one.

---

[596] Ibid
[597] Ibid P.100

Are hybrids theories coherent?

The term hybrid theory is slightly misleading because advocates usually place their faith in one position as doing the real work. Proponents still maintain allegiance to their original theoretical starting points while conceding as much to the rival theory as they deem necessary. Hybrid versions of TT and hybrid versions of ST will be treated separately. We will consider whether they make any significant alterations to the way TT and ST construe their answers to questions raised in the introduction.

Hybrid versions of TT

TT's answer to question 1 claimed we are utilising a body of psychological information to explain and predict the behaviour of others. Hybrid TT expands the idea of what counts as theoretical information. For example both Stitch & Nichols and Davies expand their definition of theoretical information to include any body of psychological information even if it is not functionally specified in terms of law-like relations. This information we are using can have a non-explicit format which need not be amenable to explicit theoretical formulation.

A hybrid answer to 1 can claim that simulation can be used in explanation and prediction, but that the psychological states themselves still derive their meaning from their position in a theoretical framework. Simulation may be employed as a heuristic device. Though we may use simulation to mind read; psychological states are defined by their position within a theory. Simulation may, for example, helps to

251

identify which information an agent considers relevant. Simulation is only used to enhance a theoretical process. It supplements our general understanding of psychological activity with context specific knowledge to produce more nuanced explanations. It is not involved in the initial attribution of psychological states to others. The type of simulation theory theorists have in mind when advocating a hybrid are the off line versions. Stich & Nichols claim simulation is employed as a tactic to predict the inferences others will make. We use the same machinery we employ when making inferences ourselves. It also enables predictions about other agent's desires. However it must operate in conjunction with theoretical strategies.

It is not convincing that assigning simulation the role of a heuristic device is really enough to create a genuine hybrid position, especially if ST is not involved in the initial attribution of psychological states. We are not really using our own psychological apparatus to save having to make theoretical inferences. The spirit of the original simulation theory has been lost. It is also unconvincing that an appeal by ST to a theoretical assumption of similarity of another person to oneself forces the ST advocate into a hybrid TT-ST position. It is not the relevant type of theoretical assumption.

Discussions of hybrid positions by Heal and Goldman also raise questions about what the appropriate realm of each component is and about how they interact. There may be significant differences between the possible ways of combining the theories. Goldman points out that it is possible for a hybrid theory to construe the

252

components as working autonomously from one another.[598] However this kind of

hybrid would threaten to undermine a key intuition supporting TT suggested by

Stich and Nichols. This is the intuition that mind reading abilities exemplify just

another instance of reality's susceptibility to explanation through employment of the

dominant explanatory strategy; through employment of an internally represented

knowledge structure.[599] A hybrid position positing co-existing but autonomous

simulation mechanisms would posit mechanisms that are not capturable in this

explanatory approach. ST components would therefore signal a deviation from the

dominant explanatory strategy, rather than mechanisms of implementation. Stich and

Nichols now concede that no monolithic theory can account for the full range of

facts about mindreading.[600] They also allow there will be processes that do not fit

either classification (though they do not really explore this avenue).[601] Botterill and

Carruthers also concede that processing a fully comprehensive theory of thinking is

too big a task for a TOM subsystem to accomplish.[602] However Carruthers still

argues TT is more fundamental because it provides knowledge of the mind reader's

own mental states.[603] However by theory here he means something comparable to

theory-laden perceptual knowledge so only maintains the fundamental statue of TT

by employing a weak notion of theory.

---

[598] Goldman 2006 P.45
[599] Stich and Nichols 1992 P.35
[600] Stich and Nichols 2003 P.211
[601] Ibid P.212
[602] Botterill and Carruthers 1999 P.90
[603] Carruthers and Smith 1996b P.26

The expansion of what counts as theoretical information undermines a key starting point of the original theoretical position. This was the idea that a grasp of psychological states consists of a clearly defined theoretical body of knowledge about functionally related psychological states. This loosening of what is to count as theoretical information sits uneasily with hybrid TT's allegiance to the original formulations of TT.

The discussions of Heal in this chapter and the previous one also provide ammunition for a more radical objection to hybrid TT. Heal is opposing the idea that psychological content is internally mapped in the way TT must presuppose. There is no need to appeal to an isomorphic inner structure or separate bodies of knowledge. On this externalist picture any hybrid appeal to an off line collaboration between theoretical information and simulation is misguided because an explanation of what we are doing when understanding others does not need to reference internal goings on. Instead what is required is an ability to understand how a given attribution of a psychological state to someone will affect the rest of their thoughts and behaviour. Our attention is directed to a shared intersubjective space not to inner occurrences. This can be the case because thinking about another's psychological states does not involve referencing a theoretical body of knowledge. It is a particular application of one's capacity to think about the content of that psychological state. For Heal, content is not internal, it is part of an intersubjective world. However, that the content of a psychological state is internal is part of the framework of ST. If we are not trying to solve a problem of how to access another's internal content then the

theory becomes redundant.

Hybrid versions of ST

There are variations on the answers hybrid versions of ST could give to question 1. One answer would be that we are employing simulation as a default but occasionally allowing it to be overridden by theoretical processing. Another would be to say that while simulation enables us to attribute some kinds of psychological states, for example emotions or propositional attitudes, other attributions require theoretical strategies. A tighter hybrid position would claim that both strategies are working in conjunction with one another. On this picture theory may be used in the final part of the simulation strategy. This may include an understanding of general psychological concepts and the principles dictating their interaction. The theoretical assumption that others are like oneself is sometimes understood as a concession to TT. However it is more specific information relating to an individual that really drives psychological attribution. Gordon suggests we are making use of a body of theoretical information within the context of practical simulation. If we attempt to employ this information too rigidly regardless of context it will be unreliable. We use simulation to guide the application of psychological rules.

Like Hybrid TT, Hybrid ST retains allegiance to its parent theory. Again however the concessions Hybrid ST makes to TT sit uneasily with the original aims of ST. A virtue of ST was that we had no need to resort to complex theoretical processes as

we simply use ourselves as a model. Resorting to theoretical processing at any stage of the simulation process undermines this virtue. Hybrid theories also lose the virtue of mutual exclusivity.

How do we unpack the empirical and conceptual elements in these revisions?

This examination of the TOM debate has uncovered difficulties stemming from the sheer variety of different formulations associated with each position and also with hybrid positions. Furthermore definitions of TT and ST have been reshaped not only by advocates of these respective positions but also by advocates of their rival positions. In such cases this often appears to be done in a way that would best suit the rival advocates own agenda. Nonetheless opponents of both TT and ST have been as influential as advocates in defining the positions of their rivals. Hybrid positions, rather than helping to settle what is meant by theory and simulation have brought out further ambiguities with the terminology. For example they have brought the ambiguous nature of the term 'theory' into sharp relief. In fact examination of hybrid positions reveals this trait becoming more pronounced; for example both Goldman and Stich and Nichols try to reframe the TOM debate in ways that will be advantageous to their original allegiances.

<u>What counts as empirical/conceptual?</u>

A related difficulty is that the rise of hybrid theories has altered the chief concerns of the debate. Now that both TT and ST camps usually allow a role for the rival, the central question is no longer which position receives best empirical support but which position can be assigned a more fundamental role in executing mind reading tasks given that there is empirical evidence for both (both Heal and Carruthers recommend this). This is less clearly an empirical question as it is not a matter of whether we find evidence for the existence of a process or not. It is a more conceptual issue as the way one views the evidence will be framed by one's philosophical allegiances. One's evaluation of the empirical evidence will be strongly led by one's theoretical preconceptions and the interpretive framework one is working within. Taking for example Heal's hybrid position, there seems to be room for different intuitions about whether a situations immanent contextual content or the general framework in which it must be placed has more fundamental importance for mind reading. Someone sympathetic with Heal might accept simulation is confined to informational enrichment (as Botteril and Carruthers suggest) without seeing this as privileging the TT position. They just have to understand informational enrichment as central to mind reading.

Similarly, concerns with defining the positions in such a way that they exhaust the possibilities in the field, and preserving the mutual exclusivity of the rival positions, seems a conceptual task rather than a strictly empirical one. As well as blocking the

formation of alternative theories there is also a vested interest in redefining one's

favoured theory in such a way that it will be as empirically robust as possible. For

example Goldman framed the debate in such a way that TT must not only establish

the involvment of theoretical mechanisms, but must also establish these mechanisms

are never involved in implementing a simulation process. ST on the other hand only

has to establish the involvement of simulation. Empirical evidence of theorizing will

not threaten ST providing it can be established that this is not the dominant method

of mind reading. Stich and Nichols advocacy of a less rigid notion of theory appears

expedient if TT and ST are to fill all the space in the TOM debate.[604]

Although there are many conceptual issues, the TOM debate is often framed as an

empirical one. Goldman and Stich and Nichols construe the TOM debate as

primarily pitched at a sub-personal level of description involving rival accounts of

the relationships between pieces of information-processing machinery.[605] However;

even the experimental data taken to be paradigmatic of one position has been re-

appropriated in support of its rival. For example in previous chapters we saw that the

results of Wimmer and Perner's false belief task cited in support of TT by the

authors have also been utilised in support of ST by Gordon. Another objection to the

empirical evidence for TOM offered by Heal is that this evidence is confined to

fairly trivial cases of mind reading which do not allow the complexity of our

thinking about others to be explored.

---

[604] Davies and Stone 2001 PP.3-4
[605] Ibid P.5

Davies has argued that it is difficult to find anything decisive between TT and ST in the developmental data.[606] He considers a further criticism of Wimmer and Perner's false belief task used as empirical support for TT put forward by Harris. Harris claims the false belief task was designed with the intention of making it impossible to succeed employing a basic simulation strategy.[607] Davies explains that this basic strategy, termed 'total projection' by Gordon, makes no adjustments for relevant differences between simulator and target. Davies argues the fact more sophistication is required for successful simulation does not arbitrate between theory and more sophisticated simulation involving for example development in imaginative flexibility rather than TOM.[608]

First-person cases of attribution initially appeared a more promising area for empirical arbitration between TT and ST. Davies argues that for TT, first-person cases are just like any other cases whereas for ST first-person attribution is unproblematically available to the subject.[609] TT should therefore predict no differences in the level of error between first and third-person attribution while ST should predict that third-person attribution will be less reliable. In Support of TT, Gopnik and Wellman appealed to evidence of an absence of asymmetry between first and third-person cases. Children who fail on a third-person version of the false belief task also fail a first-person version. At around the same time, in support of ST, Harris appealed to evidence in favour of asymmetry. Young children display

---

[606] Davies 1994 P.200
[607] Ibid P.198
[608] Ibid
[609] Ibid

259

much greater accuracy reporting what they are currently thinking, pretending, seeing, and wanting.[610]

A big problem here is that if the various pieces of empirical evidence can be accommodated by different theories it cannot still count as empirical evidence. Though proponents of both theories write as if the debate is an empirical one the issue is much less straight forward.

Where do the revised theories stand in relation to difficulties raised in previous chapters for parent theories?

This chapter will now look at whether these hybrid positions offer any progress in relation to any of the problems raised for these positions in the two previous chapters.

<u>Key issues facing the TOM debate</u>

First it will be useful to remind ourselves of the key issues that have been raised for ST and TT in the previous chapters:

- The central problem was that both theories are working with a picture of the mind as inner. Attributing psychological states to others is a matter of

---

[610] Ibid P.199

attributing unobservable inner states that causally affect the behaviour we
observe.

- Another problem which was raised for both theories was their
phenomenological implausibility if taken to describe personal level
processes. In both cases this led to the theories being relocated at a sub-
personal level. However the continuing appeal made by both theories to
personal level psychological concepts is problematic.

There were also specific problems raised for each theory:

Problems for TT

- A key issue was how psychological terms derive their meaning. TT avoids
opening itself to the problem of other minds by claiming psychological
states get their meaning from their positioning in a public theory. However
this move opens them up to another problem; it is unable to capture the
distinctive manner in which one's own psychological states are revealed.

- TT was also criticised for working with a problematic picture of what
constitute mental states as unobservable theoretical states whose meaning is
defined by the functional position they occupy in a causal explanatory
theory. This picture was not philosophically plausible.

- Another objection was that the regularities we normally take to explain our

behaviour described in everyday psychological language (he was angry about the betrayal so he kicked the door) are claimed to be superficial and behaviour is actually dictated by structural and functional features of theoretical states apprehended in a vocabulary that is different from the one used at the personal level which we have no access to. It is not clear exactly how this language of sub-personal processes is supposed to relate to our ordinary psychological vocabulary of personal level states such as beliefs, desires, and perceptions which we employ in explanations about others.

Problems for ST

- ST on the other hand assigns a privileged role to first-personal experience in assigning psychological states their meaning. This raised an issue how psychological concepts can be coherently applied to other agents.

- The idea of unattached representations that featured in sub-personal formulations were implausible and also in conflict with the standard simulation process.

Hybrid TT

Hybrid versions of TT enable the theory to deal with specific objections to TT raised above. A problem noted with TT's answer to 1 was that it was not philosophically

262

neutral. It involves a particular picture of what philosophical states are; theoretical inner states whose meaning is fixed by their place in a causal explanatory theory. Hybrid TT partially deals with this objection. It allows that these states may require 'informational enrichment' using context specific information but the setup is not fundamentally changed. ST does not play a role in giving psychological states their meaning.

Another problem was that the personal level regularities we normally take to explain our behaviour are claimed to be superficial and behaviour is actually dictated by regularities holding between internal theoretical states. Hybrid TT concedes that more specific information about an individual drives our ability to attribute psychological states. For example information about particular beliefs or goals a person may have.

A more general problem was that this answer lacks phenomenological plausibility as a personal level characterisation of what we are doing. Hybrid versions of TT are less susceptible to this criticism. Most hybrid TT denies that it is a personal level process. Instead we employ a tacit theory. This theory need not be articulable. Some hybrid formulations of TT widen the definition of theory to include any body of psychological information whether or not this information has the structure of psychological laws. However it is not clear that one should allow a hybrid TT to make this move. It abandons something fundamental to TT: that the meaning of psychological terms is fixed by their role in a functional theory. To establish a role

for TT it is not enough to point to any theoretical information that may be involved. To say the rules are not accessible is one thing – to say they do not have a rigid, potentially specifiable structure (e.g. stored in a linguistic format) is another. As Heal argues, explicit theories must be paradigmatic of theory.

The central issues raised in chapter two still stand. Hybrid versions of TT are still working with a picture of the mind as inner. They still assume it is coherent to move between personal and sub-personal levels of explanation. Sub-personal states are treated as analogous to personal level states.

Hybrid ST

In the previous chapter it was argued ST lacks phenomenological plausibility as a personal level process. Low frequency hybrid ST theories would allow that simulation is only an occasional strategy which would be acceptable as part of a phenomenological description of what we are doing. However, in this case, ST is not really an answer to question 1 anymore but only a component in TT's answer to question 3. This point also applies to hybrids in which simulation serves as a theoretical shortcut only involves simulation in an answer to question 3 not to question 1.

Another problem was that for ST meaning was anchored in 1st personal experience. This need not be the case for hybrid ST. it may grant a role to shared concepts and

interrelating principles. These can be utilised in the course of the simulation process. Like TT, ST also faced the problem that it lacks phenomenological plausibility as a personal level characterisation of what we are doing. As with hybrid TT, Hybrid ST does not specify that simulation be a personal level process. However a personal level combination of the two processes would have little more phenomenological plausibility than either strategy alone.

Hybrid theories offer small improvments in relation to some problems raised for each theory. In particular these improvements concern problems raised which were specific to each theory. However most of the general difficulties raised for the TOM debate still stand. We do not have a phenomenologically satisfying account of what is involved in attributing psychological states to others. In light of the fact that these difficulties have not been overcome we will turn to an alternative account – Direct Perception to see if these difficulties can be avoided. This position is offered as an alternative to TT and ST. However it marks a departure from these rivals in that it does not picture psychological states as hidden inner states. I shall suggest that it enables a more satisfying picture of how these states are revealed to us.

5 Direct Perception: Gallagher and Merleau-Ponty

Because this thesis has raised problems with the TOM approach, such as the philosophical picture of the mind as inner informing these positions and their phenomenological implausibility, it will now turn to alternative accounts of our relationship to the psychological states of others to see if these problems can be avoided. This chapter will examine two accounts of intersubjective perception offered by Gallagher and by Merleau-Ponty which argue we have a direct perception of other's intentional states. Both appeal to empirical evidence from developmental studies to support their case. However both accounts also draw on the phenomenological tradition. Phenomenology is the study of the structure of our experience. It involves in depth scrutiny of the way things appear in experience. This chapter will be divided into two parts. The first part will begin by looking at critical remarks Gallagher makes about ST before exploring Gallagher's account of direct perception and the implications Gallagher draws for the TOM debate. Gallagher's account of direct perception seeks to offer an empirical alternative to TT and ST. In one sense, Gallagher can be understood as continuing the debate of the previous chapters but he is also introducing an entirely new approach (although features of this approach are anticipated in claims made by certain proponents of ST, for example in claims that psychological representations denote shared phenomena that have their basis in attuned intersubjective encounters). We saw that ST's attempts to align themselves with direct perception of psychological states were incompatible with core values of simulation. However in Gallagher's case these

insights cohere more neatly with his philosophical position.

However I will suggest that Gallagher does not consistently position himself far enough away from the philosophical picture motivating TOM. Gallagher's position sits somewhere between being *a rejection of the epistemological problem* TOM approaches confront, about how we can know the intentional states of others and *a solution to it*. I will consider which of these potential readings of Gallagher's work is most justified. One reason for interpreting Gallagher as offering a solution is that his account of direct perception is presented as an alternative to two rival attempts to solve this epistemological question; TT and ST. I will suggest that there are two ways in which we can construe Gallagher's account. The first way is to understand direct perception as a third account of how we can access other agent's intentional states sitting alongside TT and ST. This involves taking seriously the idea that there is an epistemological question to be answered; on what basis are supposed claims to know about the psychological states of others justified? The second way to construe Gallagher's account of direct perception is as suggesting that there is no question to be answered. On this reading the problem of other minds is not a problem but a pseudo–problem. In other words what Gallagher offers is not a solution to the problem of other minds; it is a rejection of the problem. Gallagher claims "the direct perception approach *comes close* to suggesting that there is no problem of other minds "(my italics).[611] This raises an ambiguity as to which of the above construals of direct perception should be attributed to Gallagher. I will suggest that elements of both construals can be found in Gallagher's writing. This chapter will begin by

---

setting out Gallagher's account of direct perception and interaction theory before examining this ambiguity more closely.

In part two we will turn to Merleau-Ponty. Gallagher's account of direct perception is heavily influenced by Merleau-Ponty, particularly Merleau-Ponty's account of bodily expression. However there are important methodological differences between the two. Merleau-Ponty's work offers a clear rejection of the epistemological problem about knowing the intentional states of others.

Part 1

<u>Critical remarks on ST</u>

Before assessing Gallagher's position it is worth looking at some implications he draws for the TOM debate, particularly for ST which relate to aim C of this thesis. We saw in chapter three that some ST supporters accused TT of running an argument from analogy whereas they thought their own position was immune to this accusation. Gallagher and Zahavi make a strong case for viewing the mindreading process adopted by ST as an example of an analogical argument. They characterise the argument from analogy as an inferential process taking us from observed observable behaviour to a hidden mental cause.[612] Gallagher and Zahavi draw a distinction between what they call explicit and implicit versions of ST. Explicit versions correspond to standard ST. They involve consciously simulating being in

---

[612] Gallagher and Zahavi 2008 P.181

another's situation in order to work out how they must feel and then attributing the results of this exercise to the other person. Implicit ST conceives of simulation as a sub-personal process executed by neural mechanisms operating below the level of awareness. Gallagher and Zahavi's argument focuses on similarities between Mill's position (described in the introduction) and explicit ST. Gallagher also argues that the structure of implicit versions of ST are derivative from their explicit counterpart. This suggests that implicit ST should also bear structural similarities with Mill's strategy.

Gallagher and Zahavi draw comparisons between the explicit ST position and Mill's position. One point of agreement is that both positions assume us to have direct access to the contents of our own minds. A further point of agreement is that both positions claim we use this directly given content as a starting point in the process of understanding others. In the case of explicit ST, the simulator first considers what mental states she would be likely to be undergoing if she occupied the other person's situation before inferring that these are the mental states that the other is likely to be undergoing. Finally, she attributes these states to the target.

Gallagher and Zahavi argue that the implications of critical evaluations of the argument from analogy also apply to explicit ST.[613] They make use of a number of criticisms that have been levelled against Mill's position in their critical evaluation of explicit ST. Gallagher and Zahavi are not concerned to produce new refutations of the argument from analogy directly because they take this work to have been done

---

[613] Ibid

already. Instead they endorse a number of objections to the argument that have already been made by other prominent philosophers. For example, Scheler drew out two questionable presuppositions of the argument which Gallagher and Zahavi agree to be problematic.[614] Firstly the starting point for this argument is one's own consciousness. This approach is thought to underestimate the difficulties associated with deciphering self-experience. The second questionable presupposition is that we never have direct access to another's mind. Here the argument from analogy is thought to be guilty of overestimating the level of difficulty involved. Another problem with using an argument from analogy, originally raised by Ryle, and cited by Gallagher and Zahavi, is that this kind of inference could only allow us to place ourselves in the others situation and see how we would feel, not to understand the other.[615] It is lacking in sensitivity to the diversity of other's.

Gallagher argues that simulation is a concept primarily developed at and suited to the personal level. Implicit versions of ST take their lead from explicit predecessors. If they are structurally similar to their explicit counterparts, implicit formulations of ST will also be vulnerable to the same criticisms. Gallagher suggests that the motivation for interpreting certain neurobiological processes as simulations is that simulation theorists are taking a model of a process developed at the explicit level and trying to impose it on the sub-personal level.[616] Gallagher claims that implicit ST is modelled on its explicit counterpart. The explicit version involves a multistage process. The first stage is to create pretend states in oneself modeled on the target.

---

[614] Ibid P.182

[615] Ibid P.177 Wittgenstein will also make this point.

[616] Gallagher 2007 P.359

270

The second stage is to feed this information into a psychological mechanism to generate new output states. The third and final stage is to project this output onto the target. In implicit simulation theory, Gallagher tells us, we also find a multi stage process: first the simulator explicitly perceives the target's behaviour. Next, shared motor representations are activated. Finally the representations of mental/emotional states that are associated with the activated motor representations (i.e. the mental/emotional states that one would experience oneself when one would execute the actions) are attributed to the target.[617] Modelling implicit ST on explicit ST involves postulating a stage-by-stage process, where the final distinct stage, in which the mental state is attributed to a subject, is a stage that goes beyond simple recognition of a mental state and involves a subsequent inferential process.

According to Gallagher, the reason this inferential method is thought to be necessary is because of the way the situation is framed by (explicit and implicit) ST; the simulator observing another person confronts the problem of other minds.[618] This is the problem of how we can access the target's unobservable mental state. In this picture mental states are not thought of as accessible to perception. They are hidden behind the behaviour that we actually observe. Rather than perceiving, for example, an emotion we are speculating about what emotion the subject is likely to be experiencing. According to Gallagher, simulation is a process designed to explain how we overcome this problem. He argues that, for ST, the task is defined as that of how we get from the behaviour we perceive to the mental states supposedly hidden

---

[617] Ibid P.357
[618] Gallagher 2008 P.535

behind this behaviour and inaccessible to perception. Gallagher points out one of the initial conditions which influence this way of conceptualising the situation: both Mill and the explicit simulation theorist start by examining detached third-personal observations of another person's visible behaviour (the observer position). They then seek to explain how we can get from this type of observation to an understanding of what motivates this behaviour.

ST and perception

Gallagher makes similar points to those raised in the discussion of ST. He claims that ST, rather than envisioning the subject to be in direct perceptual contact with other people's mental states, posits an additional cognitive routine (or argument) to enable the subject to arrive at the assumption of other minds.[619] This routine is an inferential process taking us from what is observable to what is not. It is the capacity to mind read. For ST this entails a discrete process of attributing the mental state that one simulates either to a target, or perhaps to oneself, and occurs subsequent to the perception itself.

Gallagher notes that ST is not committed to a particular theory of perception.[620] However, he argues that in practice ST usually implies our perception of other's is very impoverished. He claims that implicit ST typically begins by viewing other subjects from a perspective which is disengaged from social interaction. This is why

---

[619] Gallagher 2011 P.56
[620] Gallagher 2008 P.536

ST needs to rely on inferences from perception to mental states. It is worth elaborating on what sort of information, according to Gallagher, will be unavailable in the simulationist's account of perception. Perception is here understood as the passive absorption of sensory information rather than as an interaction with one's environment. It is not therefore sensitive to cues that can be deduced from the contextual setting of one's interactions. Such cues might be taken from one's own previous experiences of negotiating with either the agent herself or with the more general conventional practices which may be embodied by the current inter-subjective situation.[621] When it is understood as a very basic process only dealing with rudimentary sensory information, perception is not a process equipped to take account of any cues that might clarify what we see. Perception will not be sophisticated enough to take into account the rich context that often frames our intersubjective encounters. This means that the sensory data has to be subsequently interpreted. The meaning of perceptual experience itself does not provide sufficient information to make a judgement about the other person's mental state.

Direct Perception

In opposition to the picture of the mind as inner and perceptually unavailable informing TOM approaches, Gallagher advocates an account of direct perception of the psychological states of others. He presents this as an alternative to TOM accounts of how knowledge of other minds is possible. Instead of having to mind

---

[621] Ibid P.540

read intentional states which themselves are unavailable to perception, Gallagher suggests we come into direct perceptual contact with these intentional states.

Gallagher claims we can directly perceive the intentional states, including the intentions feelings and emotions, of others. The term 'direct' is used by Gallagher in contrast to perception plus some extra perceptual process (such as simulation or theorizing) rather than in contrast to 'indirect' perception. The resources available to perception on their own are considered sufficient to provide an immediate understanding of the intentions and feelings of others. For example these resources allow us to see that the other person is angry and motivated to storm out. We see bodily movement as expressive of emotion, and as goal-directed intentional behaviour. Perception, then, is sensitive to intentions and feelings. Gallagher calls this smart perception; perception is smart enough to be sensitive to, and to utilise, information on its own and so is not reliant on inference mechanisms. In the case of intersubjective perception this means we are able to discern features of the others feelings and intentions. In conceiving of our intersubjective perception as discriminating he is suggesting that it is sensitive to, and able to exploit, information already discernible in the other's behaviour, not that it penetrates through this observation to something hidden behind the behaviour. Indeed; smart perception may be a more appropriate term than direct perception for describing Gallagher's position. He stresses that direct perception does not imply "pure perception" (by which he presumably means something like perception of items constituted

independently of human engagement with them).[622] He also recognises that behaviour is often ambiguous.[623] The main question Gallagher takes himself to be addressing is how smart or informed perception is about its subject matter.

A related question is whether the psychological attributions we make to others are generally correct. Gallagher's position is that in most cases the resources provided by perception will be sufficient for understanding others. This is not to claim that our psychological attributions will always be accurate. He acknowledges that in rare cases we may have to supplement perception using other strategies which might include theorizing or simulation.[624]

The resources intersubjective perception can access are not just perceptions of the bodies of others but fall under the categories of context and interaction (which will be explored in more detail shortly). Physical, social and cultural aspects of the environment contextualise the meaningful gestures and expressions and can provide us with clues about what another is thinking and doing. For example if someone is marching towards the exit buttoning up their jacket and their shoulders are tense we may directly perceive that they are angry and intend to storm out. Most of what Gallagher has to say about direct perception concerns perception of objects rather than intentional states. He uses the example of perceiving a red car. A non-smart perception would perceive a red mass of a certain shape. Smart perception immediately sees the car.[625] Shape and colour immediately appear as aspects of the

---

[622] Ibid P.538
[623] Ibid P.540
[624] Ibid
[625] Ibid P.536

car rather than unsynthesised pieces of sensory data. Gallagher claims one

characteristic of smart perception is that it is enactive. Rather than consisting in the

simple processing of sensory input it involves sensory-motor skills.[626] The car is not

perceived merely as one object among others but as something one can use. The car

'affords' certain types of action.[627] These affordances inform and partially constitute

the perceptual process. Gallagher claims one does not first see the car then judge it

as drivable. Ordinarily this judgment is built into the direct perception. Perception

also often involves an 'emotional coloration'.[628] Gallagher gives the example of a

love hate relationship with his car tacitly informing his perception of it. Gallagher

stresses he is not denying that complex neurological processing may underlie direct

perception. However he argues that to understand how the various aspects of sensory

data are integrated is a problem for the neuroscientist rather than for the perceiver.[629]

The observer position

We can clarify the direct account of perception that Gallagher advocates by

contrasting it with what he calls the observer position.[630] Gallagher describes this

position as one that typically begins by viewing other subjects from a perspective

which is disengaged from social interaction. Gallagher argues that, here, perception

means third-person observation rather than something that happens in the context of

interaction. The observer's role in the proceedings is not supposed to contribute

---

[626] Ibid P.541
[627] Ibid P.537
[628] Ibid P.538
[629] Ibid P.537
[630] Ibid P.540

276

anything to the process of understanding the other because she is conceptualised as external to the events she perceives. The observer is peripheral to the action rather than involved in it. Consequently there is a sharp distinction between the perception and one's own action. Gallagher further argues that although this type of perception is able to differentiate between other people and mere objects it is not able to do much more. It is not, for example, able to gather anything about the emotions or intentions of the agent.

Response of TOM Positions

Gallagher's characterisation of TOM as employing a very uninformed or non-smart perception is perhaps overstating the case. The fact his discussion is spelt out in terms of objects rather than psychological states makes it difficult to apply to TOM approaches but the implication he draws for TOM approaches appears to be that data concerning others is perceived under the most rudimentary characterisation and more sophisticated (psychological) characterisations have to be inferred on this basis. If this is the right way to interpret Gallagher then just as non-smart object-perception begins with a red mass of a certain shape and infers a car, non-smart agent perception begins with rudimentary data such as a curved mouth and infers a psychological characterisation such as a smile. However neither TT nor ST are committed to denying that we are perceptually sensitive to psychologically characterised data per se. In particular Gallagher's characterisation of these theories as in the observer position fails to take account of the nuances of sub-personal

versions of ST. Sub-personal versions of ST stress the role of contextual information in allowing us to decode another's behaviour. These positions want to understand perception as sophisticated enough to be sensitive to contextual features of a situation facilitated through attuned intersubjective encounters. They are not committed to claiming we observe brute sensory data. However there is still a problem with these positions. Psychologically characterised data is conceived as the input to an inference process. While ST is focused on psychologically characterised data this data is still understood as outer manifestations of hidden inner causes. If this was not the case then performing a simulation process would be redundant. (Gallagher's treatment of perception of psychological states as analogous with the perception of objects is also problematic in itself. It ignores the distinctive character of the way these states are revealed to one. This point will be elaborated on shortly).

TT would also be expected to claim it can accommodate smart perception. It is worth considering a possible objection TT could make to Gallagher as it will enable us to clarify the differences between TT and direct perception. It is open to TT to argue that all Gallagher has done is draw attention to the theory laden character of perception. This does not stop it utilising a theory. Smart perception is theory laden so TT is not undermined by these claims.

Gallagher could respond that he is not claiming perception is theory laden. We do not have to engage in inferences. One simply describes how things appear to one. This will not help though; as the point of TT's claims about the theory laden nature

278

of knowledge is that the theoretical content is non-explicit. Furthermore, this would not be quite right as perceptions come with sets of expectations about the behaviours they afford. A car is something one might drive. Gallagher also has another line of argument open to him. For Gallagher meaning is not set by the laws of a theory. As we learn more about an agent and their situation we may make adjustments to our psychological attributions. The point here is not simply that Gallagher's position has more room for manoeuvre concerning the correctness of psychological attributions; we saw that TT also implies that these are potentially revisable. However in the case of TT the revisions concern the nature of *unobservable states*. In Gallagher's case the revisions will concern *further observation*.

Gallagher is not claiming that meaning is given immediately and incorrigibly. His position makes understanding an extended process. It is an open ended process of interpretation within a contextual setting. Therefore the claims one makes go beyond the present moment or present data because they imply a pattern of data over time. This pattern will be open ended and it is possible to make mistakes in its identification. TT is also able to claim theory ladenness allows this possibility. However the way theoretical classifications transcend the present moment is different. For Gallagher this is not an inference to something which is unobservable but to a pattern of interpretation extended over time.

<u>Gallagher on Mirror Neurons</u>

Gallagher also takes issue with ST's interpretation of empirical data. He is not disputing the empirical evidence for the existence of the neural mechanisms that are being claimed to underlie implicit ST. He only disagrees with interpreting the activity of these mechanisms as an instance of simulation. He argues that a better interpretation of the role MNs have in interpersonal encounters is that they facilitate a direct perception of other agent's intentional states.[631] He thinks they therefore provide empirical support for his own position. Gallagher claims MNs may constitute empirical evidence for direct perception.[632] He argues MNs "constitute the neural correlates of a non-articulated immediate perception of the other person's intentional actions".[633] He claims this interpretation of MN activation fits the direct perception account of intersubjective understanding and interaction.[634] This suggests a relationship between sub-personal level and personal levels of ascription. This relationship would best be understood as an answer to question three. Sub-personal processing might be appealed to in an explanation of what makes personal level processes possible.

---

[631] Gallagher 2007 P.358
[632] Ibid P.541
[633] Ibid
[634] Ibid P.542

<u>Interaction theory</u>

There is also a second strand to Gallagher's position. The strand we have been

looking at involves treating perception of others as analogous with perception of

objects whereas this second strand in his thought deals specifically with agents.

Gallagher advocates a position he calls interaction theory (IT) which he again

presents as an alternative to TOM views.[635] It is based on developmental psychology

as well as phenomenology. It emphasises the strong role interaction plays in

interpersonal understanding. Gallagher suggests IT as a way of avoiding a number of

problems associated with TOM approaches. These problems arise from assumptions

shared by TT and ST. The assumptions are, firstly, that there is a problem of social

cognition due to lack of access to other peoples intentional states.[636] Secondly, mind

reading is employed to explain or predict behaviours we observe from a third-person

stance. Thirdly that mind reading is the primary and pervasive strategy for

understanding others.[637] One problem facing TOM is that there is no

phenomenological support for mind reading, at least for claims of pervasive

conscious mind reading. A second problem is referred to as the starter problem.[638]

To know what theoretical rule or simulation one should apply in a given situation

presupposes some initial knowledge about the other person. Gallagher appeals to

these problems to justify the introduction of his alternative theory.[639] IT stresses that

other minds are not hidden or inaccessible. Instead intentional states are revealed in

---

[635] Gallagher 2011 P.55
[636] Ibid P.56
[637] Ibid P.57
[638] Ibid
[639] Ibid P.58

bodily behaviour.[640] Mind reading rather than being the default strategy for understanding others is only employed in rare cases.

The everyday stance one adopts towards others is second person interaction rather than detached third-person observation. Perception can also utilize features inherent in one's own interaction with another agent. One's own involvement in the proceedings can also serve to disambiguate the other agent's expressive movements. For example one may have just insulted someone in the course of an argument which motivates their storming out of the room. Social understanding, Gallagher claims, often boils down to social interaction.[641] Social perception is the product of an already meaningful embodied engagement with others, rather than the passive reception of sensory information. Understanding of the other person is constituted within perception–action loops.[642] These define the activities one performs with, or in response to, others. Furthermore Gallagher argues perception is either innately or at least very early on tuned to socially relevant features of one's environment.[643]

The developmental evidence Gallagher appeals to in support of IT concerns our primary and secondary intersubjectivity. Primary intersubjectivity begins at birth.[644] It is constituted by innate or rapidly developing sensory-motor capacities that pull the newborn into relationships with others and allow it to interact.[645] Gallagher

---

[640] Ibid
[641] Gallagher 2008 P.540
[642] Ibid
[643] Ibid
[644] Gallagher 2011 P.59
[645] Ibid P.60

282

claims these capacities manifest themselves at the experiential level. We see

another's intentional states in their movements and expressions.[646] Very early on

infants manifest an 'interactive attunement' with others in the form of timing and

coordination.[647] Furthermore, Gallagher argues, primary subjectivity does not recede

to make way for more sophisticated mind reading techniques. It continues into adult

life (we will see Merleau-Ponty makes a similar suggestion).[648] These embodied

interactive processes continue to shape adult intersubjectivity. Gallagher explains:

"we continue to understand others in strong interactional terms, facilitated by our

recognition of facial expressions, gestures, postures, and actions as meaningful".[649]

Secondary intersubjectivity begins at around 1 year of age. By this stage infants are

in possession of "a non-mentalizing, perceptually-based, embodied and pragmatic

understanding of the intentions and dispositions of other persons".[650] They start to

co-constitute the meaning of the world in social interactions. They also gain a more

sophisticated understanding of others by being able to contextualize their actions in

terms of both pragmatic tasks and cultural practices.[651] Gallagher draws out a

consequence of this approach, he claims "Meaning and emotional significance is co-

constituted in the interaction – not in the private confines of one or the other's

head".[652]

---

[646] Ibid
[647] Ibid
[648] Ibid P.61
[649] Ibid
[650] Ibid P.62
[651] Ibid
[652] Ibid

Our movements and gestures are often synchronised with other agents and resonate to them.[653] Gallagher argues we do not choose to enter into interaction with others. Rather, he claims, it is something we find ourselves thrown into before we could be capable of making decisions about it.[654] Interaction is a fact of our embodiment. He appeals to the work of Merleau-Ponty in support of the bodily nature of interaction. Gallagher claims we are involved in interaction before we are capable of knowing it.[655] It begins even before the intersubjective processes belonging to primary intersubjectivity.[656] In fact these processes begin before birth. The developing foetus exhibits non-reflex based movement in response to the stimuli of the mother's movement. Gallagher calls this a form of intercorporeal interaction.[657] Thus the foetus finds itself immersed in interactive processes that precede those that will characterise one's primary subjectivity.[658]

This notion of strong interaction fuels a response to the problem of other minds It facilitates Gallagher's view of an agent as "someone who emerges from intercorporeal interactions, and develops in social interactions with others" as opposed to an enclosed individual.[659] The focus on intersubjective interaction facilitates a departure from conceptions of self-agency that characterise TOM approaches. TOM approaches treat self-agency as "reducible to neural or mental or

---

[653] Ibid
[654] Ibid P.63
[655] Ibid P.66
[656] Ibid P.65
[657] Ibid PP.64-65
[658] Ibid P.65
[659] Ibid P.67

strictly individual processes framed in terms of mental causation".[660] Such

approaches therefore encourage the use of analogical argument to overcome a

problem of other minds. Gallagher offers a different response to the problem. The

capacities manifest in primary and secondary interactions should not be understood

as capacities belonging to an individual. Arguments from analogy rely on an

individual possessing intentional concepts based entirely on their subjective

experiences. The understanding involved in primary and secondary intersubjective

interactions is not reducible to individual capacities. Meaning is a product of, and

created through the interactions themselves.[661]

It is not totally clear whether IT is continuous with what Gallagher has to say about

direct perception. In one way the discussion of smart perception pre-empts what

Gallagher has to say about IT. It suggests we grasp things in terms of the

possibilities they afford for interacting with them (a car is driveable). However the

possibilities for interaction that other agents afford seem to be of a different nature.

The practices we can enter into with other agents are not the same as those that

objects afford us. This topic will be examined in detail in the next chapter.

Is Gallagher engaged in Epistemology?

We will now return to the question of how to interpret Gallagher's position in regard

to the epistemological issue of knowing other minds.

---

[660] Ibid P.68
[661] Ibid P.65

An epistemological answer

In support of the first construal, in which direct perception is an answer to the epistemological issue, certain comments by Gallagher suggest he takes seriously the view that there is a task of explanation and prediction required for social interaction. On this reading social perception has a specific function; to ensure that we possess the resources to avoid being hindered by potential epistemological difficulties concerning other people's intentional states. Understanding another person's mental states allows us to explain or predict their behaviour. His dispute with TOM only concerns the way this understanding is achieved. For example he suggests that TT and ST require extra-perceptual cognitive elements "because of the way the problem is framed".[662] This suggests that it is only the framing of the problem that Gallagher is taking issue with; not the existence of a problem. Similarly he suggests, "perception is smart enough on its own, without the supplement of inference mechanisms" to provide a sense of the others intentions.[663] Smart perception is being offered as an answer to the question whether perception is able to justifiably deliver beliefs about the psychological states of others without supplementation.[664] On this picture justification for our perceptual beliefs consists in highlighting fundamental capacities to assess things accurately rather than consisting in an inferential process in which these beliefs are inferred on the basis of other beliefs. In this way it is continuous with the way other perceptual beliefs are justified. Gallagher equates the question of whether we have a direct perception of intentional states with the

---

[662] Gallagher 2008 P.536
[663] Ibid
[664] Ibid

286

question whether we have smart perception, or how 'informed' perception is.[665] To be informed here means to be informed about the other in relation to solving a potential problem about their intentional states. Although the question is not being answered independently of the perceptual process it is still being tacitly addressed. The answers are understood as built in to the perception in the same way the drivability of one's car is built into one's direct perception of it.[666]

This construal can be supported by the use Gallagher makes of Merleau-Ponty's account of bodily expression (which we will examine later in this chapter). However this is not the use Merleau-Ponty's puts it to. Merleau-Ponty's work is better understood as rejecting the other minds problem. We will see that in his writing the epistemological question of how we justifiably claim knowledge of other minds is emphatically rejected.

An epistemological rejection

An alternative reading of Gallagher does not acknowledge a special epistemological issue about other minds. Gallagher stresses that *all* perception is direct.[667] There is nothing epistemologically unique about perceptions of intentional states. Direct perception applies to cars as much as to intentional states. This more radical construal of Gallagher's approach is also suggested by his rejection of mentalizing or mind reading approaches to understanding others. He presents his approach as "in

---

[665] Ibid
[666] Ibid P.537
[667] Ibid

287

contrast to the great effort mounted by TT and ST to solve the problem of other minds".[668] Gallagher is critical of TOM approaches because these approaches 'define the problem to be solved' as how we get from perceived behaviour to hidden mental states.[669] He argues "that question is itself put into question by the direct perception approach".[670] Gallagher argues direct perception suggests this is not what happens in normal everyday encounters. This makes it look like Gallagher is disputing the epistemological set up employed in TOM. However the reason Gallagher thinks no move from perceived behaviour to hidden intentional states is required is because people "do not ordinarily need to go further than what is already the rich and complex comprehension that we gain through the perception of a situated agent".[671] It is not that the question of what is going on in the others mind does not arise, only that the resources required for answering this question are already perceptually available.

The emphasis on interaction in Gallagher also suggests a reading in which there is no epistemological problem of other minds. This is because he acknowledges that before we are in a position to explain or predict the actions of other people we are already interacting with the expressions, gestures and purposive movements we observe others making. These already reflect their intentions and emotions.[672] This picture suggests we are interacting with others rather than forming judgements about their psychological states. Gallagher appears to fluctuate between two stances. On

---

[668] Ibid P.535
[669] Ibid P.540
[670] Ibid
[671] Ibid
[672] Ibid P.542

one stance intersubjectivity is constitutive of psychological states. It is part of the make up of a psychological state that it is publicly expressed through one's transactions with other agents. On the other intersubjectivity is a tool that allows these states to be accessible to perception. This ambiguity is prominent in discussion of IT. It is unclear whether a grasp of what it is to have psychological states requires one to be engaged in 2nd person relations with others (for example one grasps anger when one is able to respond appropriately to an angry person) or whether such interactions merely function as aids which helps smarten our perceptions rather than being constitutive of the states themselves.

One way to summarize Gallagher's position might be to say the problem of minds does not arise because its creation and resolution are instantaneous. This is not the same as saying it is a pseudo-problem that does not arise. Although intentional states are not hidden, our perceptions of these states are involved in solving an epistemological problem. On this view, although problem solving is involved, there is no *special* problem of other minds. Attribution is justified because everything is perceptually available; it is simply a matter of learning to recognise patterns. This is a particular application of smart perception. Gallagher's arguments are built on his account of bodily expression. This account is influenced by Merleau-Ponty's. However, if we examine other aspects of Merleau-Ponty's philosophy alongside his account of bodily expression it will become more problematic to interpret him, as Gallagher does on my first construal, as offering an answer to the other minds problem. In making these arguments Gallagher draws on ideas that potentially lead

one stance intersubjectivity is constitutive of psychological states. It is part of the make up of a psychological state that it is publicly expressed through one's transactions with other agents. On the other intersubjectivity is a tool that allows these states to be accessible to perception. This ambiguity is prominent in discussion of IT. It is unclear whether a grasp of what it is to have psychological states requires one to be engaged in 2nd person relations with others (for example one grasps anger when one is able to respond appropriately to an angry person) or whether such interactions merely function as aids which helps smarten our perceptions rather than being constitutive of the states themselves.

One way to summarize Gallagher's position might be to say the problem of minds does not arise because its creation and resolution are instantaneous. This is not the same as saying it is a pseudo-problem that does not arise. Although intentional states are not hidden, our perceptions of these states are involved in solving an epistemological problem. On this view, although problem solving is involved, there is no *special* problem of other minds. Attribution is justified because everything is perceptually available; it is simply a matter of learning to recognise patterns. This is a particular application of smart perception. Gallagher's arguments are built on his account of bodily expression. This account is influenced by Merleau-Ponty's. However, if we examine other aspects of Merleau-Ponty's philosophy alongside his account of bodily expression it will become more problematic to interpret him, as Gallagher does on my first construal, as offering an answer to the other minds problem. In making these arguments Gallagher draws on ideas that potentially lead

to more radical conclusions. One of these conclusions is that our relation to others is not primarily an epistemological one. A problem is whether our relations to others are appropriately construed as epistemic relations in the first place. Though Gallagher emphatically rejects mind reading he can still be read as offering an account of how we construct epistemic relations.

Seeing this as an epistemic relation is further undermined if we take seriously the claim that our understanding of mindedness is necessarily intersubjective. Mindedness isn't something we could grasp independently of grasping the kind of intersubjective encounters that we find ourselves in, in which case it is difficult to see how the nature of the epistemological problem could be articulated in the first place.

Part 2

Gallagher's account is heavily influenced by Merleau-Ponty. However it still shares some of the trappings of the TOM debate. Merleau-Ponty's offers the resources to avoid these pitfalls. My discussion of Merleau-Ponty will be divided into sections. The first will examine Merleau-Ponty's criticism of the classical model. This will enable us to see faults with TOM and Gallagher. The second will outline the case Merleau-Ponty makes for the direct perception of psychological expressions as psychological, and the lack of access to physiological data including his discussion of expression. The third will examine his claims about the intertwining of self and others. One use of Merleau-Ponty's work will be to offer both an expansion and a correction of Gallagher's work. His work will also be useful in another way. Many of the themes discussed in relation to Merleau-Ponty's work have already been touched upon in discussions of sub-personal versions of ST. However the claims were shown to be in tension with the philosophical picture underlying ST and dictating its answer to question 1. Here Merleau-Ponty's position will be shown to be able to accommodate these themes because there is no conflict between his claims and the philosophical picture that he is advocating.

The classical model

Whereas it is unclear whether Gallagher's work offers a rejection of the philosophical model in which the intentions of others pose a particular

epistemological problem, or merely offers a solution to the other minds problem within this framework, Merleau-Ponty's work offers a decisive refutation of the philosophical picture behind TT and ST. I will begin by examining Merleau-Ponty's criticisms of this model that appear in *The Phenomenology of Perception*. He is attacking intellectualised accounts of how we understand others. Merleau-Ponty opposes the idea psychological phenomena could be inferred from physiological data (as suggested by accounts driven by sub-personal processing). In fact he argued that we do not have access to such data. He is resisting a certain approach to explanation typical of science. This is an understanding of explanation which involves attempting to look beneath the phenomena to try to uncover what supports it.[673] One reason for opposing an objective approach is that Merleau-Ponty is resisting a certain structuring of our knowledge (of others) in terms of building blocks formed from elementary particles of sense data. This strategy limits our ability to interpret gestures to bringing out a particular relationship between people and the world as presented to one through natural perception.[674]

Classical approaches are influenced by a Cartesian framework involving reified notions of body and soul. The body becomes the sum of its mechanical components while the soul is something 'wholly given to oneself' or transparent to consciousness.[675] As the result of a reflective procedure subject and object become theoretically detached from one another.[676] It is only the body that is the subject of

---

[673] Merleau-Ponty 1962 P.113
[674] Ibid P.186
[675] Ibid P.198
[676] Ibid

empirical/scientific investigation. This body is the objective body as described by physiology.[677] It is not something amenable to being inhabited by consciousness.[678] Merleau-Ponty argues a lesson of Cartesian approaches is that it is not comprehensible how significance and intentionality could reside in molecular or cellular structures.[679] On this picture there is a straightforward dichotomy between two modes of being: 'being for itself' deals with the realm of objects arrayed in space and 'being in itself' deals with the realm of consciousness.[680]

Merleau-Ponty is also opposing the traditional subject-object dichotomy.[681] This can be related to the philosophical picture facilitating the problem of other minds. This dichotomy construes observable behaviour, including linguistic behaviour, as 'signs' of unobserved thought in the way smoke signals fire.[682] Scientific objective approaches entrench a theoretical division between an epistemological subject and an object. It carries a conception of the natural world as existing 'in itself' autonomously of the existence it has for experiencing subjects.[683] An implication of dichotomous approaches is that one must employ analogical arguments to understand others.[684] One must search within oneself to decipher the meaning of the gestures one witnesses others making.[685] Other people; conceived as empirical beings, become "mere pieces of mechanism" while the psychological subject is not

---

[677] Ibid P.351
[678] Ibid
[679] Ibid
[680] Ibid P.349
[681] Ibid P.174
[682] Ibid P.181
[683] Ibid P.154
[684] Ibid P.352
[685] Ibid P.184

something we can encounter.[686] There is no place for a plurality of consciousness and other people in objective thought. Consciousness is understood as what exists for the individual. This does not allow for the possibility of encountering another consciousness in experience.[687]

Merleau-Ponty, however, thought it was a confusion to posit the involvement of a cognitive operation whenever we encounter others.[688] Understanding others does not consist in an intellectual interpretation of their behaviour.[689]

The inner psyche

I will now consider some complementary remarks Merleau-Ponty makes in 'The Childs Relations to Others'.[690] Merleau-Ponty suggests that the way psychology traditionally construes questions about the nature of our relations with others and how they come about leads to difficulties.[691] One interesting aspect of this discussion concerns the operative notion of the psyche. Merleau-Ponty's ideas are radically opposed to traditional conceptions of the psyche or consciousness. He opposes a conception of the psyche as "what is given only to one person".[692] My mind is only accessible to me. Other people only have an indirect access mediated by my bodily appearances. It is not difficult to see the negative implications this

---

[686] Ibid P.349
[687] Ibid
[688] Ibid P.185
[689] Ibid
[690] Merleau-Ponty 1964
[691] Ibid P.113
[692] Ibid P.114

294

carries for knowing other minds; how are we supposed to reliably know the psyche of another when by definition it is what is given only to that person. The psyche is something incommunicable, accessible only through introspection, and the psyche of another is radically inaccessible. The best one can do on this picture is to guess what another is thinking on the basis of the others bodily appearances including facial expressions, gestures and speech. Merleau-Ponty argues that traditional psychology does not provide us with the tools we need to account for our understanding of others.[693]

The conception of the body implicit in the above picture poses a second difficulty. Merleau-Ponty asks what would justify one's assumption that the others body encloses a psyche.[694] We have to consider what it is that is exclusively available to each individual on this picture. This is the mass of sensations informing the subject about the states and functions of his bodily organs.[695] One only knows one's body through the mass of sensations it provides one with. This cenesthesic sense, Merleau-Ponty points out, "is as individual as the psyche itself".[696] It is not something visible to an observer. So while we talked a moment ago of mediating bodily appearances, in fact we cannot talk about a shared medium of representation here. My representation of my bodily sensations and my representations of your bodily sensations are not given to me in a common language. I have an introceptive image of my body and a visual image of yours. Those mediating bodily signifiers

[693] Ibid P.113
[694] Ibid P.114
[695] Ibid
[696] Ibid

cannot strictly speaking be equated with the sensations experienced. Classical psychology is said by Merleau-Ponty to get round this by taking the bodily signifiers en masse to provide a kind of decoding. One observes the others bodies characteristic gestures and utterances and projects what one feels in one's own body.[697]

Four components are involved in this problem of our awareness of the experiences of others. These are one's own psyche, the introceptive or sensory image of one's own body, the others visual body and the hypothetical psyche of the other person. Merleau-Ponty notes a number of difficulties for this picture. Firstly, there is a problem how one comes to associate one's own intimate bodily experiences with one's visual experiences of the other. This is an implausibly intellectualised account of our understanding of other minds. Translating the visual data we get of the others bodily expressers would mean bringing this data under concepts associated with my body's reactions to my own (introceptive) experience. But we are able to perceive expressions before we would plausibly be developed enough to perform this intellectual feat. Very young children are sensitive to facial expressions. It is implausible they are able to perform this complex processing in order to gain an understanding of, for example, a smiles global meaning and learn that a smile is a reliable indicator of a benevolent feeling. The process would involve first comparing the visual perception of another's smile with the facial movement one makes when one feels benevolent. Secondly it would involve projecting onto the other a feeling of benevolence of which one has intimate experience but cannot directly perceive in

[697] Ibid P.115

296

the other. Merleau-Ponty points out that this is hard to reconcile with the relatively

preconscious character of our perception of others.[698]

Furthermore the attribution, for example, of a benevolent emotion to a smiling face,

would have to begin through drawing analogies between the facial gestures one

observes and those one executes.[699] But this presupposes a way of comparing one's

experience of the motor processes one has when smiling with the visual experience

of the smile of the other.[700] Merleau-Ponty asks: "Have we the means of

systematically comparing the body of the other as seen by me with my body as

sensed by me".[701] He notes that for this to work there would have to be a reliable

point by point correspondence between the two occurrences. Merleau-Ponty argues

these difficulties facing the traditional picture can be brought out through

considering the phenomenon of imitation. Imitation involves performing a gesture in

the image of another's gesture. On the above picture; this would require that one

translates the visual image of another's gesture into a motor language. In the case of

smiling one would have to activate one's facial muscles in a way that reproduces the

visual expression. However; one will not have the others internal motor feeling in

one's own face.

To summarise; the classical model involves Cartesian conceptions of mechanistic

physical bodies and immaterial minds. This results in behaviour being understood as

---

[698] Ibid
[699] Ibid
[700] Ibid P.116
[701] Ibid

the signifier of hidden intentions. On this model we understand others by analogy. Our own conscious experience is assigned a privileged role in intentional attribution. We only have indirect access to other psyches mediated by bodily appearances. Furthermore Merleau-Ponty provides reasons why such a strategy will be problematic; we are left without a common medium of representation for psychological states of self and other.

Direct Perception

Merleau-Ponty offers an alternative definition of psyche as "a relation to the world".[702] His replacement of the conception of psyche found in the classical model allows the construction of a less problematic and more direct picture of our relation to the minds of others. It involves the abandonment of the idea that psyche is only accessible to oneself and cannot be observed by others. Rather than thinking of conscious experiences as involving a series of closed off episodes they are modes of the agent's comportment towards the world and its contents. In the same manner; the psyches of other people are primarily ways of orientating themselves to the world. Because consciousness is directed at things external to itself, including the actions of other psyches, it is able to experience the things it meets in a meaningful way. As Merleau-Ponty puts it they are "themes of possible activity for my own body.[703]

---

[702] Ibid P.117
[703] Ibid

Early consciousness first encounters what Merleau-Ponty calls conducts. These are

our own actions and those of others. Conducts are ways of grasping our natural and

cultural surroundings. Actions are not alien to the agent-observer in the way that the

visual body as it appears in traditional conceptions of psyche must be. In meeting

another, one is not confronted with a mass of sensations but a corporeal schema. A

corporeal schema is the overall experience one has of one's body and its position in

surrounding space. This serves to orientate an individual towards the world. Instead

of a mass of inner sensations, what is perceived is the body's position in relation to

important coordinates of its environment. Merleau-Ponty argues that it is in another

person's conduct that one can discover his consciousness.

We saw Merleau-Ponty was sceptical about using a point by point correspondence

between one's own body as sensed and the others body as seen. He argues that a

more likely supposition is that the subject is identified globally.[704] Merleau-Ponty

argues it is far less problematic to understand how *conducts* can be transferred from

another person to oneself than it is to understand how one could represent a radically

foreign psyche. An action 'speaks directly' to one's own unique motility. The author

of the action is not at this stage understood. However a perspective on the author is

created when self and other are both understood as conducts working in the world.[705]

Merleau-Ponty stresses that this involves changing the traditional conception of the

understanding we have of our own body as well of the psyches of others. To

---

[704] Ibid P.116
[705] Ibid P.117

appropriate the conducts of others requires a corporeal schema not just a mass of private sensations.

With the introduction of the postural or corporeal schema we are in a position to see the advantages to this alternative conception of psyche when it comes to grasping other minds. The various sensory components of experience are understood at the level of "an already organised totality".[706] The sensory domains involved in perceiving one's body are not presented to the subject as totally distinct regions. They have a common style of action or gestural meaning which binds them together. At this level the possibility of having common bodily experiences with others looks much more coherent; we do not meet the same unbridgeable gap opened up by the absence of a shared representational vehicle for representing one's own experience and that of others (e.g. translating from motor sensation to visual data). As Merleau-Ponty explains: "The visual image of the other is interpreted by the notion I have myself of my own body and thus appears as the visual envelopment of another corporeal schema".[707]

Indeed Merleau-Ponty suggests one's perception of one's own body will be problematic on the traditional picture. If our access to the body was really composed of individual cenethesic experience the body would be swallowed in this cinesthesia. A corporeal schema on the other hand will be transferable across sensory domains just as it is across agents. We have one system with two terms, 'my behaviour' and

---

[706] Ibid P.118
[707] Ibid

'the others behaviour'. This functions as a whole. One experiences one's own body more fully when it can be submerged in dealings with others. Other people, understood as conducts, offer themselves to one's motor intentions.[708] Merleau-Ponty makes reference to Husserl's idea of an 'intentional transgression' in which one animates and pervades another person. In perceiving another, the bodies of perceiver and target are coupled. One lives the others conduct 'at a distance'.[709] One also knows that one's own gestures can be taken up by another's intentionality. This transfer of intentions across bodies, along with the mutual alienation of self from other (which we will shortly explore), make perception of others possible.

Phenomenology and subjectivity

Merleau-Ponty is drawing on a philosophical tradition in which our experience of another person is an experience of another subjectivity and is quite different to experiencing a pattern in an object. This tradition starts with Heidegger who has a notion of everyday being-with. The everyday way in which we experience our environment includes traces of this. He gives an example of walking along the perimeter of a field. Here the field is experienced as belonging to a particular person and well maintained by them.[710] Another example concerns the workspace of a craftsman. Heidegger thought that upon entering this space the craftsman would encounter not just equipment but also a sense of the people for whom the work was intended. For example if one is engaged in making clothes these carry with them an

---

[708] Ibid
[709] Ibid
[710] Heidegger 1962 P.153

301

"essential assignment or reference to possible wearers".[711] Heidegger claims the

others attached to such products are not to be understood as add-ons to some objects

which are merely available to oneself in one's experience. They are encountered

within an experience of the world which includes an appreciation that these objects

are available to others. He also claims this world is already one's own too.[712]

Heidegger claims that when we encounter the presence of others through such

aspects of our environment they are neither ready to hand nor present at hand but as

Daseins like oneself. Dasein is Heidegger's term for human being which has a

distinctive type of being because this being is conscious of its own existence.

Daseins are in the world and alongside one.[713] Heidegger disputed a picture in which

as he puts it, one must:

"start by marking out and isolating the 'I' so that one must then seek some way of

getting over to the Others from this isolated subject".[714]

Others are not defined against oneself. Rather they are "those from whom, for the

most part, one does *not* distinguish oneself—those among whom one is too".[715]

The world Dasein inhabits is always a shared world and inhabiting the world is itself

being-with other subjectivities.[716] To put this point another way "others are

---

[711] ibid
[712] Ibid P.154
[713] Ibid
[714] Ibid
[715] Ibid

encountered environmentally".[717] One often encounters the Dasein of others in terms

of things that are ready to hand within the world. However Heidegger stresses that

people are never experienced as present-at-hand person-things.[718] We meet them "at

work" in the world. Even if they are just standing around they are not apprehended

in this manner. Standing around is itself grasped as a mode of being in the world.

The reason Daseins encounter one another in this way is because Dasein is

essentially Being-with.[719] This is an existential characteristic of Dasein which holds

true even when Dasein is alone. Heidegger claims even Being-alone is a mode of

being-with.[720]

It is because each Dasein has this essential structure of being-with that it can be

encountered by other Daseins.[721] Heidegger claims :

"Dasein, as Being-in-the-world, already is with Others".[722]

For Heidegger other Daseins are objects of solicitude. This denotes a kind of

concern with others which can manifest itself in terms of being for, against or even

indifferent to others.[723] Even if a Dasein takes itself to be entirely independent of

others this is still a way of being-with them. Others are still disclosed to one as other

---

[716] Ibid P.155
[717] Ibid
[718] Ibid P.156
[719] Ibid
[720] Ibid
[721] Ibid P.157
[722] Ibid P.162
[723] Ibid P.158

303

Daseins.[724] The world such a Dasein inhabits is still a world for others even if no others are currently present. Because our fundamental mode of being is being-with, this presupposes an understanding of others.[725] This understanding is not a form of knowledge derived from acquaintance; it is a primordial existential mode of being which makes such acquaintance possible. Even acquaintance with oneself is grounded in being-with.[726] For Heidegger being towards others is not the same as being towards objects because others share the same type of being as us.[727] We are orientated towards others in a relationship of being-with that is irreducible and autonomous.[728]

Sartre is another influential thinker in this philosophical tradition. He thinks it is an experience of one's own objectivity that reveals the presence of other subjectivities. He claims that other people, such as a man passing by in the street, are revealed to one as objects.[729] However he resists the view that this is the primary relation by which the presence of other subjectivities is discovered.[730] This would be problematic because it does not, for example do anything to rule out the possibility that the man is merely an automaton. Sartre argues that this apprehension of the other as an object essentially refers one to a more fundamental apprehension of the other as a "presence in person" rather than a mere object.[731] A perception of the other refers to "a primary relation between one's consciousness and the others" in

---

[724] Ibid P.160
[725] Ibid P.161
[726] Ibid
[727] Ibid P.162
[728] Ibid
[729] Sartre 1943 P.252
[730] Ibid P.253
[731] Ibid

which the other is given directly to one as a subject.[732] Sartre denies he is appealing

to any kind of mystic experience. The other is shown to us "in the reality of

everyday life".[733]

Sartre illustrates his position with the example of standing in a park which has

benches positioned around the edge of a lawn. A man passes by these benches.

Sartre argues the passerby is simultaneously apprehended as both an object and as a

man.[734] Sartre suggests there are important differences involved in apprehending a

passing thing as an object and as a subjectivity. If the thing were just an object we

could classify it using conceptual apparatus suited to spatio-temporal objects. His

relation to the other objects round the park would be of a purely additive type. He

would be *beside* the benches; at a given distance to the lawn and so forth. A

consequence of this would be that were the man-object to be removed from the

scene, the relations holding between oneself and the various other objects would not

be disrupted.[735] The arrival of this man-object would not alter one's current

relationships with the other things in one's surroundings. However Sartre's point is

that the case is not like this. The passerby does not appear to one as just another

additive relation like the juxtaposition between the bench and the lawn. According to

Sartre we "register an organization *without distance* of the things in my universe

around that privileged object".[736] The passerby's presence has an impact on one's

relation to the other things in one's environment. Sartre claims the man stands in a

---

[732] Ibid
[733] Ibid
[734] Ibid P.254
[735] Ibid
[736] Ibid

305

spatio-temporal relation to the lawn but the lawn is also "bound to him in a relation which at once both transcends distance and contains it".[737] The man brings to Sartre's environment a spatiality which is not Sartre's spatiality:

"for instead of a grouping toward me of the objects there is now an orientation which flees from me".[738]

Whereas a moment ago the bench was apprehended as a thing Sartre could go and sit on, or walk past, or turn away from, it is now apprehended in relation to the intentions of the passer by who may move past it, or stop and sit in it etc. The various new relationships are given to Sartre as a whole and at the same time they escape him.[739] The man's presence causes the relations Sartre has established between the objects of his universe to disintegrate. As he puts it "an object has appeared which has stolen my world from me".[740] The others appearance prompts a decentralisation of the world which comes in to conflict with the centralisation that Sartre believes one is constantly striving to effect.

Although the other is an object it is unique among objects in that it sees what one sees.[741] There is also, for Sartre, the permanent possibility of being seen by the other. Sartre claims that it is this possibility of becoming an object of observation for the

---

[737] Ibid
[738] Ibid
[739] Ibid P.255
[740] Ibid
[741] Ibid P.256

306

other that allows one to apprehend the presence of his being-as-subject. Just as one is able to apprehend the other as object the other is also able to objectify oneself.[742]

Sartre argues the relation of being seen by the other represents an irreducible fact which could not have been deduced either from the essence of the other in so far as they are an object, or from one's own subjective being.[743] In fact the possibility of seeing the other as an object is derivative of the apprehension of this relation and represents a degradation of the original relation. He claims "being-seen-by-the-other is the truth of seeing-the-other".[744] There is no possibility of thinking of another is an unreachable subjectivity. The other person is defined by his relation to the world and to oneself.

Sartre argues that the look of the other is a constant feature of our experience. This look sets in motion the process of our becoming an object for others and reveals the existence of the other as beyond doubt.[745] Sartre's notion of the look encompasses more than just the moment when someone literally turns their eyes towards oneself.[746] It can also be denoted by occurrences such as a rustling of branches, the slight opening of a shutter or even by the windows on a building.[747] These all represent the eye which is first apprehended as support for the look rather than a

---

[742] Ibid P.257
[743] Ibid
[744] ibid
[745] Ibid P.282
[746] Ibid.257
[747] Ibid P-257-8

sensible organ of vision. One apprehends immediately that one is seen rather than that someone is there.[748]

Sartre illustrates his position with the example of looking through a keyhole with one's ear pressed to the door. At first in this example Sartre is alone and uninhibited. Sartre claims that in such a case there is no self behind the action; one is entirely absorbed in the spectacle presented. For Sartre the situation is totally transformed through the sound of footsteps in the hall. Sartre claims "I now exist as myself for my unreflective consciousness.[749] Sartre can see himself because another can see him. Sartre argues there is an affective experience of shame or pride that reveals the look of the other to one and also reveals oneself at the end of the look.[750] Shame comes from recognition that one is the object being observed and judged by the other. Shame here does not refer to a particular shame one might feel for a guilty act but stems from the recognition that one has been transformed into the degrading state of objectivity through the look of the other.[751] Pride is an alternative reaction built on the foundation of shame. It recognizes the other subjectivity as the cause of one's objectification but also recognizes one's own responsibility for this state and embraces it.[752] The others look embraces not only Sartre but also the door and keyhole which were previously instrumental objects for Sartre but are now orientated towards the other.

---

[748] Ibid P.259
[749] Ibid P.260
[750] Ibid P.261
[751] Ibid P.288
[752] Ibid P.290

308

It is not the actual look of the other that is doing the work here so much as the ever present possibility of the others look. If Sartre were to discover the hallway was deserted the sound of footsteps will nonetheless have accomplished the revelation of the presence of the other who is present all around him; above him below and in the neighbouring rooms.[753] While Sartre can be mistaken about a particular presence it is the others facticity that matters. Furthermore the absence of another human being is just a mode of being present because the presence of the other is experienced in terms of human space rather than in terms of spatio-temporal relations.[754] Sartre claims the presence of the other is an original presence. Being for others is a constant fact of human reality.[755] The other is always present to one as the entity through whom one becomes an object. One can discover that what one takes to be a figure watching you is just a tree trunk. However this would not alter the truth of the presence of all men to oneself. The certainty one has about the existence of others is independent of such experiences and makes them possible.[756]

Sartre claims that the self that is revealed to one through the look of the other possesses a degree of indeterminacy.[757] Correspondingly one cannot totally know the other. The other is grasped as an undifferentiated totality.[758] The others freedom imposes limits on Sartre's own. The very presence of another reveals to Sartre that he has an outside, placing him in an intersubjective world.[759] Prior to the presence of

---

[753] Ibid P.277
[754] Ibid P.278-279
[755] Ibid P.280
[756] Ibid
[757] Ibid P.262
[758] Ibid P.292
[759] Ibid P.263

309

another subjectivity in the hallway Sartre grasped his environment in terms of the potentialities items offer for his purposes, for example a dark hall was a place to hide. However the look of the other superimposes new meanings over those Sartre has bestowed on his surroundings.[760] The others look reorganises the world which can no longer be seen by Sartre purely in terms of his own interests and projects which are now succeeded by those of the other. To apprehend oneself as seen is to apprehend oneself as seen in the world and in relation to its other contents. The hallway, which was a place to hide is now also somewhere where the other may direct his flashlight on to Sartre.[761] The others look also pulls one into a spatio temporal relations to the world.[762] It causes one to live one's own experience as "fixed in the midst of the world".[763] The other, then is a precondition for having a world. The other "causes there to be a world by temporalizing himself towards his own possibilities".[764]

Sartre explains the other is the "being through whom I gain my objectness".[765] To conceive of oneself in objective terms presupposes the other not as an object but as a subjectivity. The experience of the other is immediate and is not to be understood as an item in the world.[766] Sartre claims the other is in "an original relation of being

---

[760] Ibid
[761] Ibid P.264
[762] Ibid P.266
[763] Ibid P.268
[764] Ibid P.271
[765] Ibid P.270
[766] Ibid P.273

310

with me" whose indisputable and factual necessity belongs to one's own

consciousness .[767]

For Sartre our experience of others always involves conflict. While the look of the

other turns one into an object the converse is also true. One reinforces one's sense of

self by objectifying the other which is a refusal to acknowledge their subjectivity.[768]

One constantly strives to contain the other within the bounds of their objectivity in

all one's relations to them.[769] However when one finds oneself on the receiving end

of the look this deprives one of any power to objectify the other.

Merleau-Ponty does not accept that the look will make you feel shame or conflict

but accepts the immediacy of the gaze of another subjectivity. Even though he does

not accept Sartre's conflicted account of our relations with others, he still accepts

that we get an immediate experience of another subjectivity.

Throughout the phenomenological tradition individual subjectivity requires,

constitutively, an affective awareness of the consciousness of others. This is the case

for Heidegger's everyday being- in- the-world. It is there in Sartre's account of the

experience of the world as an intersubjective world and ourselves as available to the

perception of others. This viewpoint is also found in Merlaeau-Ponty's account of

the intertwined subjectivities necessary for psychological concepts, and concepts of

self to emerge. Consequently there is not cognitively privileged relation to one's

---

[767] Ibid P.275
[768] Ibid P.283
[769] Ibid P.297

own states on which a special epistemological doubt about the states of others could be based. Rather there is an affective intertwining from which conceptions of self and other; and psychological characteristics for both emerge.

Expression

Merleau-Ponty thinks we have a direct perception of psychological expressions as psychological phenomena. He focuses on expressive gesture as a way in which human beings are able to read the intentional states of one another. To understand why it is useful to consider what is involved in performing a gesture. Merleau-Ponty argues that the intention manifest in a gesture is not a thought prepared within one.[770] Merleau-Ponty also suggests one does not perceive the meaning of their gesture through a bodily signal. The movement is already bound up with the world it occurs in. Perception and movement form a system. Merleau-Ponty gives the example of gesturing to a friend to approach. Features of the world, including the friend's response, are themselves apparent in the gesture. The friend's reluctance to approach infuses the gesture with an impatient quality. He also uses the example of smiling. Here the smile, the relaxed face and the gaiety of gesture, embody the rhythm of action or way of inhabiting the world which constitute joy.[771] The same is also true of linguistic gestures. The spoken word is a genuine gesture.[772] Speech, Merleau-Ponty emphasises, is not a sign of thought but a vehicle of meaning.[773] A

---

[770] Merleau-Ponty 1962 P.111
[771] Ibid P.186
[772] Ibid P.183
[773] Ibid P.178

'phonetic gesture' brings about the structural coordination of experience both for the speaker and for the listener.[774] This parallels the way in which bodily behaviour endows surrounding objects with an intersubjective significance.[775] The faces of other people are also instruments of expression. They can carry another's existence in a manner parallel to that in which one's own existence is carried by one's body.

The role of the body

Merleau-Ponty argues it is one's body that perceives the bodies of others. It is as embodied subjects that people confront one another. Merleau-Ponty argues a conception of the body as object is arrived at by a process of impoverishment of our primordial experience of "the body for us" by which he means the body of human experience.[776] He claims one's body discovers "a miraculous prolongation of my own intentions, a familiar way of dealing with the world" in the bodies of others.[777] Merleau-Ponty understands the body as a "focal point of living meanings".[778] The body and the psyche are involved in a relationship of reciprocal expression.[779] The body opens one up to the world and to other selves by placing one in a particular situation.[780] To have a body is to be connected to a particular world.[781] The body is

---

[774] Ibid P.193
[775] Ibid
[776] Ibid P.351
[777] Ibid P.354
[778] Ibid P.151
[779] Ibid P.160
[780] Ibid P.165
[781] Ibid P.148

313

also the basis of coexistence.[782] It is through one's body that one understands other people, just as it is the body that enables one to perceive things.[783]

For Merleau-Ponty the expressive qualities of human beings are bodily features, but they are physiological features but physiognomic ones.[784] We encounter smiles not movements of the lips. Furthermore, rather than being understood through an intellectual process gestures are understood through one's own body. This is possible because other people's bodies are already apprehended as like one's own. This apprehension is pre-intellectual. This can be seen in cases of imitation. Another agent's body parts are immediately identified with one's own.[785] Merleau-Ponty argues the imitating subject "projects himself or loses his separate reality in the other".[786] It is possible to employ gestures communicatively and to comprehend others gestures because a reciprocal relationship holds between one's own intentions and the gestures of others, and between one's own gestures and the discernable intentions of others.[787] Merleau-Ponty claims that "It is as if the other person's intention inhabited my body and mine his".[788] A visible gesture "outlines an intentional object" that is genuinely present and comprehended by means of one's body.[789] The body achieves this by adjusting itself to this object and overlapping it. The meaning of a gesture is interwoven with the structure of the world elucidated

---

[782] Ibid P.165
[783] Ibid P.186
[784] Whereas physiology deals with the mechanics of bodily animation physiognomy is concerned with reading human characteristics including emotional and moral qualities that are written into the expressive qualities of the human body and particularly the face.
[785] Ibid P.141
[786] ibid
[787] Ibid P.185
[788] Ibid
[789] Ibid

through the gesture.[790] The meaning is "arrayed all over the gesture itself" rather than behind it.[791]

Merleau-Ponty claims there is a cultural aspect to emotional behaviour. There are behavioural variations to the ways in which emotions are manifested across cultures. Merleau-Ponty's example perhaps overstates the case; he suggests Japanese people express anger by smiling. He suggests such differences correspond to subtle differences in the character of the particular emotions. Cultural influences affect the way one meets a particular situation.[792] Rather than being purely a matter of the body understood at a biological level of description, the use made of one's body in comprehending another is also important. Body and world are simultaneously revealed through emotional gesture.[793] The use human beings make of their bodies transcends the biological relations.[794] Bodies are expressive and gestural. They signify one's intentional relations towards others and towards the world at large. The patterns manifested in one's bodily behaviour bestow the world around one with significance both for oneself and for others.[795] In this way they facilitate a shared intersubjectivity. Comprehension of other gestures presupposes perception of a common world.[796] Furthermore expressive bodies allow direct recognition of others as intentional entities. Merleau-Ponty argues that the bodies of others are not inhabited by psyches but are themselves living bodies experienced as manifesting

---

[790] Ibid P.186
[791] Ibid
[792] Ibid P.189
[793] Ibid
[794] Ibid
[795] Ibid P.193
[796] Ibid P.194

consciousness.[797] Merleau-Ponty claims that if one experiences the inhering of one's consciousness in its body and in its world the plurality of consciousnesses presents no difficulties.[798] In other words if one's fundamental understanding of one's own consciousness is of an embodied phenomena that is tied to an intersubjective world, the fact that other consciousness's enter into one's experience is not miraculous.

Merleau-Ponty denies our understanding of others involves anything resembling reasoning by analogy.[799] Such reasoning would involve an ability to draw comparisons between the emotional expressions of others and one's own. In fact, Merleau-Ponty argues, perception of others is a prior condition of such observations.[800] He claims that a fifteen month old infant immediately experiences both its own mouth and the mouths of others as apparatus for biting. Biting already has an intersubjective significance.[801] There exists an internal relation between the way one experiences one's own phenomenological body and the way one experiences other agent's bodies which causes other bodies to appear "as the completion of a system".[802] Bodies appear as "manifestations of behaviour" and not as objects.[803] We find ourselves "already in communication with others taken as similar psycho-physical subjects".[804] As soon as we are confronted by the body of another it becomes for us a particular viewpoint on to the world. Merleau-Ponty

---

[797] Ibid P.349
[798] Ibid P.351
[799] Ibid P.352
[800] Ibid
[801] Ibid
[802] Ibid P.252
[803] Ibid
[804] Ibid P.253

316

describes this viewpoint as "the theatre of a certain process of elaboration".[805] Other bodies are never experienced as mere components of the world.

In Summary, Merleau-Ponty views psyches as ways of inhabiting the world. Rather than attempting to impose meaning on discreet units of sensory data we perceive an already organised totality. This makes space for the possibility of common bodily experiences between self and others. It is embodied gestures that exhibit intentional states. They do not act as mere signifiers of unobservable intentions. Expressive qualities belong to the body and they are understood through the body rather than through an intellectual process. The ability to gesture presupposes a common language between an observer's own intentions and the intentional bodily displays of other agents, that bestow the world with a shared significance.

The intertwining of self and others

Merleau-Ponty does more than close off the gulf between self and others, for him one's own experience and that of others are not just in close proximity. They begin, for each individual, as intertwined and are only separated later when we are capable of making finer differentiations. It is crucial to Merleau-Ponty's account of understanding of others that we do not assign an epistemic privilege to the first-person. Merleau-Ponty claims we begin in a state where we are not aware of ourselves and others as different beings. Strictly speaking we cannot at this stage

---

[805] Ibid P.353

317

attribute to the child either a full sense of himself or of the other. Though there is some understanding of the other as a psyche, the self-other distinction is not yet robust enough for genuine communication to be possible. Genuine communication presupposes a sharp distinction between a communicator and the subject to whom they communicate. Our initial state is one of pre-communication. In this state there are no individuals but only an "anonymous collectivity" undifferentiated by the concepts of self and other.[806] A child's personality at this stage is immersed in the situation. The situation will be a product of the child or of the others who share its existence.[807] For example children of this age only recognize their father if he appears in a customary setting. Merleau-Ponty argues children confuse themselves with their situation.[808] Children gradually become aware of themselves as distinct entities as they become aware of their body and of what radically distinguishes it from the others body. They come to understand that it is closed in on itself. This realization is facilitated by acquisition of the visual image of their body, for example in mirrors. This furnishes children with an objectified image of their own body. This discloses the insularity and difference of their body, and correspondingly those of others.

As this happens Merleau-Ponty claims: "I begin to live my intentions in the facial expressions of the other and likewise begin to live the others volitions in my own gestures".[809] This differentiation is what enables there to be genuine communication

---

[806] Merleau-Ponty 1964 P.119
[807] Ibid P.146
[808] Ibid
[809] Ibid P.119

between two distinct individuals; but it is the undifferentiated commonality preceding this stage of development that provides the grounding. For Merleau-Ponty consciousness of oneself as a unique individual is not primitive. It only occurs at a late stage in our development. Merleau-Ponty also claims children begin by imitating actions or conducts rather than imitating other people. The process of coming to isolate oneself from others in the understanding, which involves both the objectification of one's own body and the constitution of others in their difference, is never completed. We retain the ability to transcend this isolation through our capacity for sympathy.[810] Merleau-Ponty claims sympathy presupposes the absence of a genuine distinction between self and other.[811] It does not imply a distinction between self consciousness and consciousness of the other. It is a manifestation of what Merleau-Ponty terms the 'me and other' system. He argues "It is the simple fact that I live in the facial expressions of the other as I feel him living in mine".[812] Merleau-Ponty distinguishes between child and adult versions of sympathy. Initially, sympathy in children stems from ignorance of one's own boundaries rather than from perception of others. Adults on the other hand, are aware of their boundaries but retain the power to cross them. Adult sympathy does not abolish the differences between oneself and the other.

Merleau-Ponty explains that becoming aware of one's embodiment and becoming aware the bodies of others are animated by psyches are not just logically related operations, they form a real system. Both involve becoming conscious of

---

[810] Ibid P.120
[811] Ibid P.146
[812] Ibid

'incarnation'.[813] They are complimentary operations, the experience of one's own

body and of the bodies of others 'form a totality and constitute a "form"'.[814]

Although they are internally linked operations, this does not mean they must occur at

the same time. In fact the perception of one's own body occurs first.[815] Children pay

attention to their own bodies earlier that they pay attention to the expressions of

others. The operations form a system that is articulated over time, developing in

accordance with a law of internal equilibrium. Exclusive focus on perception of

one's own body creates an imbalance as it develops and fosters in a phase where a

focus on perception of the other dominates. There is a pendulum like movement

back and forth between these two phases. They can only be emphasised in

succession. Each phase of development contains the seeds of its successor.

Throughout development our living relationship with others is the vehicle of our

intelligence. This has important implications for the methodology of any

investigation into our perception of others. We must understand our relation with

others "not only as one of the contents of our experience but also as an actual

structure in its own right".[816]

To sum up, a common idea of intentionality precedes the ability to properly

distinguish self and other. Distinguishing oneself from others is a developmental

achievement, so one's own case is not epistemically privileged in Merleau-Ponty's

---

[813] Ibid P.120
[814] Ibid
[815] Ibid P.121
[816] Ibid P.140

account. The conception of self as a private and bounded entity is an abstraction from this initial state. To arrive at this abstraction an infant must utilise a grasp of others as fellow intentional subjects. The individualised understanding of oneself arrived at is an understanding of an embodied subject rather than the understanding of a Cartesian psyche which would require further abstraction. The process of abstracting oneself from others is never complete.

Understanding of others informs all perception

One reason there is never a possibility of privileging one's first-personal experience in an account of intentional understanding is because awareness of others comes in at ground level. To have a world at all is already to have a world one shares with others. Merleau-Ponty gives the process of learning to perceive others a privileged place in the development of perception in general. He argues that a child's perception is not mere reflection on external phenomena or a straightforward sorting of sense data.[817] It is a more fundamental process in which the child organizes his experiences of external events. This process is not a logical or a predictive one.[818] The function of organizing experience is prior to what are understood as 'standard' cognitive functions like perception as well as intelligence and imagination.[819] Affectivity is not subordinate to these more general functions. Merleau-Ponty argues that even the most seemingly neutral and abstract aspects of our perception, those dealing with sense qualities and space are fundamentally shaped by the child's

---

[817] Ibid P.98
[818] Ibid
[819] Ibid P.99

321

personality and the interpersonal relationships in which that child is involved.[820] He

stresses that there is no pure perception, separable even theoretically from social

conditioning.[821] Functions of the intellect and the subject's relation to society are

part of a unified global project.[822] This project establishes the subjects relations with

the neutral perceptual fields provided in experience simultaneously with his relations

with his human and social surroundings.[823] Intellectual assembly of one's world

receives constant guidance from the affective assembling of one's inter-human

relationships.[824]

The point Merleau-Ponty is making is that cognitive functions including perception

are correlated with the child's structuring of the social world. He is not suggesting a

causal relationship.[825] In fact he suggests that questions about the temporal ordering

of these occurrences are meaningless. There is never a point where either

phenomenon occurs in isolation from the other. There is never a pure state of

perceiving in isolation from social conditioning or vice versa.[826] An individual is not

externally moulded but "takes a position in the face of external conditions".[827]

Merleau-Ponty examines the origins of our perception and understanding of others.

He makes a case for seeing these as intrinsic to perception and understanding in

general. Merleau-Ponty warns against thinking of this study as a specialised or

---

[820] Ibid P.100
[821] Ibid P.108
[822] Ibid
[823] Ibid
[824] Ibid PP.112-113
[825] Ibid P.107
[826] Ibid P.108
[827] Ibid

secondary study to be answered following a more general investigation into understanding. This would be a mistake because the way in which we become aware of others is constituted interdependently with the way in which the world and the self is disclosed.

How can Merleau-Ponty be used to sophisticate Gallagher's position?

A difficulty raised with Gallagher's position concerned an ambiguity whether we confront an epistemological question about perceiving others' psychological states. This ambiguity does not arise in Merleau-Ponty. In his work no epistemological question is possible. Perceiving others involves discovering oneself in relationships to other consciousnesses. Other agents are grasped within the context of the intersubjective possibilities they afford us. There is no question of first satisfying ourselves that they have minds. In fact the epistemological question becomes unintelligible for Merleau-Ponty as we cannot even formulate it without assuming the truth of its conclusion. We have no independent standpoint of 3rd person observational data about which we can speculate as to whether it includes evidence of other subjectivities. The world of empirical sense data is itself shown to be a kind of fiction. The world we actually experience is already meaningful and significant. It is also a world which is experienced as intersubjectivly shared. An intersubjective world is itself a requirement for perception.

Gallagher and Merleau-Ponty share a common strand; both view interpersonal

meaning as constituted within social interaction and view the expressive character of our embodiment as key to facilitating these interactions. However there is also a crucial difference between the two thinkers. Gallagher does not distinguish perception of people in a significant way from perception of objects instead treating them as analogous to perception of objects. This makes identifying psychological states analogous to identifying physical patterns in the environment. For example seeing a distressed facial expression is like seeing a red car. However for Merleau-Ponty the way we experience and engage with other subjectivities has a completely different character. His ideas belong with a phenomenological tradition which envisions our perception of others as radically different to our perception of objects. For Merleau-Ponty our primordial experiences begins within an intersubjective space.

What are the implications of Merleau-Ponty's insights for TOM?

It is time to consider whether what Merleau-Ponty has to say enables criticisms of TT and ST. There are limitations to the usefulness of Merleau-Ponty's ideas for the TOM debate. His equation of the classical model with scientific theories is outdated. His arguments are constructed against empiricism whereas scientific thinking has progressed. Neither TT nor ST is committed to the idea that our knowledge of others is constructed using elementary particles of sense data. They do not claim to use a point by point correspondence between the felt body and the observed body of the other. In fact TT employs a notion of theory laden perception, and sub-personal ST

suggests we are perceptually sensitive to psychologically characterised data rather than brute sensory data. Both then are capable of starting with data grasped under psychological characterisations. This raises a question: if Merleau-Ponty has assumed an empiricist philosophy of perception and is therefore constructing an argument from analogy on that basis then does this assumption invalidate what he has to say given that TOM has no commitments to this framework?

I will argue that Merleau-Ponty's key arguments can still be applied to the TOM debate.

ST

Merleau-Ponty's work raises problems for the central features of ST. He undermines the first-person/third-person dichotomy ST is working with which allows the problem of other minds to be raised. This negates the need for analogical arguments including the one employed by standard ST. The expressive nature of bodies allows other agents to appear as carriers of intentional significance. There is no need to attempt to forge comparisons between radically different phenomena (introspective experience and observational data). Merleau-Ponty goes so far as to suggest perceiving others as psychological agents is a precondition for observation in the first place. Observation is an embodied activity that already presupposes coexistence with other living bodies. Awareness of others is part of the structure of awareness itself.

A related aspect of this is the role of first-person introspection in understanding others. This is central in ST as well as in the traditional argument from analogy. Merleau-Ponty does not assign a privileged epistemic role to first-person experience when making sense of others. This distinguishes his account from mainstream simulation accounts. Merleau-Ponty undermines this starting point decisively by rejecting the notion of a psyche that is uniquely disclosed to an individual. By replacing this notion with the idea of psychologically expressive bodies he also removes a key motivation for inferential strategies because the minds of others are no longer conceived as radically inaccessible. Instead psychological states themselves provide a common medium of representation between the self and others. They enable us to inhabit a shared public space and instil it with mutual significance. The notion of a private self used in ST is an abstraction from an original intersubjective state.

TT

Merleau-Ponty also offers insights that also have implications for TT. Merleau-Ponty attacks the assumption that the bodily appearances of others provide only indirect access to their psychological states. He also shows how if this assumption were true we would be unable to make any progress towards understanding others. Merleau-Ponty opposes the idea hidden psychological data could be inferred from physiological data.

Merleau-Ponty also opposes the idea understanding others should be understood as

a cognitive or intellectual feat which have implications for personal level version of TT. Instead it is the body that perceives the bodies of other agents. Furthermore as embodied subjects our focus is not on the other agent but on the common world.

Furthermore residual features of the classical model still structure the TOM debate. Consciousness is still understood as a private inner event. Merleau-Ponty shows that psychological attributions neither privileges first-person experiences nor involves any kind of inference. He thus brings the validity of the whole TOM debate into question.

In the next chapter I will explore how such validity is also brought into question by the writing of Wittgenstein.

6 Wittgenstein

In this chapter I will put Wittgenstein's ideas to use in two ways: firstly I will use him to formulate a number of objections in relation to TOM. I will show that we can extract a number of arguments against the possibility one can get from a conception of the mental conceived as inner to the attribution of psychological states to others. Much of this is in the same spirit as objections Gallagher raises in relation to the TOM debate which were examined in the previous chapter and can serve to enhance these arguments. However I will also be using Wittgenstein in a second way. This is to draw attention to and to correct some over-simplification in Gallagher's account of direct intersubjective perception. Whereas Gallagher portrays perception of intentional states as on a par with perception in general, Wittgenstein chooses to emphasise important differences between intersubjective understanding and perception of objects. Though both philosophers are concerned with showing that there is nothing mysterious about our ability to perceive intentional states; Wittgenstein's account still allows for the distinctive character of psychological states. Wittgenstein's position bears similarities to aspects of Merleau-Ponty's work explored in the previous chapter. Both seek to make an appreciation of other psychological agents prior to any epistemological doubts we might later raise, although there are also subtle differences between the two.

<u>The inner/outer distinction</u>

Wittgenstein drew attention to a misleading conception of psychological states prevalent in philosophy. It is a conception that is present in at least some versions of TOM. This conception arises because we adopt an introspective approach to understanding psychological states. This introspective approach leads us to treat them as radically different from the other kinds of phenomena we experience. We are seduced into thinking psychological terms describe unobservable inner states. Once we are seduced by this picture it becomes very difficult to conceive of alternatives to the dilemma it presents to us.[828] Because of this misleading conception of the psychological it comes to seem paradoxical to speak of psychological states and physical states together even though such reports as "he suffered great torments and tossed about restlessly" feature unproblematically in everyday discourse.[829] The torments experienced by the protagonist and the tossing about become viewed as quite different kinds of thing.

To clarify the misleading conception of psychological states he wishes to reject, Wittgenstein approaches the question of what is meant by the psychological from a different angle. He considers what is meant by the belief that men have souls.[830] He examines what happens if we bring an introspective analysis of psychological states to bear on this question. He suggests that "a picture is conjured up which seems to

---

[828] Hutto 2008 P.199
[829] Wittgenstein 1968 421
[830] Ibid 422

fix the sense unambiguously".[831] It is a picture of what McGinn terms the inner.

Sensations and thoughts are the product of an internal realm of states and

processes.[832] However when we try and apply this picture during the stream of life

its application turns out to be problematic. This picture suggests our access to the

psychological states of others will at best be indirect.[833] This picture then comes into

friction with the ease with which we normally grasp other people's psychological

expressions. If psychological states are something known only from inner

experience then it is hard to see how they could be given an external application.

Instead of realising this problem appears because we are working with a flawed

picture of the psychological we erroneously conclude that the psychological itself is

problematic. Another problem with picturing the psychological as involving

unobservable inner states is that it goes hand in hand with an impoverished picture

of our perception of the physical bodies of other people. These come to be thought

of as objects of mere physical description unconnected with psychological

phenomena.

Extending your feeling into others

Wittgenstein's position can be illustrated through attention to a particular issue.

There is an interesting difficulty attached to the conception of psychological states as

inner: if psychological states were really inner phenomena known by introspection,

how could they get attributed to other people? Wittgenstein is clear that the answer

---

[831] Ibid 426
[832] Mcginn 1996 P.147
[833] Ibid P.149

to this problem could not be to mentally transfer what one has oneself to another person.

Wittgenstein's work offers a challenge both to the idea that we can anchor the meaning of psychological states in first-personal experience and to the idea that such anchorage could be employed in a resolution to the problem of other minds. He asks where we could get the idea that other people can feel. He considers the following suggestion: "Is it that my education has led me to it by drawing my attention to feelings in myself, and now I transfer the ideas to objects outside myself?".[834]

He goes on to reject this suggestion. One reason this suggestion is unsatisfactory is that it does not take into account what is special about living things and the way they appear to us. As Wittgenstein notes we do not transfer psychological concepts to stones or to plants.[835] But this is exactly the sort of thing human bodies become on the inner /outer picture. There is no essential connection between psychological states and living bodies. The point here is not that an analogical argument from one's own case will be weak but that there is nothing on which such a strategy can get a grip.

As McGinn notes this is also to attribute a special role to the imagination in assigning psychological states to others (she is specifically interested in sensations but I think the point of these remarks can be broadened to include other types of

---

[834] Wittgenstein 1968 283
[835] Ibid

psychological state). A concern is how one's own experience can serve as a model

for someone else's experience.   Wittgenstein argues:

"If one has to imagine someone else's pain on the model of one's own, this is none

too easy a thing to do: for I have to imagine pain which I do not feel on the model of

the pain which I do feel".[836]

This involves a particular picture of what is involved in understanding another's

psychological states. We form an image of the psychological state which we then

transfer to the other person.[837] But to form an image of a psychological experience is

just to imagine experiencing it.

The difficulty of transferring inner states to others is not confined to processes

involving the imagination though; it affects any form of mind reading based on first-

personal experience. Wittgenstein's view was that modelling third-person attribution

of intentional states on first-personal ones will not give you what you need:

"Believing that one can simply extend the idea of what is immediately felt into other

people's bodies will give you only an idea of having feelings in their bodies, not of

their having feelings".[838] In support of this Ter Hark cites a 1932 manuscript from

Wittgenstein:

[836] Ibid 302
[837] McGinn 1996 P.172
[838] Ter Hark 1991 P.235

"when I apply this concept in the case of pain I will never come from my pain to his but from my pain in my tooth to my pain in his tooth".[839]

One problem is that imaginatively transferring a pain one experiences in one's own tooth to someone else's tooth is not like imaginatively transferring a pain one experiences in one's tooth to one's arm. We can be misled by a superficial similarity between the two cases. The latter case is coherent but the former is not. Elsewhere Wittgenstein makes the same point with a different example. He imagines being told:"But if I suppose that someone has a pain, then I am simply supposing that he has just the same as I have so often had".[840] Wittgenstein compares this remark with: "You surely know what 'It is 5 o'clock here' means; so you also know what 'It's 5 o'clock on the sun' means".[841] This latter statement is superficially like admitting you know what it is to be five o clock in Wales. In such a case one can base this understanding on what it is to be 5 o clock here. However in the cases of another's having the same pain and of it being 5 o clock on the sun Wittgenstein argues the "explanation by means of *identity* does not work".[842] One does not know what it means to speak of it being the same time here and on the sun. In the same way closer inspection of the idea that a supposition that another is in pain is equivalent to a supposition that he has the same thing you do will reveal that we lack any story to anchor this statement. This can be seen when we realise an equally legitimate extension from one's own case would be to say "the stove has the same experience

---

[839] Wittgenstein 1932 manuscript (quoted in Ter Hark 1991)
[840] Wittgenstein 1968 350
[841] Ibid
[842] Ibid

333

as I".[843] Although our imagination may be able to sketch in some details in such cases (for example imagining a grandfather clock on the sun pointing to 5)[844] there is no explanation of how we could have *public* content.

The point is not that we cannot imagine another's intentional states, only that we would not be able to imagine them if the model of the inner were accurate. If psychological states were private objects that had no essential connection with bodies then their application could not coherently be extended to other bodies. To imagine another person in pain, for example, does not involve appealing to an image of your own pain.[845] In summary, if we accept a conception of psychological states as inner then the only way to attribute psychological states to other people would be to try and extend these inner concepts to the outer bodies of others. However this cannot be achieved as bodies here are not the sort of thing on which psychological concepts can get a hold. The best we can do is to imagine having our experiences inside their bodies rather than inside our own. This links to question 4 in the introduction which asked how we could ever perform such a miraculous feat as to successfully attribute psychological states to others.

Private language

A collection of remarks that have become known as Wittgenstein's private language argument also reveals concerns that any attempt to locate the meaning of

---

[843] Ibid
[844] Ibid 351
[845] McGinn 1996 P.172

psychological terms in first-personal experience will not be coherent. Wittgenstein raises the question of whether it would be possible to possess a language in which one could record or express their inner experiences but where the individual words refer to immediate private sensations which can only be known to the user.[846] This would be a language which was by definition unavailable to others.

Wittgenstein attacks the idea that our sensations are really private. He unpacks what this would mean. The idea would be that only the individual can know he experiences a particular sensation whereas other people can only speculate.[847] Wittgenstein argues against this on two fronts both of which are based on an appeal to ordinary usage of language. Firstly it is incorrect to claim that other people could not know that one is, for example, in pain. Secondly it is nonsensical to claim that one *knows* one is in pain. To claim one knows this would add nothing to an assertion that one is in pain. Doubt and certainty belong to the linguistic practice of talking about other people's sensations but not to the practice of talking about one's own.

Wittgenstein imagines a case where a child invents a name for his sensation of toothache. The child in this example inhabits a world in which there is no outward expression of sensation. People do not characteristically groan or pull faces when they are in pain. Even supposing we (provisionally) allow the child has accomplished this feat of 'naming'; Wittgenstein asks us to consider what has really been achieved. In the actual world, naming only has significance because it is

---

[846] Wittgenstein 1968 243
[847] Ibid.246

inextricably linked to a plethora of other practices. Naming assigns a word with a

role in grammar. Wittgenstein explains "when we speak of someone's having given a

name to pain, what is presupposed is the existence of the grammar of the word

"pain"; it shows the post where the new word is stationed".[848] In any case we need

not even provisionally concede that such a practice is coherent as we will see

shortly.

Wittgenstein next imagines a case where one tries to associate one's inner sensation

with an outward sign 'S' in a diary.[849] He wants to see whether such a process can be

given any practical application. He suggests one might be interested in doing this for

the purpose of keeping a record of the occurrence of the sensation. With that end one

writes S on a calendar every time one has the sensation. Wittgenstein argues that this

does not enable one to define S and thus to make it a meaningful term. Introspecting

one's sensation while saying S will not be sufficient to accomplish this. The aim

would be to impress a connection between sign and sensation on oneself in a way

that would allow one to remember the connection for use on future occasions. The

problem is that this process does not provide any criterion for correct usage within

that framework whatever seemed correct to one was correct. And if there is no

distinction between seeming correct and being correct then we have not succeeded in

naming anything. Our psychological concepts could not get their meaning in this

way. In fact we could not even describe the act of writing S on the calendar as

making a note of it. A note has a function but S has not been assigned any

---

[848] Ibid.257
[849] Ibid.258

function.[850] Wittgenstein resists the idea that assigning a private definition could involve an undertaking to use a word in a particular way.[851] He asks whether one is meant to invent the technique of using a word for oneself or to find it ready made. Wittgenstein argues that if the private language user intends to use a word to stand for sensations in the way that they ordinarily do then the user must take his sensation words to be connected with one's natural expressions of sensation. If so we do not have a private language.[852] In principle this language could be adopted by anyone.

We began by provisionally accepting that it at least made sense to think one might invent a name for one's sensation within a private language even if this is not what we in fact do. However we can now see that Wittgenstein is pulling the rug from under this idea. The private language user does not have the right to call S the name of a sensation in the first place. Sensation is a word belonging to common language; it does not denote anything in a language only intelligible to one person.[853] Even if one tries to get out of this difficulty by denying S has to stand for a 'sensation' to be intelligible and instead just has to be a 'something' the problem simply recurs. Something is also a term belonging to public language. Nor will it help if one withholds any attempt to translate S into public language at all and treat it as an inarticulate sound. If words are simply sounds then, according to Wittgenstein, no private operation can convert these sounds into names. A name requires a criterion

---

[850] Ibid.260
[851] Ibid.262
[852] Ibid.256
[853] Ibid.261

for correct and incorrect use otherwise it cannot be picking out a determinate entity.[854]

Wittgenstein's knowledge argument

Wittgenstein argues that it is either wrong or nonsensical to claim one has knowledge about one's own sensations. He is taking issue with the idea that the meaning of sensation words such as pain derives from identification of inner states accessible only to the experiencing subject. On this view "only I can know whether I am really in pain; another person can only surmise it".[855] Wittgenstein argues we often know whether others are in pain. In fact expressions of doubt and certainty are only coherent within the practice of talking about the sensations of other people. To claim one knows that one is in a particular psychological state actually adds nothing to the assertion that one is, in fact, in this state. Making knowledge claims about one's own psychological states is not a coherent move within our shared practices of talking about psychological states because knowledge claims only make sense if there is a logical possibility of them being true or false. In one's own case there is no coherent possibility of 'I am in pain' being made sincerely but incorrectly. It must be stressed that, for Wittgenstein, the reason is not because the speaker has incorrigible access to a private inner state which allows the speaker to provide an accurate account of this. Rather, in such cases, what the speaker is doing is affirming or expressing their pain. They are not offering a description of a state of affairs. If this

---

[854] Ibid
[855] Ibid. 246

is what the speaker was doing they might indeed mis-describe the state. An assertion one is in pain is a modification of the more primitive practice of crying.[856] It is not normally a description of anything. Rather, it is part of the criteria for being in pain. That it gets expressed in this way partially constitutes what makes the experience one of pain.

First/third-person asymmetry

Wittgenstein's philosophy acknowledges an asymmetry in the ways in which the psychological states of others and one's own psychological state are disclosed to one. However the distinctive relationship in which one stands to one's own psychological states is not an epistemologically privileged one. The view that one can claim to know about one's psychological states belongs within an epistemic picture in which these states are private inner states which we have privileged access to. The other side of this picture is that the psychological states of others are irretrievably hidden from us. This picture is the result of misconstruing differences in the way that we use first-personal and third-personal concepts.[857] There is a temptation to think that because one has to pay attention to the behaviour of others to ascribe a psychological state but not to one's own behaviour that this is the result of privileged access. This temptation is reinforced by the observation that one can be mistaken about others but not about one's own psychological states.

---

[856] Ibid.244
[857] Child 2014 P.2

On the contrary Wittgenstein thought it was incoherent to make knowledge claims concerning one's own psychological states. This is because a claim to know, for example that one is in pain, implies a logical possibility of being mistaken but we cannot make sense of this possibility here. For Wittgenstein, the primary uses of expressions such as "I'm in pain" are not descriptions of inner events but avowals. They express one's state of pain rather than describe it. In fact the expression is part of what constitutes being in that particular state. It is a modified form of a natural expression of pain. The fact that the phrase "I'm in pain" is an expression rather than a description of a particular state of affairs explains why one can't be incorrect. As we saw a moment ago expressions of doubt and certainty are only found within the practice of attributing psychological states to others. The minds of others are not hidden but the way they are revealed to one (through their behaviour and what they say) is different and there is room for error in one's psychological attributions.

Although this is the primary use of avowals there are still cases where it is coherent to make assertions about one's own psychological states. For example one might inform one's doctor of a pain during a medical check-up. While there are contexts where one might appropriately describe one's psychological state, this use of psychological terminology is secondary to the expressive practice.

It should be stressed that this asymmetry in Wittgenstein's treatment of psychological states is not equivalent to an asymmetry resulting from a Cartesian inner realm, transparent to the first-person and absolutely private. In fact for Wittgenstein it does not make sense to talk about knowing in the case of one's own

experience.

Wittgenstein's alternative to the inner

This chapter will now outline Wittgenstein's alternative suggestion as to how we gain understanding of the nature of psychological states. Rather than attempting an introspective analysis of psychological states Wittgenstein suggests we should look at how psychological terms such as pain or anger appear in ordinary discourse. This will remove the temptation to think such concepts always function in a uniform way. One reason the picture of the inner is not coherent is because "an 'inner process' stands in need of outward criteria".[858] This is not a claim that we require contingently related outer behaviour to justify the attribution in inner states. Rather, Wittgenstein is pointing to the internal relations psychological states bear to their external manifestations. As Hutto explains:

"For Wittgenstein the kind of behaviour that serves as a criterion for the psychological is not the mechanical or thoughtless variety; it is always and everywhere already enlivened, expressive and mindful".[859]

The difficulty explaining our ability to ascribe psychological states to other people when such states are conceived as inner does not arise on Wittgenstein's position:

---

[858] Wittgenstein 1968 580
[859] Hutto 2008 P.201

"The expression of doubt has no place in the language game but if we cut out human behaviour, which is the expression of sensation, it looks as if I might *legitimately* begin to doubt afresh".[860]

To combat the sense that statements which reference psychological and physical states simultaneously are paradoxical, Wittgenstein recommends comparing them to statements of the form "these three struts give the building stability".[861] In this statement we can see that 'struts' and 'stability' describe different aspects of the same phenomena. Stability is not an additional object that can be placed alongside the three struts. Yet there is nothing problematic about speaking about struts and stability in the same sentence. It is clear that we are still talking about the same thing; the building. We can describe different aspects of a thing at the same time.

Attention to real situations in which we confront other human beings shows that a picture in which we confront physical bodies unconnected with psychological phenomena is not true to life: "think of the recognition of *facial expressions*. Or of the description of facial expressions – which does not consist in giving the measurements of the face!".[862]

When Wittgenstein refers to giving the measurements of the face part of his point is that recognizing psychological states in other people's faces is not a question of deciphering another's expression through a bottom up analysis of the physical

[860] Wittgenstein 1968 288
[861] Ibid.421
[862] Ibid.285

relations that hold between the different parts of the face. Like Merleau-Ponty he thinks such a methodology would be doomed to fail. We see first of all the expression of another human being. A human face is already permeated with the expression rather than merely associated with it. In support of this claim Wittgenstein points out that imitating another's facial expression is not a matter of attempting to match the spatial coordinates of his face but a top down process led by the others expression itself: "Think, too, how one can imitate a man's face without seeing one's own in a mirror".[863] What we are sensitive to is not bodily movement. We successfully imitate another without an awareness of the requisite bodily movements. A related point is that the level at which we initially engage with others is not the level of mechanical movement. As Cockburn notes, we spontaneously respond to smiles and angry glances.[864] If asked to describe the smile we spontaneously responded to a moment ago in terms of bodily movement one would probably struggle. We do not experience other bodies as conglomerations of parts. It is not at this level that others are accessible to us.

Wittgenstein is also interested in what it is that makes something an expression of subjectivity. He stresses that it is other people we react to and interact with, not physical body parts: "if someone has a pain in his hand, then the hand does not say so (unless it writes it) and one does not comfort the hand, but the sufferer: one looks into his face".[865] It is human beings, not body parts, which have experiences like pain. Human faces are of special significance for Wittgenstein in demarcating people

---

[863] Ibid
[864] Cockburn 2001 PP.95-96
[865] Wittgenstein 1968 286

from mere objects. For example one looks into the sufferer's face.[866] In the majority

of cases the face serves as the expressive core for a person's psychology.

Wittgenstein also emphasizes the distinctive ways in which we respond to other

people. These responses are radically different to the ways in which we interact with

and think about inanimate objects.

Wittgenstein claims intersubjectivity is constitutive of psychological states. It is tied

in with certain kinds of practice that we share with others (I return to this below).

Wittgenstein describes our attitudes towards another human being in the following

way "My attitude towards him is an attitude towards a soul. I am not of the *opinion*

that he has a soul".[867] An attitude is not an epistemological relation. Wittgenstein

regards these 'attitudes' we have towards others as fundamental. They do not rest on

some further grounding. In particular they do not rest on anything like analogical

arguments to hidden mental states based on similarities between another's bodily

behaviour and one's own. Instead our attitudes are rooted in the practices in which

we engage with other people. An attitude is constitutively tied to a wide range of

practices. These practices embody different ways in which we are interested in and

respond to other people.[868] Cockburn explains that Wittgenstein was reacting against

a tradition in which it was thought to be coherent to make a distinction between

one's thoughts about others and about material objects by claiming to have beliefs

that the former have minds but the latter do not. As Cockburn argues, Wittgenstein is

drawing attention to "the fact that we *feel about* and *act towards* other human beings

---

[866] Ibid
[867] Ibid 2 iv P.178
[868] Cockburn 2001 PP.98-99

in ways that are utterly different from those in which we feel about and act towards, for example, stones".[869] This difference manifests itself as a 'practical orientation' towards others.[870] This means that their behaviour elicits feelings and actions in one and rules out others as entirely inappropriate.[871]

The point is not to contrast opinion with knowledge and argue that we can have sound knowledge of the psychological lives of others. It is to deny we primarily stand in any kind of epistemological relation to such states. This becomes clear if we remind ourselves of the typical practices in which the psychological states of others feature. As Malcolm argues there can be situations where doubt as to the authenticity of another's psychological state do not arise, for example, when a doctor observes a patient groan in pain. This is not because alternative explanations for what the doctor is observing, for example that the patient is acting, have been refuted. In situations like this, questions about the authenticity of what is observed would be dismissed as absurd by anyone involved in the situation. It is not that doubts are impossible but that they are not appropriate. The doctor acts with a kind of certainty in responding to the patient's groans with medical assistance, but this certainty is not anchored in a process of epistemic justification. Rather such certainty is a precondition for the kind of practices in which human beings typically participate, such as treating sick people.

---

[869] Cockburn 1990 P.6
[870] Ibid
[871] Ibid PP.6-7

Alternative explanations for the groaning behaviour are potentially limitless (for example it could be caused by deception, hypnosis or a mindless physical abnormality) although the doctor may have no particular reason to consider any of these possibilities. Malcolm argues that if someone were to doubt that the groan was sincere then this doubt would not be illogical although it would be highly abnormal. However, if our normal practices are to continue, then entertaining such doubts must constitute the exception rather than the rule.[872] In order to sustain the possibility of shared practices such as sympathising with, or tending to another's pain, we are closed to the possibility of such doubts.

Another reason why an attitude is not reducible to an epistemological relation is that there is more than one attitude one could take to the phenomena we observe. Malcolm discusses a thought experiment by Wittgenstein in which a tribe take their slaves to have no souls or feelings. Many of their practices will parallel our own. For example when a slave is sick they will allow the slave time to recuperate. However these practices would be intermingled with other more alien practices such as discussing slaves as if they were machines and discarding them when they became worn. If observing a slave writhe in agony, a member of this tribe would not exhibit any emotional distress. Malcolm explains that the difference in attitude is not a matter of believing or expecting different facts. A belief that another is suffering would be an epistemological relation which could be changed by learning further facts. Instead what we have are characteristic ways of reacting to suffering such as

---

[872] Malcolm 1954 P.547

viewing the sufferer with compassion and trying to comfort them.[873] There is no intermediate process between observing the writhing sufferer and responding in which we have a thought that the pain behaviour is caused by the sufferers pain. There is not even a tacit assumption that the writhing expresses pain.[874] Such an assumption would leave room for doubt and such doubt is not a feature of our pain-practices. This is not to say that our reaction could not take different forms such as a refusal to acknowledge the suffering.

Malcolm also explains that the characteristic expressions, language and activities that constitute our ways of responding to others in pain constitute a form of life. In order to understand our psychological concepts we must examine the behaviour activities and natural expressions surrounding our use of psychological concepts.[875] Malcolm cautions that it would be a mistake to attempt to seek justifications for these forms of life. For example we should not argue that we pity an injured man because we know or believe that in addition to the groans and writhing there is pain.[876] Rather we have to understand that the responses such as pity are taken as appropriate within the forms of life within which the concept of pain has an anchorage.

---

[873] Ibid P.549
[874] Ibid
[875] Ibid P.550
[876] Ibid

McGinn argues that Wittgenstein's use of the term attitude is intended to mark a distinction between the intellectual or judgmental and the visceral.[877] The existence of other psyches is part of the fabric of the forms of life involving others. The fact of other psyches does not enter one's consciousness as an item of knowledge but provides form to the world one inhabits. An attitude is prior to an opinion. Unlike an attitude an opinion is an evidence-based judgment and allows room for disagreement and error.

A distinguishing feature of these practices is that they are normative. Even if one decides not to comfort the sufferer with a painful hand one must at least grasp the appropriateness of such a response. For Wittgenstein this is part of what it is to grasp that somebody is in pain. This differentiates such recognition from the purely causal sequence of events which might facilitate recognition of an object. Cockburn offers a useful example of the normative demands involved in recognising another as angry: "When I learn that someone is angry I learn things such as the following: if I want to avoid a nasty confrontation it would be a good idea to keep out of his way for a while, a good idea not to raise that touchy subject that we need to discuss sometime, and so on".[878]

Furthermore the role a psychological state occupies within a practice is necessary for its identification as a psychological state in the first place. Our perceptual capacities

---

[877] Mcginn (1998) P.54
[878] Cockburn 2001 P.39

could not pick out a psychological state independently of the normative practices in which such states are embedded. Something is identified as, for example, an instance of grief because it calls for certain kinds of sympathetic responses from others. This makes intersubjective perception different from the perception of cars. Again Cockburn puts the matter well:

"the *significance* for us of the fact that someone is in a certain mental state is not exhaustively described in terms of the behaviour that we can expect from her…..
that another has a severe pain in her foot is not simply a reason for thinking that she won't make it to the cinema on foot; it is also a reason for feeling sorry for her and trying to help her".[879]

It is a normative framework rather than a causal one that anchors psychological attributions. One thing to note here is that we cannot focus exclusively on the behavioural patterns manifested by the individual but also have to take into account the social implications of this behaviour.

Wittgenstein is suggesting there is something special about the way we perceive other human beings. He argues that "only of a living human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees;

---

[879] Ibid.P.40

is blind; hears; is deaf; is conscious or unconscious".[880] This difference between human beings and objects is not explained by a distinction between unobservable intentional states and observable mechanical behaviour. Nonetheless the kinds of interactions we engage in with other human subjects have no analogue in our interaction with objects. In fact Wittgenstein goes further in arguing for the impossibility of sustained perception of others as automated objects: "Seeing a living human being as an automaton is analogous to seeing one figure as a limiting case or variant of another; the cross-pieces of a window as a swastika, for example".[881] Although arguably this can be done for limited periods if one feigns a detached attitude to others, Wittgenstein challenges the reader to try and hold on to this way of looking at others during the course of interaction. The most likely outcome of such an attempt is that the idea becomes meaningless. If we do temporarily achieve a detached view of other bodies as mere conglomerations of mechanical parts then our distinctive range of attitudes no longer get a grip. For example we cannot sympathise with a damaged hand but only with a sufferer.[882] When we reduce the hand to the status of body part we are tempted to say that it is something other than the body that has experiences. If we had to rely on inferences from mechanistically envisioned movements to intentional causes then inferring other minds would indeed be problematic. In fact it is only from within the grip of a misguided picture in which minds are inaccessible that we can even begin to see others as automata. Reid connects this with retreating into a kind of fantasy in which

---

[880] Wittgenstein 1968 281
[881] Ibid.420
[882] Ibid.286

350

minds are inner and private.[883] This makes to easier to deny another person their

subjectivity. The point is that there is something deliberate about a refusal to

acknowledge the other as a subjectivity (which imposes ethical demands on us)[884]

that makes the attitude insincere. It is significant that Wittgenstein finds this much

easier to accomplish alone than in the presence of other people.[885] When confronted

with a human face the illusion will tend to vanish or one will experience a sense of

uncanniness. Reid argues one significance of the comparison with seeing the cross

piece of a window as a swastika is that to achieve this effect one has to make oneself

blind to the surrounding window frame.[886] It is the denial of something one does

observe that lets something else appear. As one denies a window one sees a swastika

and as one denies a human being one sees an automaton.

To summarise, Wittgenstein shows that we are not just identifying patterns within

spatial coordinates when attributing psychological states to others. He emphasises

that our sensitivity to others is closely bound up with our engagement in practices

that are unique to psychological beings. Because it is the practices that are central

there is no starting point for raising epistemological questions about the existence of

other minds.

---

[883] Reid (2010) P.597
[884] Reid is arguing that the idea of a private inner realm leads to negative ethical and political
consequences as it facilitates the dehumanization of selected others. It helps us escape the fact of their
humanity making it possible to treat them in an inhuman manner.
[885] Wittgenstein 1968 420
[886] Reid (2010) P.603

General implications for TOM

Wittgenstein's attack on the traditional inner/outer distinction has implications for TOM. It has been claimed that the mistaken picture of the inner that Wittgenstein identifies is central to modern cognitive science and in the philosophy of psychology.[887] This metaphor "wrongly encourages us to imagine a distinct locale wherein special events and processes reside and causally interact, behind the scenes... To fall in with such metaphors of the mind is to promote a mistaken understanding of the "inner".[888] Hutto argues supposedly neutral terms like TOM, mind reading and mentalizing embody precisely the philosophical picture of the mental Wittgenstein wishes to draw our attention to.[889] All positions in the TOM debate take it for granted that psychological ascription proceeds by the positing of inner mental causes.[890]

We can see that this is the kind of position that Wittgenstein had in mind because it is a picture that prevents us from seeing alternative possibilities: "Many researchers are aspect-blind to the fact that these are questionable assumptions and especially to the possibility of other descriptions of this practice".[891]

Wittgenstein claims that, from the perspective of theories working with a picture of the mind as inner, human bodies are treated as analogous to stones and plants. It

---

[887] Hutto 2009 P.200
[888] Ibid
[889] Ibid P.204
[890] Ibid
[891] Ibid P.205

352

would not be totally fair to apply this claim to either TT or ST. In versions of both these theories human bodies invite the attribution of psychological states in a way that stones and plants do not. However in another sense people and stones really are treated as analogous because, in the case of people we are supposed to be responding to mere physiological data, and on the basis of this reach mental attributions through inference.

Wittgenstein's position and TT

Wittgenstein connects the idea of the inner with the idea that states are grasped through first-person introspection. However for TT it is not the case that psychological states are known through introspection. Whereas Wittgenstein suggests attempts to attribute psychological states to others are doomed because they involve a conception of bodies that will not allow psychological predicates a foothold; thus forcing an attempt to extend self experience to others, TT suggests inner psychological concepts can still be applied to bodies. TT is not falling foul of the problems besetting the argument from analogy in the sense that it does not start with the first-person. It starts with theory laden data.

Nor does TT ground the meaning of psychological terms in first-personal experience in the style of the private language argument. The meaning of the psychological terms one employs is public and in principle available to others. This raises a question about the applicability of Wittgenstein's remarks to this theory.

However there are still significant ramifications for TT. TT does still envisage psychological states as inner. Our access to the psychological states of others is still understood as indirect, mediated through theoretical apparatus utilising observation of public behaviour. The relation between the behaviour and the inner states remain contingent and in need of an *inference*. Mental states for TT are still hidden, our own as well as others, and our access comes only via the application of a theory.

TT pays a high price to avoid difficulties associated with the assumption that psychological states are known through introspection. As we saw in chapter 2, TT treats all psychological terms as third-personal. This leads to the problem that it does not capture the distinctive relation a person stands in to their own psychological states. Unlike Wittgenstein, TT offers no account of the special relation to the first-person. Wittgenstein however accepts a $1^{st}/3^{rd}$ person asymmetry. However, he does not make 1st person acquaintance with psychological states a matter of knowledge.

TT and Wittgenstein operate inside contrasting frameworks. The theoretical framework within which psychological states are placed by TT is a scientific one. For Wittgenstein the framework is everyday intersubjective normative behaviour. A term is intersubjective so the response one makes is part of the meaning of the term. To show that a certain response is appropriate is to locate the meaning of the terms within a normative intersubjective framework. This can't be accommodated by TT. It becomes entirely external to, for example, pain that there is a social norm that says we should comfort the person who is in this state.

<u>Wittgenstein position and ST</u>

If (as certain proponents explicitly stated in an earlier chapter) ST purports to offer a solution to the epistemological problem of other minds then its strategy will involve starting from the first-person perspective. Wittgenstein shows that the most this perspective can allow us to accomplish is to attribute copies of our own experiences to others. It also involves the highly dubious practice of treating one's own experiences as items of knowledge which can then be used in inductions about others.

Standard ST conceives of simulation as involving the imagination. It therefore also falls foul of Wittgenstein's more specific objection that assigning to the imagination the role of transferring psychological states to others will not do the job that is required. This question is raised by Wittgenstein's 5 o clock on the sun example. This expresses puzzlement about how we could coherently attribute psychological states of others if our understanding of these states was entirely first-personal.

<u>Wittgenstein position and sub-personal ST</u>

The application of Wittgenstein's diagnosis to sub-personal ST is less clear cut and needs more attention. The question is whether sub-personal ST is trying to address the epistemological question of other minds, or whether it is simply giving a description of empirical conditions which enable our practices of invoking mental

states to others to get off the ground. This thesis has been arguing that the epistemological question is answered at the personal level. It was argued that if ST is simply offering an answer to question 3 (what is necessary empirically for us to be able to engage with others) then it does not confront a philosophical problem. However it will not shed any light on the epistemological question.

There are a number of ways in which sub-personal ST departs from the traditional simulation picture.

1. Imagination

Gallese's version of ST offers an explanation of how we come to understand others that takes place at the sub-personal level. Simulation theorists who adopt this strategy are not guilty of portraying our grasp of the psychological states of others as the result of an implausible imaginative exercise. The theory does not assign a role to the imagination. However, an analogue of Wittgenstein's objection to modelling third-person attribution of intentional states on first-person intentional states still applies to processes occurring at a sub-personal level. This is because, as Gallagher points out, sub-personal ST takes the imaginative process of simulation we find in standard ST and postulates a sub-personal analogue of this process.[892]

Although not conscious representations, MNs are conceived as representations of psychological states. The same representation occurs when you experience being in a given psychological state and when other people undergo the same type of

---

[892] See Gallagher 2007

experience. For example Gallese & Goldman claim: "other people's mental states are represented by adopting their perspective: by tracking or matching their states with resonant states of one's own".[893] Similarly Hurley claims "the simulator copies the states of the other and uses the copies in her own decision-making equipment".[894]

2. The inner/outer model

If ST is using MN activations to offer an account of how we overcome the epistemological problem of other minds then Wittgenstein has raised some problems. If we take MNs to function primarily as representations of our inner psychological experience, which are also transferable to others, then there are issues about attributing what we have to other people. One problem is that this looks like a very weak argument from analogy. Another is that it will not give us what we need. We saw that if we begin by understanding intentional states from an entirely first-person perspective this kind of attribution to others can provide an idea of having feelings in their bodies but not of other agents having feelings. More problematically still the attribution may not be coherent. Recall that Wittgenstein argues attribution to others in terms of identity with oneself is not coherent if we are working with a model of psychological states as inner. If we only conceive psychological states in terms of this model it makes no sense to attribute them to others.

---

[893] Gallese & Goldman 1998 P.493
[894] Hurley 2006 P.20

A defender of ST might object that the use their theory makes of MNs does not fit with the inner/outer picture. There are reasons for thinking that not all versions of ST are adopting the false picture of the inner and the outer that Wittgenstein criticises. I will lay out some of these arguments. This will involve a brief recap of material reviewed in earlier chapters.

One reason is that they do not want to claim psychological understanding is based in first-personal experience. Not all versions of ST place us in a position where we must struggle to apply exclusively first-personal psychological concepts to 'mere bodies'. One reason is that other people are not perceived as 'mere bodies' at the neurological level. MNs are responding to basic intentional states (for example they respond to specific, context embedded intentional actions such as grasping rather than movements, and to sensations and emotions).

A second equally important reason we do not struggle to apply first-personal psychological states is that psychological states are not conceived as exclusively first-personal on all versions of the ST model. We saw that MNs, which are thought to provide representations of psychological states at the neural level, are claimed by implicit ST to provide "shared representations".[895] To recap: MNs fire in one of two conditions: when undergoing a particular intentional state and when observing the equivalent state in another person. MN activation is envisaged by implicit ST to represent the intentional state itself. It makes as much sense on this view to assign a psychological state to someone else as it does to oneself.

[895] Gallese 2008, Jeannerod & Pacherie 2004

358

Evidence for shared representations is presented in the context of we-centric space: "interpersonal relations are established when a full-blown self-conscious subject of experience is not yet constituted".[896] 'We-centric space' helps to establish interpersonal relations prior to the development of a fully self-conscious subject. This intersubjective space provides the individual with a powerful tool to overcome the epistemic gap between him/herself and the other, produced by the establishment of a self-centred perspective".[897] We have an attenuated intersubjectivity before developing our subjectivity. During development, our initial experience of psychological states is not one that clearly demarcates between a state of oneself and a state of another.

We saw earlier that the model of the inner Wittgenstein opposes severs the connection between the bodies of others and their psychological states. Simulation theorists also have a response to this related objection. Sub-personal versions of ST are not committed to the idea that we confront mere bodies which we must then somehow attribute psychological states to. At the neurological level specialised regions distinguish walking, head-turning, torso-bending and arm movement as well as hand-object interactions when observing others.[898] These versions of ST claim the brain is able to distinguish biological occurrences from merely physical ones (even though this is not quite the same as the distinction Wittgenstein wishes to draw between expressive human bodies and material objects).

[896] Gallese 2006 P.16
[897] Ibid
[898] See Sinigaglia & Sparaci 2010 P.5

359

To summarise I have suggested three ways in which a defender of sub-personal ST might distinguish this position from the one Wittgenstein is opposing. Firstly psychological understanding is not based in first-personal experience. Secondly the realm of the psychological is not understood as exclusively first-personal. Thirdly ST is not suggesting we confront mere bodies which we must attribute psychological states to.

Ways in which sub-personal ST involves a model of the inner

I will now argue that features of the model of the inner are still operative in sub-personal versions of ST and therefore Wittgenstein has valid objections against these theories. If the sub-personal data were simply being utilised in an answer to question 3, as an explanation about how our personal level practices are enabled, then this use of the data would not be problematic. However the data is being taken to provide an answer to questions 1 and 2 because it is thought to dictate the nature of the personal level practices. It is a mistake to think the personal level practices can simply be read off from sub-personal data.

We have seen that sub-personal versions of ST employ representations or copies of psychological states rather than accessing the authentic psychological states of others. If we are employing internal representations of psychological states, even if these are not states of the imagination, we still face the difficulty of how to

360

justifiably transfer these states from one's own mindreading equipment to other minds. And Wittgenstein has suggested that this cannot be done by extending your inner process into others.

However Gallese suggests a way in which internal psychological representations might be transferred to others:

"By means of an isomorphic format we can map others' actions onto our own motor representations, as well as others' emotions and sensations onto our own viscero-motor and somatosensory representations. This is what I mean by embodied simulation".[899]

The idea here is that we exploit similarities between ourselves and others. These similarities include the fact that other people are the sort of things to which psychological states get attributed. In fact the relationship is reversed in the above quotation by Gallese; instead of speaking about extending your inner psychological representations outwards, Gallese speaks of mapping outer psychological manifestations onto inner representations. However the question remains how we can legitimately link psychological representations to other people or relate the behaviour of others to our own internal states. On Gallese's view it does make sense to attribute instances of your own experience to others because they are qualitatively comparable. However this only provides a story of what might causally facilitate the linking of first and third-personal representations. The issue Wittgenstein has raised

[899] Gallese 2009 P.524

is how to justify making such connections, which he suggests will be incoherent. This links to the above issue about what kind of questions sub-personal ST is offering an answer to.

On the face of it Wittgenstein's warning that psychological states defined entirely through inner experience would lack a firm basis for ascribing to others should not be applicable to sub-personal formulations of ST. Other bodies, as they figure in the implicit ST story, are something on which psychological predicates already get a legitimate hold (although it should be borne in mind that these predicates are conceived as subjectless which is highly problematic if we are sympathetic with Wittgenstein's account of how psychological predicates get their meaning). Gallese claims: When we 'directly recognise others' it is as 'selves like us' rather than as bodies we must then satisfy ourselves are 'endowed with a mind'.[900]

Furthermore the ST claim that we have shared representations appears to suggest that the concern about extending private inner processes outwards does not apply here. If psychological representations are understood as shared then psychological states cannot be the sort of thing I only know about from my own case. We saw that, in the case of deliberate imaginative exercises of psychological attribution that Wittgenstein attacks, extending the idea of what is immediately felt into other people's bodies will only provide an idea of our having feelings in their bodies, not of their having feelings. The reason for this was because we were working with a model of the inner which sharply distinguished psychological processes from

[900] Gallese 2007 P.147

362

observable 'biological' ones. If instead we have a model of psychological states as shared then this problem need not arise.

Provided ST is justified in appealing to an empirically establishable similarity in form between oneself and others it could claim to avoid Wittgenstein's worries about the coherence of applying what we know from our own case to others. The case for justification is based on the idea of 1. shared representation, 2. we-centric space, 3. a basic biological/non biological contrast. I will evaluate these three factors in turn.

1.   On closer inspection, the representations provided by our neural machinery would better be described as merely neutral rather than shared. MNs provide a representation of the action or emotion which is not tied to a particular agent. It needs subsequently to be assigned either to the agent or to an observed target. We saw that this task is supposed to be executed by the Who System.[901] This system attributes a psychological state to the self or to a target. Alongside the shared representations afforded by MN activation, we have other non-overlapping representations whose activation only occurs in the specific case of the agent's own psychological states. These serve as a cue for attributing the psychological state to the self or to the other.

---

[901] Georgieff & Jeannerod 1988, Jeannerod & Pacherie 2004

363

It is not clear that 'shared' is really the best description in this context. For a representation to be literally shared, both agent and target must experience the same representation. For Wittgenstein psychological concepts are shared because they belong to shared language and practices but this does not imply we simultaneously undergo identical psychological states. Wittgenstein's discussion is designed to raise questions about how we get this neutral content that could belong to oneself or to another agent. If this content is ultimately derived from our own experiences then there is no straightforward way of extending it to others and claiming another agent is undergoing a psychological episode is comparable to claiming it is 5.o clock on the sun.

Furthermore it is not even clear what kind of content is under discussion. It is unclear how we can coherently attribute a grasp of psychological content to these sub-personal states. Grasping the meaning of psychological concepts is a personal level accomplishment and it remains wholly unclear how this is linked to the occurrences of mirror neurons at the sub-personal level. In the same way the epistemological question only arises at the personal level so does the question how one grasps the meaning of a psychological state. If MNs are helpful to understanding the process by which we grasp meaning then that is useful. However if the claim is that MNs constitute grasping meaning then it is clear they cannot be characterised in this way

2. According to the developmental story offered by Gallese, a grasp of

psychological states begins as undifferentiated with regard to whether an experienced state belongs to the self or to another. Our understanding of these states develops within we-centric space. However this does not ensure a representation can still be called shared once we reach a stage where a decision whether it belongs to self or other has to be taken. Because a psychological representation is not attached to either agent does not mean it is shared by both. Again it would be more accurate to say that it does not yet represent either the perceiving agent or the observed subject. The issue raised by Wittgenstein is whether the process of grasping another person's psychological states involves a grasp of anything outside the agent's own machinery. By employing the idea of representations implicit ST risks locking us inside our own psychological experiences.

It is also debatable whether the notion of an unattached representation that sub-personal versions of ST begin with is really coherent. One thing Wittgenstein tries to bring out is the grammatical connection between the concept of a sensation and that of a subject who feels the sensation.[902] One cannot proceed by imagining a psychological state without a subject who experiences it and then add either that another person or oneself experiences that state. To imagine another in a psychological state it is not necessary to produce a representation of that state. However Implicit ST appears to be offering a sub-personal analogue of just such a process:

---

[902] McGinn 1996 P.172

"The integrity of the sensorimotor system indeed appears critical for the recognition of emotions displayed by others, because the sensorimotor system appears to support the reconstruction of what it would feel like to be in a particular emotion, by means of simulation of the related body state".[903]

3. Sub-personal versions of ST could make a stronger case that they can incorporate the idea that we, or at least our brains respond differently to what is living and what is non-living. MNs are shown to be sensitive to basic goals which they equate with simple intentions of an agent. Although this contrast may not be as powerful as Wittgenstein's contrast between living and non-living it might suffice to show that attributing intentional states to others does not involve an incoherent extension of one's own inner experience. This process of attribution is not locked inside first-personal experience. These formulations of ST do not ignore the fact that there is something unique about living things. However, what is distinctive about other agents cannot be satisfactorily captured because ST does not allow us to perceive other agents as agents in psychological states. The issue here is that either ST allows direct perception of psychological states in which case a simulation process becomes redundant or there is still some hurdle involved in attributing inner states to others in which case simulation is required. Either we have ST in which case we begin by employing 1st personal states or we have direct perception of other agent's psychological states in which case it is not evidence for ST.

---

[903] Gallese 2006 P.19

1st/3rd person asymmetry

Although Wittgenstein does away with the idea of a private inner realm he does not deny there is an asymmetry between first and third-personal psychological ascriptions. As we saw in chapter two, one of the problems with TT is that there is no account of the $1^{st}/3^{rd}$ person asymmetry we find in psychological ascriptions. This avoids problems facing Cartesianism and ST. ST begins with the first-person and then tries to make sense of the idea that what you've accepted in the first-person might also apply to some other body. It shares with Cartesianism the idea that knowledge of psychological states is derived from one's own experience. Although TT avoids this position it fails to account for the distinctive relationship we stand in to our own psychological states and instead puts our knowledge of both as equivocal.

Wittgenstein's philosophy is able to account for this asymmetric relation by capturing what is distinctive about 1st person psychological exclamations. In these cases the person who is, for example, in pain is the one expressing it. The expressions therefore play a constitutive role in the experience itself. As McGinn explains we should not view the fundamental asymmetry of self and other that is reflected in the grammar of psychological concepts as a limitation. Wittgenstein does not deny that there can be uncertainty about another's psychological state. However we should not let our reflections on this psychological truth encourage us

to accept a false philosophical picture of the others metaphysical privacy.[904] The purpose of Wittgenstein's contrast between attitude and opinion is to show that the place of intersubjectivity in our lives is out of range of scepticism. The feasibility of scepticism is made untenable not by individual judgments about another's psychological state but by the existence of the social world at a level more fundamental than that of making judgments. Wittgenstein's insight was that if one looks closely at cases where we might doubt the authenticity of another's psychological expression this doubt is not connected with a strong philosophical doubt but with the complexity of our psychological language games. These incorporate the possibility of masking emotions and pretence as well as sincere expressions of thoughts and feelings.[905]

Wittgenstein and Gallagher

Like Gallagher; Wittgenstein offers an account of how we are able to directly grasp the intentional states of others and Gallagher would be sympathetic with many of the claims he makes. However Wittgenstein's ideas can not only be used to supplement Gallagher's but also to correct them. In particular Wittgenstein can be used to overcome a difficulty with the account of direct perception Gallagher advocates.

There are, of course, many similarities between Wittgenstein's and Gallagher's ideas. Wittgenstein's observation that we do not encounter others at the level of

---

Mcginn 1998 P.49
Ibid P.55

physiological movement is in the same spirit as Gallagher's ideas about smart intersubjective perception. When we detect other peoples intentional states we do not first engage with their bodies in terms of their physiology, we directly detect, for example, patterns of happiness or grief and we also directly detect intentional actions as actions. This is an issue about what level of description our perceptual capacities are responding to when we encounter other agents. It is also a general point that perceptual capacities need not be attuned to the most basic level of physiological description. Gallagher makes this point when he claims we do not see a red sports car as a red mass with a particular shape.[906]

A dilemma was raised in the previous chapter during the comparison of Gallagher with Merleau-Ponty as to whether Gallagher is rejecting the philosophical model in which the intentions of others are in principle problematic, or merely accepting that there is an epistemological issue here but providing a solution to it. For Wittgenstein, as for Merleau-Ponty, epistemological questions do not arise because our relations to others are not primarily epistemological. We may have particular questions about a given individuals psychological state but there is not a general sceptical question about the minds of others which needs answering. For both our mode of awareness of others is not an epistemological one. Wittgenstein and Merleau-Ponty are also close on the issue that psychological ascriptions are anchored in the practices we find ourselves in with others.

In the previous chapter the question was raised whether intersubjectivity is constitutive of psychological states for Gallagher. This chapter drew attention to an ambiguity as to whether a grasp of what it is to have psychological states requires one to be engaged in 2nd person relations with others or whether such interactions are merely a tool for smart perception. In Wittgenstein's writing it is much clearer that a grasp of such interactions is required if we are to grasp what is involved in the attribution of psychological states to ourselves or others.

It is worth emphasising a difference between Wittgenstein and Gallagher. Whereas Gallagher wants to show that perception of others is unproblematic because there is nothing atypical about perception of intentional states compared to perception of physical objects, Wittgenstein suggests there is something unique about the way we perceive other human beings. Differences between the way we perceive psychological manifestations in human beings and the way we perceive sports cars are not explained by Wittgenstein in terms of a distinction between unobservable intentional states and observable mechanical behaviour. As Cockburn explains our attitudes are fundamental in one's relations to others:" I respond to what I see — to the human being — with pity, resentment, gratitude and so on: that is, in ways that mark off my thoughts and feelings about others from my thoughts and feelings about trees or *cars*" (my italics).[907] Although Gallagher's notion of smart perception is a step in the right direction, it still leaves us with a problem. It does not bring out what is unique about intersubjective perception. As Cockburn puts it "a person is an extended, tangible being that exists in the same world as tables, trees and mountains.

[907] Cockburn 2001 P.101

370

But this formulation conceals differences that are every bit as important".[908]

Gallagher's account makes perception of psychological states just another instance of object perception. Because, for Gallagher, perceiving psychological states amounts to detecting patterns there is not the sense of encountering an experiencing subject. However patterns of intentionality are not brute facts in nature in the way objects are. For Wittgenstein, as for Merleau-Ponty, engaging with another psychological subject is very different from engaging with an object. We saw that for Merleau-Ponty an appreciation of others as intentional subjects was a precondition of perceiving them at all. Our experience of subjectivities as subjectivities is immediate. The experiences we have are only coherent in terms of the presence of another subjectivity. In Wittgenstein this difference is centred round the kinds of practices that are exclusive to our interactions with other human beings.

However the starting points for Wittgenstein and Merleau-Ponty are not identical. In the previous chapter we saw that Merleau-Ponty begins with the world from the point of view of the subject. Other subjectivities are presupposed in having a world at all. For the world itself is experienced as something onto which other points of view are possible. For Wittgenstein the step which rules out epistemological questions comes a little later than it does for Merleau-Ponty and it involves a more specific focus on the distinctive character of the other human beings we encounter. For Wittgenstein the intersubjective practices we engage give us our grasp of psychological concepts. The subjectivity of others is presupposed in the nature of the practices. Wittgenstein also lays greater emphasis on the face as the nucleus of our

[908] Ibid P.37

experiences of intersubjectivity. Faces do not simply function as sources of evidence for another's psychological states; they automatically immerse us in intersubjective experiences. For example the suffering on another face can make one a partner in their distress.[909] One's own facial expressions also constitute a response to the suffering.

Our immediate experience of subjectivities as subjectivities marks Wittgenstein off from Gallagher who reduced our grasp of intentional qualities to a species of pattern recognition. Wittgenstein draws our attention to the normative character of the practices in terms of which our psychological concepts are anchored. This normative quality is lacking in Gallagher who treats intentional phenomena causally. Gallagher's position does not properly account for the difference between people with minds and objects whereas Wittgenstein's notion of an attitude does. One's attitude towards another soul is quite different from an attitude towards a chair. Each comes with a battery of associated practices and expectations. For example one might offer assistance to a visibly distressed soul whereas one would not do this for a worn-looking chair. The practices that one would engage in with a minded entity are very different to those one would engage in with an object. These differences are spelt out in terms of the normative requirements other minded entities can exert for us. On this matter, Gallagher is positioned closer to TOM in that he is working with a broadly scientific pool of concepts anchored in the scientific practice of making sense of things whilst offering Direct Perception as an answer to questions about making sense of others.

[909] Ibid P.104

372

Wittgenstein's account then does not see psychological attributions simply as a kind of pattern recognition. As noted earlier he writes "My attitude towards him is an attitude towards a soul. I am not of the *opinion* that he has a soul".[910] One way to understand an attitude would be a disposition to pick up on particular types of pattern and to respond to these patterns in certain ways. Wittgenstein still offers an improvement on Gallagher by emphasising the unique characteristics of these patterns as against those involved in identifying physical objects. However the attitude Wittgenstein claims we have towards others is part of the structure of the intersubjective world. To see people in psychological states such as pain or joy is necessarily to be affected by what we observe. These experiences are part of the fabric constituting experience in general. They are preconditions of being able to take part in practices such as that of making sense of others behaviour. In this respect Wittgenstein is closer to Merleau-Ponty than to Gallagher.

Wittgenstein and the phenomenological tradition

The phenomenologists examined in the previous chapter could be understood as offering an answer to the question of what it is to experience another as a minded entity rather than an object. The chapter drew attention to continuity in the answers given by Merleau-Ponty and Sartre to this question, both of whom highlighted the immediate character of the recognition one has that one is confronted with an

---

[910] Wittgenstein 1968 2 iv P.178

experiencing subject who also has a world. Wittgenstein's notion of an attitude towards a soul might also be understood as addressing the same issue. Like Merleau-Ponty, Wittgenstein does not think that there is an epistemological issue here. Understood in this way Wittgenstein's answer to the question of what it is to recognise something as minded involves an appeal to the range of shared practices associated with psychological experience. The ways we might respond to a person who has hit himself with a hammer are totally different to the responses that might be evoked by witnessing a nail and wood being hit with the same hammer. Typically we would exhibit concern and feel compelled to offer aid in the case of the former whereas this would be absurd in the latter case. Wittgenstein thinks the way to resolve these kinds of questions is by pointing to the practices themselves in order to scrutinise them. We will see that many of them are normative in character. He also thinks we should take care not to go beyond these practices in our search for an answer.

Although this is certainly enough to distinguish his position from Gallagher's, Wittgenstein does not stress in the same way as the phenomenologists the essentially intersubjective character of the world. It is difficult to see how Wittgenstein methodology could allow us to arrive at this sort of insight. For Merleau-Ponty, to be an experiencing subject is to have a world and to experience a world is to experience an intersubjective space. Even before we encounter particular entities we still experience the marks of their presence. A chair can only be a chair because it is something that can be sat on by others as well as oneself. This idea that we

encounter other subjects even through the presence of objects is not found in Wittgenstein's work.

Summary

Wittgenstein's philosophy undermines key ideas in the TOM enterprise. He successfully resists a picture in which physiological data might be used to infer the psychological states of others by showing that adopting this picture leads to incoherence. We saw TT is not vulnerable to the full thrust of Wittgenstein's attack because it understands psychological states as theoretical rather than as products of introspection, so avoids attributing analogies of self experience to others. However it does not escape the problems Wittgenstein raises with treating these states as unobservable inner states only contingently related to behaviour. TT's avoidance of the problems associated with introspectively identified states also comes at the cost of raising a new problem about how the distinctive character of first-personal experience can be captured.

Wittgenstein also raises problems with the approach adopted to explaining our understanding of others psychological states adopted by standard ST because he shows that starting from an exclusively first-personal perspective cannot give us the result we seek. Wittgenstein also shows that the imagination cannot be assigned a pivotal role in an explanation of our ability to attribute psychological states to others.

The issue with sub-personal ST was more complex. Although sub-personal ST does not involve the imagination it is modelled on a process involving the imagination found in standard ST. However the attempt to relocate this process at a sub-personal level is incoherent. Furthermore sub-personal ST does not sit happily in the inner outer model in terms of which the problem of other minds at first arises. One reason is because it does not conceive psychological states as exclusively first-personal. Furthermore sub-personal ST does not claim we confront mere bodies which we must attribute such psychological states to. However features of the inner outer model, which this chapter has been using Wittgenstein to undermine, are still present in sub-personal ST. There is still the problem of how to transfer internal representations to others; be they states of the imagination or not. Wittgenstein's challenge to the coherence of extending one's inner processes to others still stands. ST is not able to get round this worry by introducing the idea of shared representations because they cannot be described as shared in any meaningful sense of the term. Although sub-personal ST can say with Wittgensteinian that what we confront is not mere bodies divorced of any psychological characterisation, it is not able to capture what is distinctive about psychological beings. Any move that allows us to be sensitive to others as agents in psychological states brings us into the remit of direct perception and makes the simulation process redundant.

Wittgenstein's ability to capture what is distinctive about our dealings with other human beings also makes him a useful corrective to Gallagher who fails to account for this and instead plays down the differences between perception of psychological

states and perception of things in general. Wittgenstein's work also emphasises the normative nature of the practices from which psychological concepts derive their meaning and enables us to move away from the causal treatment psychological states receive in both the TOM debate and in Gallagher's work. Wittgenstein's position bears important similarities to Merleau-Ponty's in that for both an appreciation of others as psychological agents precedes any epistemological questions we might raise about particular cases.

7 Conclusion

This PhD thesis examined ostensibly empirically based theories of how we understand other's psychological states; Theory Theory, Simulation Theory and Direct Perception. It raised a number of issues for these theories in relation to the answers they give to questions raised in this thesis. These questions probed the relation these theories stand in to the problem of other minds: which is the problem of how one can be sure that other agents undergo psychological experiences like one's own. It looked at four different ways this question might be cashed out involving a mixture of philosophical and empirical considerations. To recap:

1. What are we doing when we attribute mental states to others?

2. What justifies claims we make about the mental states of others?

3. How is attribution of mental states to others achieved?

4. How we could ever perform such a miraculous feat as to successfully attribute psychological states to others?

It was noted that the so called other minds problem was connected with a picture of psychological states as private inner states. According to this picture psychological states have to be inferred on the basis of observable data. The discussion in the introduction suggested that a theory's approach to these questions would be influenced by preconceptions about the appropriate kind of answer. The dominant picture underlying the philosophical problem of other minds allows no conceptual

link between psychological states and behaviour. This makes any attribution of psychological states to others appeal to something that is in principle unobservable. The introduction also introduced a commonplace though flawed solution to the problem, the argument by analogy, which makes an induction on the basis of the first-person. We resolved to see if the theories under examination could offer a successful alternative to this method, or whether they implicitly rely on it.

Question 1

TT s answer to question 1 (what are we doing when we attribute mental states to others) is that we are trying to causally explain and predict behaviour. Giving this answer involves acceptance of 2 problematic philosophical assumptions:

Firstly the assumption that psychological states are inner. This answer belongs to a picture in which psychological states are not empirically observable and are inferred on the basis of empirical data.

Secondly the assumption that psychological states are theoretical posits which derive their meaning from the myriad causal interrelations articulated in the theory.

This thesis has been challenging both these assumptions.

An issue raised in the introduction was whether the role of one's own experience in

379

making psychological attributions to others is adequately accounted for by any of the theories under examination. TT's answer does not account for an asymmetry in the ways in which the psychological states of others and one's own psychological state are disclosed to one. It therefore fails to account for the distinctive relation in which we stand to our own psychological states.

This answer also suggests there could be an alternative way of describing the behaviour of other agents, (for which we postulate unobservable psychological causes,) without recourse to folk psychological vocabulary but it is hard to make sense of this claim. In abandoning the language of everyday psychological expression it is not clear that we have a theory neutral description of the phenomena to be explained.

Furthermore the idea that in our dealings with others we are always in the business of forming explanatory hypotheses about their behaviour lacks phenomenological plausibility. TT's answer to question 1 is therefore implausible if we take it as a description of a personal level process we engage in when we encounter other agents. A criticism that emerged in the examination of TT was that it takes a disengaged approach to understanding other rather than seeing this understanding as emerging during interpersonal exchanges.

ST's answer to question 1 is that we are using our own psychological capacities to simulate another agent's psychological processes in order to assign the end state of

this process to the agent. On the face of it, ST should be immune to some of the criticisms this thesis has leveled at TT's answer to question 1. In particular it appears opposed to the view of human beings as participants in a disengaged process of figuring out the mental states of others. The chapter on ST evaluated what at first appears to be an improvement over TT's account of how we understand others found in ST; an attempt to portray this understanding as the outcome of a process in which we are in direct contact with the psychological states of others. However it was argued that this attempt to construe our contact with psychological states as direct is undermined by the use ST makes of offline representations of others to drive the process of psychological attribution rather than their actual current psychological states.

In fact it is unsurprising that ST is unsuccessful in its attempt to portray an agent as in direct contact with other agents' psychological states when one considers that classical ST comes much nearer than TT to constituting a modern reformulation to the argument from analogy. It is suggesting we make attributions to others on the basis of our own experience, which is employed as a model of the experiences of others. This in itself implies a disengaged stance towards other agents. The very idea of mind-reading central to ST implies a picture of mental states of others as unobservable and in need of inference.

In the case of personal level simulation we are mentally placing ourselves in the other person's situation to see how we would respond while making allowances for

relevant differences between ourselves and the other agent. However this answer means that this formulation of ST faces the same problem facing the analogical strategy for knowing other minds: if meaning is of necessity derived from first-personal experience then how can one coherently extend it to others? (This leads to the problem expressed by question 4). Sub-personal formulations of ST portray our initial encounters with psychological representations as encounters with unattached representations so does not face this particular difficulty although this answer looks even less plausible as it suggests we are confronted with unattached representations and have to decide if they are our own or other peoples.

ST's answer to question 1 offered another apparent improvement over TT's answer. Whereas a problem raised with TT was that it carried the implausible implication that our psychological concepts are replaceable, ST on the face of it seeks to offer us a vindication of everyday psychological concepts. Gordon, for example, is motivated by the phenomenological implausibility of supposing we make inferences based on theoretical premises. However, anchoring our psychological concepts both in first-person experience and in unattributed inner representations is problematic.

ST seeks an explanation that can account for the immediacy with which we are able to identify psychological states. This is a worthy aim. However if ST is understood as a personallevel activity involving hypothesis formulation then it is just as implausible from a phenomenological perspective as an answer to question 1 as the answer given by TT. We still employ an inferential theory albeit a simplified one.

Heal has a partial defense to this answer. She could argue that because the focus of the replicator/simulator is not on the agent but on the surrounding world[911], one is simply attempting to experience what the world must be like from the subject's perspective. Psychological explanation is not central for replication.[912]

The implausibility of ST as a prevalent personal level methodology also prompts a retreat to the sub-personal to avoid accusations of phenomenological implausibility. However this thesis has been arguing that it is inappropriate to interpret sub-personal processing as providing an answer to an epistemological problem about how we know other minds (see discussion of question 2).

This thesis also argued that the sub-personal data is ill-fitting in the classical simulation process for which it was originally appealed to as empirical support. It is worth stressing that the sub-personal versions of ST this thesis examined are not, in fact, analogues of personal level ST at all. They offer a different answer to question 1; they describe a separate process in which we have to assign unattached psychological representations to other agents. The problem of disentangling personal experience and representations of other agent's experience which is supposed to arise because both types of experience occur at the same neurological location is a particularly salient example of what happens if one attempts to make sub-personal processes do personal level work.

---

[911] Heal (1986) P.48
[912] Ibid P.52

Furthermore the modified simulation process in which we confront unassigned psychological representations, some of which are subsequently assigned to other agents also lacks phenomenological plausibility if understood as an answer to question 1. The idea of having to decide whether to assign an unattached representation to oneself or another agent is not able to account for the sophisticated role that context plays in revealing these states to us.

TT and ST share a problematic movement back and forth between personal and sub-personal levels of explanation. An aim of this thesis was to draw attention to a tendency exhibited by both theories to treat these levels of explanation as isomorphic and to show why this failure to treat the distinction more seriously is problematic. In fact this tendency is more pronounced in ST. In Gordon's early formulation for example, personal level simulation is conjectured to represent the tip of the iceberg implying that there could be a transition between personal level simulations and automatized sub-personal processes. This assumption has been treated as unproblematic in subsequent reformulations of the theory. One reason that this is a problem is because it leads to confusion about what level of description the theories are trying to account for, and what concepts can coherently be made use of at each level. ST construes MNs as involved in answering an epistemological problem rather than giving an account of how our understanding of others might be facilitated at the sub-personal level, running together questions 2 and 3 above. MNs are being used to state what the practice of understanding others involves rather than to show how the practice is supported at the sub-personal level. However sub-personal processes

cannot capture phenomenological level interpersonal negotiations.

The discussion in chapter 3 of the role ST assigns to the who-system also brings the problematic slippage between personal and sub-personal levels of explanation in this theory into relief. ST characterizes neurological data as solving a personal level epistemological problem rather than answering a question about how our understanding of others could be facilitated by sub-personal processing (question 3).

The issue of whether the role of one's own experience in making psychological attributions to others is adequately accounted for by TT also arises for ST. It is true that ST does acknowledge an asymmetrical relationship to one's own psychological states. However it still does not have a plausible account of what this relationship consists in. An objection raised against ST in this thesis is that it treats our understanding of our own mental states as given. ST also glosses over the differences between the way we experience our own psychological states and the way we experience those of others treating them as analogous which is implausible.

Hybrid positions of both TT and ST complicate the answers these theories must give to question one. They claim that what we are doing is utilizing some combination of theorizing and simulation in order to attribute hypothetical states to others. While seeking explanations and predictions about others behaviour, either we are relying on a causal theory that appeals to a body of theoretical folk psychological concepts while employing a simulation process as a heuristic shortcut, or we are using our

own psychological capacities to simulate another agent's psychological processes and supplementing this process with theoretical knowledge.

Both these answers come at a price. For one thing the original aim of the TOM debate was to provide mutually opposed theories that exhaust the possible explanations of what we are doing and this aim must now be abandoned. For TT allocating any substantial role to simulation will undermine a conceptual picture motivating TT in which theorizing about the psychological states of other is just another instance of an all-encompassing causal explanatory strategy. Furthermore incorporating simulation involves a dilution of the original conception of what constitutes a theory, which was originally understood as an interwoven body of knowledge. This concession makes it a lot less clear what we are to make of claims that we are using a theory to understand others. Hybrid ST also loses a key virtue of the parent theory, which was its simplicity. It claimed TT's answer to question 1 was implausible in supposing we had to make use of complex theoretical information.

Whereas a criticism of the answer TT gives to question 1 made in this thesis was that it presupposes an implausible notion of what psychological states consist in (psychological states which have their meaning fixed by their functional role in a theory), hybrid TT's answer waters down this conception of psychological states allowing that functional status alone may not be capable of doing all the work. However this functional status is still central to what psychological states are.

Hybrid ST's answer to question 1 is not vulnerable to the criticism made of ST's answer that it is anchoring meaning exclusively in personal experience, thereby making an answer to question 4 difficult. Neither of these hybrid positions answers to question 1 suffers from the phenomenological implausibility attached to personal level formulations of the parent theories as they allow that the work is done at a sub-personal level.

However more general problems still apply to the answers of both hybrid positions. The main problem for both their answers is the same problem facing the parent theories: the mind is still understood as inner and unobservable. Another problem is that there is still an incoherent movement between personal and sub-personal levels of explanation with sub-personal states evoked to provide an account of what personal level interactions consist in.

Question 2

In examining TT's answer to question 1 this thesis exposed philosophical assumptions underlying what purported to be an empirically based answer. There is an assumption in some versions of TT that the development of folk psychological theories is continuous with the development of science as a whole. Unsurprisingly given this assumption TT's answer to question 2 (what justifies our attribution of psychological states to others?) is that psychological attributions are justified by their success in enabling explanation and prediction of other's behaviour. In other

words they are justified in just the manner all scientific theories are justified.

A dissatisfaction raised with this answer to question 2 is that it has the implication that we should accept that folk psychological language is in principle replaceable. However it is implausible to think of psychological states as hypothetical entities that could be replaced with a better theory. Attributions of psychological states to other's, according to TT's answer to question 2, are only justified to the extent they are successful. This means we would not be justified in continuing to attribute these states to others if they failed to prove explanatorily useful. However it is not clear-cut that we would give them up if this turned out to be the case. It is not even clear what could count as a competing explanation. To revert to the explanation of observable behaviour without appeal to psychological states, or to referencing neurological processing, would just be to abandon any attempt at explanation.

An aim of the thesis was to unravel the differences in the way TT and ST answered the questions about knowing the psychological states of others raised in this thesis; in particular in respect to the empirical and philosophical aspects of their answers. This thesis aimed to assess the extent to which they are philosophical rather than empirical. Empirical data for TT is often concerned with the acquisition of theoretical capacities and includes false belief tests and studies of autism. However there are also philosophical assumptions motivating this research. It was found that a philosophical assumption motivating TT is that our strategy for understanding of others psychological states will be continuous with our understanding of the world

388

in general. There is also an assumption that states derive meaning from theoretical relations.

ST's answer to question 2 is that attributions of psychological states to others are justified by similarity in the psychological constitutions of other agents to our own psychological constitution. Again this similarity to the classic argument from analogy raises the same problem in that we are basing justifications on inferences from a single case. It was noted that justification is not a big concern for ST.

ST is formulated as a response to TT. It considers this response to be motivated by empirical evidence. Some of the evidence standard ST relies on is the same data utilized by TT. However they interpret the evidence in a different way. The empirical data sub-personal formulations claim to be motivated by revolves around the recent discovery of MNs. It is claimed that MNs illustrate that shared mechanisms are operating in first- and third-personal experiences which is something ST would predict but TT would not. However ST is also motivated by theoretical considerations albeit empirically inclined ones such as an expectation of evolutionary continuity and a preference for simplicity in explanation. A problem is that the sub-personal data is appealed to in order to provide a justification for the answer ST gives to question 2. This involves treating questions 2 and 3 as equivocal and offering sub-personal processes as epistemological justifications for the validity of personal level descriptions of what we are doing rather than using this data to explain how these processes might be facilitated at a sub-personal level.

Question 3

Whereas the previous questions are philosophical question 3 (how the attribution of psychological states to others is achieved) is asking for an empirical explanation relating to the particular answer given to question 1. TT answers question 3 by offering details regarding the acquisition of our folk psychological theory. However there is disagreement amongst the various factions of TT about what answer to give to this question. Answers vary as to whether the capacity is acquired during development, explicitly learned or genetically bestowed. There is also disagreement as to whether to locate the various answers at the personal or sub-personal level. A motive for locating them at a sub-personal level is to avoid accusations of phenomenological implausibility which can be levelled at their answer to question 1. We do not experience ourselves as testing hypotheses about the causal role played by mental states in our observations of others. Therefore the claim that TT describes a ubiquitous personal level process lacks plausibility.

A key point to stress is that TT's answers to question 3 are empirical explanations of their answer to question 1. If the answer TT gives to question 1 is inaccurate then the various answers to question 3 cease to have any meaning as one is left with no theoretical process to empirically instantiate. If we are not preoccupied with a process of explaining and predicting unobservable mental states then we do not need an account of how this is achieved. Although the answers to question 3 are empirical they are mired in philosophical assumptions. One problematic assumption is that it

390

is coherent to flit between personal and sub-personal levels of explanation. The philosophical assumption is that low level processes such as neurological processes are somehow continuous with high level processes including conscious experience. Lower level processes are characterised as already quasi-intentional.

ST also takes question 3 as an empirical question about how the answers it gives to question one are made possible. Attribution of psychological states to others can be achieved in one of two ways. It can be achieved through a personal level imaginative capacity for simulating others and tweaking the simulation to allow for personal differences. Alternatively it can be achieved through a sub-personal analogue of this capacity automatically triggered by observation of others whose observed psychological states are 'matched' with one's own psychological states, and subsequently completed offline by the neural networks which underlie first-personal psychological experience. As with TT's answer to question 3, a problem here is that this answer is only coherent if the correct answer has been given to question 1. If ST is offering an incorrect answer to question 1 it makes little sense to seek an empirical substructure for this explanation.

Another problem is that ST is not really able to offer us an empirically based substructure in any case. The mechanisms described by ST's proponents cannot plausibly be interpreted as instantiating a standard simulation routine in which first-personal capacities are being exploited to simulate another's psychological processes. Indeed the appeal made by proponents of ST to the notion of shared

391

representations would more accurately be understood as undermining the idea that people possess any dedicated first-personal psychological capacities. Indeed it is far from obvious that the alleged neurological underpinnings presented as evidence for ST really behave like a simulation process. It is difficult to read into MN activity a simulation process. Furthermore the evidence itself does not motivate any interpretation but is open to competing interpretations one of which, DP, we explored in the second half of this thesis.

Both psychological studies, including studies of autism, and neurological data have been cited in support of the respective theories. However none of this evidence has been decisive regarding one TOM position over another or regarding either TOM position over the alternative positions discussed in this thesis. This thesis has argued that discoveries about what is happening at the empirical level do not dictate the answer to the philosophical question about how we should describe what is happening at the personal level. A question raised through evaluating these answers to question 3 is what kind of support they can be expected to provide for the respective answers to question 1. The issue is whether empirical discoveries can influence or constrain the personal level descriptions of what we are doing when we ascribe psychological states to other agents.

While the relationship between personal and sub-personal levels is not as straightforward at the answers given by TT and ST would suggest, this does not mean we should conclude that there is no meaningful relationship between these

levels or that empirical discoveries are powerless to influence our conception of the personal level. It only means that these levels correspond to one another in more complex ways.

Question 4

TT does not owe us an answer to question 4 (how on earth could we successfully attribute psychological states to others when these are something we only know about from our own case?) as it does not presuppose a miraculous picture in which psychological states are attributed to others based on introspectively derived knowledge of first-personal psychological states. However TT avoids this problem at a high cost. It is left with no way of accounting for the distinctive character of people's relation to their own psychological states.

Question 4 is a very salient question raised by standard ST which does not have a satisfactory answer to it. The meaning of psychological states is tied to a first-personal perspective, which makes it seemingly inexplicable that they could be attributed to other agents. Sub-personal formulations of ST are able to escape raising question 4 because they make psychological representations third-personal. However they pay a very high price for this move as we end up with a very incoherent picture of what is going on. Furthermore, like TT, these accounts are unable to capture what is distinctive about first-personal psychological experience.

Both TT and ST also share a key philosophical assumption that the task we are involved in is attempting to 'mind read' unobservable inner states of other agents on the basis of empirical data. TT and ST are presented as competing hypotheses about what enables us to attribute mental states to other agents, which are to be answered through empirical investigation. However this task of unravelling the philosophical and empirical aspects has shown that things are in fact less straightforward.

The answers given by hybrid theories to the questions raised in this thesis also cast doubt over the empirical nature of the debate. Chapter 4 examined cases where the boundaries between TT and ST become blurred as well as deliberate hybrid formulations of TT and ST. Although various attempts have been made to keep TT and ST mutually exclusive none of them have been universally adopted. As the theories are no longer presented as being in direct competition the question which theory receives most empirical support is eschewed in favour of the question which process, theorizing or simulation, plays the most fundamental role in our practice of attributing mental states to others. This is as much a conceptual issue as it is an empirical one and the answer will be influenced by one's theoretical starting point. This discussion of modifications to the central theories advanced by hybrid positions served to focus attention on ambiguities about what the central claims of each theory are, rather than bringing us any closer to an empirical resolution of the debate.

Another aim of the thesis was to evaluate the extent to which TT and ST are related to the strategy exemplified in the classical argument from analogy which gives rise

to question 4, and how dependant they are on the picture motivating this strategy. TT is methodologically distinct from the classical argument from analogy. However it still shares a significant feature of the picture motivating it which is a conception of the mind as inner. The psychological states of others are still treated as unobservable. The fragmented nature of the various versions of ST is brought out through consideration of their approach to the problem of other minds. Classical ST is a method of *overcoming* this problem through the use of one's own psychological apparatus. It therefore has a lot in common with the traditional argument from analogy. It takes access to first-personal states as given and the states of others as in need of inference. It also attempts to employ this privileged access to first-personal states to infer the psychological states of others. It shares the pitfall of assuming similarity between a simulator and other agent based on a single case.

However more recent sub-personal formulations of ST attempt to distinguish themselves from this process and to show that the process is in fact redundant because we are directly responding to intentional qualities of others. Therefore there is no need for an inferential process like the one employed in the argument from analogy. Advocates of this position are motivated by recent neurological data and attempt to portray the simulation process as *a dissolution* of the problem of other minds through closing an epistemic gap between self and others. However ST cannot have its cake and eat it. If we accept that no inference is required than any simulation process itself becomes redundant. If one accepts that we are in direct contact with the psychological states of others than we must reject the picture of

minds as inner that gives rise to the problem of other minds which ST is an attempt to answer.

Furthermore attempts by recent sub-personal formulations of ST to distinguish themselves from the classical ST's methodology are at best only partially successful. At a deeper level the different versions of ST are still united insofar as more recent formulations are still involved in tackling an epistemological problem, which is now reformulated as the problem of how one separates oneself from others. Like TT all versions of ST are seduced by a picture of the mind as inner so have to treat the psychological states of others as unobservable. Furthermore, although in the grip of this picture neither theory takes question 4 sufficiently seriously instead treating it as an empirical puzzle. TT and ST are both guilty of assuming the coherence of attributing psychological states to others although this picture does not entitle them to do so. They then simply ask how this is achieved.

Summary of first part of thesis

A number of features emerge in the discussion of the theory of mind debate that suggest we must now look elsewhere for more satisfactory answers to the questions raised in this thesis. Both TT and ST provide implausible answers to question 1 if taken as descriptions of personal level processes of attributing psychological states to others. They both claim we are mind reading in order to attribute inner psychological states to others in order to produce explanations and predictions

about their behaviour. We therefore require a more phenomenologically plausible account of what we are doing when making psychological attributions to others than engaging in a process of causal explanation and prediction. A reason they can make this mistake is because both positions are seduced by a philosophical picture of mental states as unobservable. The way to avoid getting seduced by this picture is through closer phenomenological investigation of intersubjective encounters with other agents. This process is illustrated through examination of direct perception, and in more detail through the work of the phenomenologists and Wittgenstein.

An upshot of this discussion of the TOM debate is that the role of one's own experience in making psychological attributions to others is not adequately accounted for by TT, ST nor by their hybrid offspring. None of these answers offer a satisfying account for the 1st 3rd person asymmetry. We are unable to account for the distinctive relation in which we stand to our own psychological states. While TT cannot allow there is anything distinctive about the relation we stand toward first-personal states, ST exclusively anchors the meaning of psychological terms in first-personal states. Some formulations of ST introduced the idea of shared representation but on closer examination this was revealed to be too crude. This justifies our turning to other approaches to answering question 1 to inquire if they can offer a more satisfying account of this role.

Another feature of the discussion that prompts us to look further afield is the use the theories make of the evidence. The use TT and particularly ST attempt to make of

sub-personal data open the way to consider alternative interpretations of the role of mirror neurons and other sub-personal data in later chapters. A disanalogy between the supposed evidence for ST and the classical simulation process was discovered which invites consideration of a prominent alternative to the theory, DP, which contains within it a reinterpretation of the activity of MNs. Furthermore, even without this disanalogy there are good reasons for rejecting ST. Once we reject ST's stipulation of the practice we must find a more plausible interpretation of the empirical data. A dissatisfaction was also raised with the type of empirical evidence which is utilised in the debate. In particular it was noted that no context was built into the empirical setting. This sets the scene for a discussion of the importance of context in allowing us to accurately attribute psychological states in later chapters. The next task the thesis set itself was to see whether Gallagher's direct perception position was akin to TT and ST or whether it was an attempt to pull the rug from under these positions. The discussion of Gallagher also serves as a bridge between the destructive critique of the TOM debate and the introduction of a philosophically robust response to questions about knowing others psychological states that utilizes insights drawn from the work of the phenomenologists and Wittgenstein; one that is not dependent on a picture of mental states as private.

Part 2

The second part of this thesis begins by looking at another theory proposed as an alternative to TT and ST: Gallagher's theory of DP. It begins by examining criticisms Gallagher makes of ST which echo criticisms made in this thesis including the criticism that first-personal attribution is simply taken for granted by ST.

It was noted that DP is in a better position to account for some of the virtues ST attempts to accommodate including the intuition that our perception of other agent's psychological states is direct rather than being mediated through theorizing or simulation. DP offers an account of what we are doing when attributing psychological states to others which has more phenomenological plausibility. Unlike ST, DP can account for the immediacy with which the psychological states of others are presented to us. Prior to discussion of DP this thesis suggested that it is a mistake to take the problem of other minds seriously because there is good reason to think it is a pseudo-problem. It also evaluated ST's attempt to distance itself from the strategy of offering a solution to the problem of other minds and instead offer a dissolution of the problem. The debate about whether ST is involved in a solution or dissolution of the other minds problem is picked up again in Gallagher's argument that sub-personal theories are running a version of the argument from analogy.

This thesis noted a number of improvements DP offered over TOM positions. One positive outcome of this discussion is that context is given a prominent role in

enabling us to pick out the psychological states of others. Gallagher introduces the term 'the observer position' to describe the detached stance TT and ST take in their quest to disambiguate the psychological states of others and Direct perception also echoes and develops the criticisms this thesis has made of this stance. It emphsises the interactive character of intersubjective perception. Another theme of the previous chapters that is picked up again is the way the neurological data supporting mind-reading theories does not determine its own interpretation. We saw that Gallagher is able to appeal to the same data in support of his alternative position.

However Gallagher shares with TT and ST positions the mistake of allowing - personal data to dictate what is happening at the personal level. Nonetheless Gallagher offers a more phenomenologically plausible account of what we are doing when we attribute psychological states to others, in which we are responding directly to psychologically coloured aspects of behaviour within a rich psychological context facilitated by our own interactions with other agents. For this reason the use Gallagher makes of sub-personal data is more innocuous. Data is not being evaluated independently of a context. This thesis conceded that Gallagher's interpretations of MNs as facilitating a direct perception of other agents' psychological states enables the empirical data to fit as least as well in his account as it does in any ST account.

This thesis agreed with Gallagher that perception of the psychological states of others is direct. However other issues raised for ST in this thesis also still arise for

DP. Again elements of both a solution and dissolution of the other minds problem are found in DP. Insofar as DP is presented as a third alternative to TT and ST it is an attempt to answer an epistemological question about how we can claim knowledge of the psychological states of other agents. Social perception is sophisticated enough to provide resources needed to answer the epistemological question and enable us to generate explanations and predictions concerning other agents' behaviour. However this way of presenting DP is in tension with this theory's ambition to distinguish itself from TT and ST by showing that there is no special problem about perceiving psychological states, instead treating social perception as equivalent to perception in general. It is also in tension with DP's professed opposition to mind reading and to the picture of minds as hidden that prompts the mindreading strategy. Attempting a solution to an epistemological problem also sits uneasily alongside the developmental story Gallagher wants to give in which we are interacting with others before we are plausibly capable of making epistemological deductions. Gallagher wishes to claim the capacities people possess for interaction with others cannot be framed as capacities of an individual. He also claims these capacities do not allow the conception of agents as enclosed individuals inherent in the analogical strategy.

A frustration was raised with the way Gallagher's exposition of DP is illustrated through perception of objects rather than perception of the psychological states of others. An objection to this methodology is that this leads to the two phenomena being treated analogously. A criticism that was made of the previous theories was

401

that both were unable to capture what is distinctive about psychological states. Capturing the distinctive character of psychological phenomena will require closer scrutiny of the practices in which such states are found.

The discussion of DP served as a bridge to the work of Merleau-Ponty and Wittgenstein. Although I suggest that ultimately DP shares with TOM the pitfall of proposing a solution to the problem of other minds rather than undermining the problem, DP does offer substantial improvements over TOM positions and these are picked up again in the subsequent discussions of phenomenology and Wittgenstein. Although DP identifies flaws in the epistemological enterprise it does not successfully avoid raising an epistemological question itself. Instead it treats this question as unproblematic because sufficient resources are available for answering this question. A second reason why the question is seen as unproblematic is because justifying the claim perception is smart enough to furnish an understanding of its subject matter without recourse to additional inference mechanisms is a problem that applies to all instances of perception rather than being unique to the psychological.

DP's answer to question 1 is ambiguous. On the one hand it is claiming that what we are doing is directly reading another's psychological states. DP's answer to question 2 is then that we are justified by features of the situation which enable us to produce this reading. On the other hand DP is saying that we are simply interacting with others whose psychological states are presented to us in virtue of the rich social context in which they appear. On this reading question 2 does not arise because we

are not engaging in an epistemological enterprise. DP's answer to question 3 is that this direct recognition is facilitated by a capacity for smart perception underpinned by our sensory motor skills. DP does not raise question 4 as it is not claiming knowledge of psychological states is derived from introspection.

This thesis turns to the work of Merleau-Ponty to challenge the importance of the epistemological framework the previous theories are working with. This thesis also found that some of the insights Gallagher draws from Merleau-Ponty concerning the directness of our perception of psychological states and the bodily nature of interaction do not sit well in the empirical framework Gallagher sometimes appears to be using. This provides another motivation to explore Merleau-Ponty's philosophical position; which can accommodate these insights where ST and DP have been unsuccessful.

Although the scientific framework Merleau-Ponty is attacking is not the same as the one TOM operates in, his work still carries implications for TOM. There are still residual features of this model operative in these theories. Merleau-Ponty's critique of the classical model was examined in order to clarify problems with the framework in which all three of the previous theories are operating. Merleau-Ponty bars us from assigning a justificatory role to physiological data. There is no bottom-up process in which we infer psychological facts from physiological ones. He also undermines the subject object dichotomy that facilitates the problem of other minds. In the same vein he opposes conceptions of conscious experience as uniquely available to an

individual and inferred in others on the basis of outward behaviour. An important insight of Merleau-Ponty is that if this was our starting point, the idea of inferring psychological states known through inner experience on the basis of bodily features could not even get a grip. The analogies could not get off the ground. This framework deprives us of a common medium of representation and thus no basis on which to draw analogies. Merleau-Ponty would oppose the intellectualised character of inferential accounts of how we attribute psychological states to others such as found in personal level TT because for him it is the body which allows us to perceive other psychological entities and this understanding is pre-intellectual.

Merleau-Ponty's work can be used to undermine ST by resisting the inclination to assign an epistemic privilege to the first-person and avoiding a $1^{st}/3^{rd}$-person dichotomy. This avoids raising an other minds problem which ST is an attempt to solve. We do not inhabit a word of psyches which are exclusively available to an individual. Rather we inhabit a world of psychologically expressive bodies. It is the body itself which is the source of expression. The notion of a body object confronting the arguer from analogy is a later impoverishment of this fundamental understanding of the human body. This impoverished notion blocks the insight that it is the expressive characteristics of human bodies that we first encounter rather than movements in objective space. We encounter smiling faces rather than curved lips. The bodies of others are perceived as akin to one's own body. Expressive gestures are used by Merleau-Ponty to illustrate the fact that we directly perceive psychological expressions. These are themselves argued to be the vehicles of

meaning rather then signifiers of intentions lying behind them. Furthermore it is intrinsic to such gestures that they unfold in a shared world.

Merleau-Ponty's intertwining of self and other prevent one from assigning an epistemic privilege to the first-person. Awareness of self emerges in the context of a shared world. Our grasp of psychological phenomena precedes the development of an ability to distinguish self and other. There is no period of development where perception is neutral as to the existence of other psychological agents. Merleau-Ponty thinks we never fully reach an absolute conceptual separation of oneself from others which is the starting point for inferential approaches. The relations one stand in to others feature so prominently in experience that they are best understood as part of the structure of experience rather than mere items of experience analogous to objects.

Merleau-Ponty's own account of our direct perception of the psychological states of others can be used as a corrective and an embellishment of Gallagher's position. Both thinkers think psychological meaning emerges inside social interaction and assign a key role to expressive behaviour in providing this information. One way Merleau-Ponty's ideas can be used to sophisticate Gallagher 's position is through the clear distinction he draws between people and objects. Because Gallagher treated perception of psychological states as analogous to perception of objects he does not capture the fundamental way in which other psychological beings enter our experience. We begin in meaningful relationships with other psyches. It is the

405

actions of other agents and their orientation towards their environment that help shape our perceptions of them. This is quite different from seeing mere objects or even having smart perceptions of objects. Other people form part of the structure of experience rather than appearing to us as items (or perceptual *objects*) in our experience about which we might speculate as to their mindedness.

A second way in which Merleau-Ponty's ideas offers an improvement on Gallagher's is than in his position there is no ambiguity as to whether he provides a dissolution or solution to the problem of other minds. The epistemological question simply cannot arise. We find ourselves in myriad relations with other subjectivities. Psyches are relations to the world and ways of navigating this intersubjective space. Other agents are part of the structure of experience rather than later perceptual discoveries about which we might speculate as to their mindedness. Conducts differ from private psychological episodes in that there is no incoherence in supposing they might belong to other agents as well as to oneself. Whereas smart perception implies sensitivity to intentional patterns, Merleau-Ponty is working within a tradition in which another agent is experienced as a subjectivity.

This is related to the way in which one understands oneself which is less individualistic than the framework supporting the previously examined theories would suppose. Understanding of one's own embodied existence includes a realisation that one is embedded in a world shared by other embodied entities. For all the phenomenologists discussed in this thesis, self-understanding already

presupposes an awareness of others as psychological agents. For Merleau-Ponty awareness of self and others are initially merged and the process of achieving a conceptual separation occurs later. Furthermore psychological states are not first conceptualised as exclusively first-personal items belonging to an individual consciousness. Therefore there could never be a requirement to bestow a person-object with subjective states via an inferential strategy.

Merleau-Ponty's writing makes question 1 problematic. An awareness of others qua psychological entities is a precondition of being able to ask questions about others or about anything else. We are not operating within an epistemological framework in which we are trying to work out whether others are the subject of psychological experiences. We are orientated towards others primarily as other psyches, not as material bodies. For this reason question 2 does not arise. It is not appropriate to seek a justification because an appreciation of other psychological agents is a precondition of experience itself. With regard to question 3 Merleau-Ponty makes clear that we cannot read psychological descriptions from physiological facts. The physiological facts leave a range of possibilities open at the psychological level. Nonetheless he seems to suggest that the 'natural' and the 'cultural' are intertwined and cannot be regarded as floating entirely free of each other. Question 4 does not arise because it only arises on a particular answer to question 1 in which psychological states are grasped on the basis of introspection.

Wittgenstein's ideas are also used to question the validity of the way in which the

TOM debate is set up. His explication and critique of the inner/outer distinction helps clarify the mistaken framework in which TOM is operating. It begins with a truncated idea of unobservable psychological states and tries to infer their presence in others on the basis of an impoverished conception of the physical. Although versions of both TT and ST have attempted to sophisticate the conception of human bodies they are working with in order to allow that they are the kind of objects on which psychological predicates can get a hold, the framework in which the psychological is ultimately inner makes these sophistications unconvincing.

Wittgenstein's ideas have the least to offer in the way of ammunition against TT because TT does not subscribe to the flawed methodology Wittgenstein is attacking in which one extends one's own inner psychological experiences to bodies which are understood in purely physiological terms. However TT is unable to encompass Wittgenstein's insights into the distinctive character of the asymmetrical relationship between first- and third-personal psychological states. Wittgenstein captures the distinctive relation people stand in to first-personal psychological declarations where a claim to be in a particular state is partly constitutive of being in that state.

The application of Wittgenstein's insights to standard ST is more straightforward. Wittgenstein's argument concerning the inadequacy of using one's own experience as a model for the experience of others provides a strong argument against classical ST. This theory is guilty of attempting to solve an epistemological problem whilst beginning with a first-personal perspective. It is therefore vulnerable to

408

Wittgenstein's criticism that all this can achieve is to attribute copies of our own psychological states to others. Wittgenstein also undermines the idea that we could begin with private sensations by showing that the idea of a private sensation is itself incoherent. This strategy will also involve treating one's psychological experiences as items of knowledge that are suited to playing a role in inductive reasoning. Furthermore the job of transferring psychological states to others is assigned to the imagination and Wittgenstein has shown that the imagination will not be capable of fulfilling this role.

Wittgenstein's arguments are also applicable to sub-personal versions of ST. The theory is still offering an epistemological answer to question 1. Although not vulnerable to Wittgenstein's criticism of assigning an unrealistic role to the imagination, sub-personal ST still postulates a representative role for sub-personal states analogous to the role personal level states play in standard ST, which is incongruous at a sub-personal level. The problem concerns treating them as transferrable to others. In any case the whole enterprise cannot even get off the ground if we are interpreting sub-personal activity within a ST framework. If psychological states are inner states known introspectively it is not coherent to assign them to other agents.

This thesis next evaluated potential counter-arguments based on the premise that sub-personal formulations of ST are not in fact buying in to the flawed inner/outer picture. This thesis responded by drawing attention to a feature of the inner that is

still operative in sub-personal formulations of ST. It employs internally produced copies or representations of other agent's psychological states rather than accessing the genuine states. Nothing outside the agent's own psychological machinery seems to be playing a role.

Another line of defence open to sub-personal versions of ST is to accept that they are working with internal representations but resist Wittgenstein's argument that internal representations cannot justifiably be transferred to others. This is based on an appeal to empirically established similarities between oneself and others. However they fail to provide a satisfactory account of what *justifies* the attribution of psychological states to others .Two potential sources of justification were rejected outright. Purported evidence for shared representations is more accurately characterised as evidence for neutral representations, a notion which is itself philosophically problematic. An appeal to we-centric space also fails to ensure that 'shared' representations get assigned to others. A third potential source of justification concerning a fundamental contrast between the biological and non-biological was partially successful. It was conceded that this portrayal of our brains as selectively responsive to living bodies may mean it is not incoherent to extend one's inner representations to others. Ultimately though ST is unable to capture Wittgenstein's distinction between our practices involving what is living and non-living because it does not enable us to perceive agents as agents in psychological states.

Wittgenstein's arguments complemented and enhanced Gallagher's arguments against TOM. Both seek to demystify our ability to perceive psychological states. Both claim we are able to directly perceive psychological states. Both deny that in perceiving psychological states we are first confronted by physiological expressions or actions from which psychological states need to be inferred.

Wittgenstein is also employed as a corrective to Gallagher's Direct Perception account. In particular he provides useful insights into differences between the perception of psychological states and perception of objects, which have been glossed over in Gallagher's account. Whereas Gallagher portrays perception of psychological states as just another case of perception in general, Wittgenstein stresses important differences between perception of objects and perception of psychological states. His treatment therefore allows us to capture the distinctive character of these states. Wittgenstein draws attention to the distinctive character of our practices involving others. His notion of an attitude allows him to bring out the distinctive body of expectations and practices that are embedded in our psychological interactions. He also draws attention to the fact it is the normative features of these practices which we pick out when identifying psychological states rather than physical qualities. This is in contrast to Gallagher who treats perception of psychological states as a causal process. In Wittgenstein's work it is clear participation in 2nd person interactions plays a role in constituting our ability to grasp psychological states whereas it is ambiguous in Gallagher's work whether this is the case or whether such interactions are merely an aid for smart perception.

Wittgenstein's philosophy also leaves no room for the ambiguity that surfaced in Gallagher as to whether we require a dissolution of the epistemological model giving rise to the problem of other minds or merely offering a solution to the problem. For Wittgenstein our fundamental relations to other's are not epistemological ones.

This thesis also drew attention to significant similarities between Wittgenstein's and Merleau-Ponty's philosophies. Neither thinks our primary way of relating to other agents is epistemological. Wittgenstein also seeks to show that we have an understanding of others psychological states that is prior to the possibility of epistemological doubt. Both also view psychological ascriptions as emerging from and anchored in intersubjective practices. The ideas of both can also be used to clarify what is distinctive about psychological states. For both an experience of a subject is very different from an experience of an object.

Wittgenstein's work has similar aspirations to Merleau-Ponty's insofar as he offers further argument why any strategy for making psychological attributions which is embedded within a picture of psychological states as inner will fail. Whereas an implication of Merleau-Ponty's writing was that these strategies are working within a picture which can allow no common ground between one's own states and the states of others on which psychological attribution could be predicated, an implication of Wittgenstein's work is that if we begin with a notion of psychological states as inner then there is no way these states could not be given any external application. As with Merleau-Ponty's objection though, part of the point is that it

presupposes a view of bodies as unconnected with the psychological which will not allow psychological attributions to get a grip.

There are also differences between the two philosophies. For Merleau-Ponty and the other phenomenologists examined in this thesis our experiences of other subjects is immediate and a precondition of experience itself. For Wittgenstein an experience of other subjects is constituted through the body of practices that are exclusive to other agents. Wittgenstein also attaches more importance to encounters with a human face as facilitating our intersubjective experiences, often drawing one into intersubjective relations despite oneself. One cannot help being effected by the psychological expression of others. This brings Wittgenstein closer to Merleau-Ponty in that sensitivity to these normative qualities of a person is a precondition for participating in the associated practices. However this is still not as strong a claim as the phenomenologists are making that the world is already essentially an intersubjective one. For them this will still be the case even if one is in total isolation from others. Things in the world can only have meaning because they can also potentially have meaning for others too.

Wittgenstein's answer to question 1 is that we are participating in shared psychological practices that presuppose the existence of shared psychological states. Like Merleau-Ponty, Wittgenstein does not portray us as involved in an epistemological enterprise. Wittgenstein's answer to question 2 is that it is the practices themselves and the attitudes to others embedded in them that justify our

413

psychological attributions. There is no epistemological question to be answered so no deeper justification is required. Wittgenstein's answer to question three can be elucidated through consideration of the following quote discussing the manner in which concepts and 'facts of nature' correspond:

"I am not saying: if such-and-such facts of nature were different people would have different concepts (in the sense of a hypothesis). But: if anyone believes that certain concepts are absolutely the correct ones, and that having different ones would mean not realizing something that we realize – then let him imagine certain very general facts of nature to be different from what we are used to, and the formation of concepts different from the usual ones will become intelligible to him".[913]

Wittgenstein was not of the view that empirical investigation of natural facts would reveal what psychological practices consist in. On his view this can only be discovered through attending to the practices themselves. Nonetheless Wittgenstein would not deny natural facts play some role in facilitating our psychological practices. The above quotation recommends an exercise intended to make us reflect on the relationship between natural facts and our understanding of our personal level practices. By asking us to consider the natural facts being quite different he aims to unseat any complacency we might have about the way we now conceptualise matters. Very different facts would likely lead to a different set of concepts or practices. It is unproductive to speculate in advance exactly how our views would change so Wittgenstein does not attempt to offer a hypothesis in which particular

[913] Wittgenstein 1953 P.230

414

facts lead to modifications in particular beliefs about the world. This is because the relationship between natural facts and our beliefs and practices is more complex. Our knowledge of natural facts contextualize the setting against which our psychological practices take place and indirectly serve to shape the practices themselves. Wittgenstein does not owe us an answer to question 4 because he does not answer question 1 in a way that makes this question arise.

In this thesis I hope to have shown that the contemporary debates between TT, ST and DP are predicated on certain kinds of answer to question 1 and that their answers to this question about what we are doing when attributing psychological states to others give rise to a number of problems that have been articulated in this thesis. The positions offer us inaccurate descriptions of the practices and in particular an inaccurate portrayal of the relations we stand in to other agents as primarily epistemological. I have also been contrasting these answers with answers drawn from the writing of Wittgenstein and Merleau-Ponty. These writers provide us with a different kind of answer to question 1 which involves a different model of our relation to others, which is not an epistemological one. Both offer useful insights into the true nature of our psychological practices. The problems raised in relation to the other answers to question 1 do not arise for these philosophers. As a consequence of their alternative approach to question 1, Merleau-Ponty and Wittgenstein have a different approach to the empirical data. This data is relevant to a question about what kinds of physiological or sub-personal processes would facilitate the possibility of the practices then described. However the sub-personal

is assigned no role in the justification for a particular account of the practice.

Bibliography

Aydede, M (2010) "The Language of Thought Hypothesis" *The Stanford Encyclopedia of philosophy* Zalta E (ed.) Available at: http://plato.stanford.edu/archives/fall2010/entries/language-thought/ (accessed April 2012)

Baron-Cohen, S, Leslie AM & Frith U (1985) "Does the autistic child have a "theory of mind"" *Cognition* Vol.21, 1 pp.37–46

Baron-Cohen, S Tager-Flusberg, H & Cohen, DJ (eds.) (1993) *Understanding other minds: perspectives from autism* Oxford University Press, Oxford

Bishop, M & Downes, S (2002) "The theory theory thrice over: the child as scientist, Superscientist or social institution?" *Studies in History and Philosophy of Science* Vol.33 pp.121–136

Botterill, G & Carruthers, P (1999) *The Philosophy of Psychology* Cambridge University Press, Cambridge

Carruthers, P (1996a) "Introduction" in Carruthers, P & Smith, P.K (eds.) *Theories of theories of mind* Cambridge University Press, Cambridge

Carruthers, P (1996b) "Simulation and Self-knowledge" in Carruthers, P & Smith, P.K (eds.) (1996a) PP.22-38

Carruthers, P (2006) The Architecture of the Mind Clarendon Press, Oxford

Carruthers, P (2006b) "Review of Alvin I. Goldman, Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading" *Notre Dame Philosophical Reviews* (11) Available at: https://ndpr.nd.edu/news/25164-simulating-minds-the-philosophy-psychology-and-neuroscience-of-mindreading/ (Accessed December 2013)

Charles D and Lennon K (Eds.) (1992) *Reduction, Explanation and Realism* Clarendon Press, Oxford

Child, W (2014) "Wittgenstein on Inner-Outer and Avowals" in H-J Glock, H J and Hyman, J (eds) (forthcoming) Wiley-Blackwell, Oxford

Churchland, P (1981) "Eliminative Materialism and the Propositional Attitudes" *Journal of Philosophy* Vol.78 pp.67-90

Cockburn, D (2001) *An introduction to the philosophy of mind* Palgrave Hampshire

Cockburn, D (1990) *Other Human Beings* Macmillan, Hampshire

Currie, G and Ravenscroft, I (1997) "Mental Simulation and Motor Imagery"

Philosophy of Science Vol.64, 1 pp.161-180

Davies, M and Stone, T (eds.) (1995a) *Folk Psychology: The Theory of Mind Debate*

Blackwell Publishers, Oxford

Davies, M and Stone, T (1995b) "Introduction" in Davies, M and Stone, T (eds.)

(1995a) pp.1-44

Davies, M & Stone, T (2001) "*Mental Simulation, Tacit Theory, and the Threat of*

Collapse" Available at:

http://www.nyu.edu/gsas/dept/philo/courses/content/papers/davies.pdf (accessed

May 2014)

Davies, M (1994) "The Mental Simulation Debate" *Philosophical Issues* Vol.5

pp.189-218

Dennett, D (1969) *Content and Consciousness* Routledge, London

Dennett, D (1987) *The Intentional Stance* M.I.T, Bradford

Edie, J M (ed.) (1964a) *The Primacy of Perception and other essays on*

phenomenological psychology, the philosophy of art, history and politics
Northwestern University Press USA

Fadiga, L Fogassi, L Pavesi, G and Rizzolatti G (1995) "Motor facilitation during action observation: a magnetic stimulation study" *Journal of Neurophysiology* Vol.73, 6 pp.2608-11

Fodor, J A (1983) *The modularity of mind: an essay on faculty psychology* MIT, Cambridge Massachusetts

Fogassi, L Ferrari, P. F. Gesierich, B Rozzi, S Chersi, F & Rizzolatti, G (2005). "Parietal lobe: From action organization to intention understanding" *Science* Vol.308 pp.662–667

Freedberg, D and Gallese, V (2007) "Motion, emotion and empathy in esthetic experience" *Trends in Cognitive Sciences* Vol.11 pp.197-203

Gallagher, S (2007) "Simulation Trouble" *Social Neuroscience* Vol.2 pp.353-65

Gallagher, S & Zahavi, D (2008) *The Phenomenological Mind an introduction to philosophy and cognitive science* Routledge, USA

Gallagher, S (2008) "Direct Perception in the intersubjective context"

Consciousness and Cognition Vol.17 pp.535-543

Gallagher, S (2011) "Strong interaction and self-agency" *Humana-Mente: Journal of Philosophical Studies* Vol.15 pp.55-76

Gallese, V and Goldman, A (1998) "Mirror neurons and the simulation theory of mind-reading" *Trends in Cognitive Sciences* Vol.2, 12 pp.493-501

Gallese, V (2003) "The roots of empathy: The shared manifold hypothesis and the neural basis of intersubjectivity" *Psychopathology* Vol.36, 4 pp.171-180

Gallese, V (2006) "Intentional attunement: A neurophysiological perspective on social cognition and its disruption in autism" *Brain Research* Vol.1 pp.15-24

Gallese, V Eagle, M & Migone, M (2007) "Intentional Attunement: mirror neurons and the neuronal underpinnings of interpersonal relations" *Journal of the American Psychoanalytic Association* Vol.55 pp.131-175

Gallese, V & Freedberg, D (2007) "Mirror and canonical neurons are crucial elements in esthetic response" *Trends in Cognitive Sciences* Vol.11 p.411

Gallese, V (2007) "Empathy, embodied simulation and mirroring mechanisms. Commentary on "Towards a neuroscience of empathy" by Doug Watt"

Neuropsychoanalysis Vol.9, 2 pp.146-151

Gallese, V (2008) "Empathy, Embodied Simulation, and the Brain: Commentary on Aragno and Zepf/Hartmann" *Journal of the American Psychoanalytic Association* Vol.56, 3 pp.769-781

Gallese, V (2009) "Mirror Neurons, Embodied Simulation, and the Neural Basis of Social Identification" *psychoanalytic Dialogues* Vol.19 pp.519–536

Georgieff, N and Jeannerod, M (1998) "Beyond consciousness of external reality: A "who" system for consciousness of action and self-consciousness" *Consciousness and Cognition* Vol.7, 3 pp.465-477

Glock, H J. & Hyman, J. (eds.) (forthcoming) *The Blackwell Companion to Wittgenstein* Wiley-Blackwell, Oxford

Goldman, A (1989) "Interpretation Psychologized" *Mind and Language* Vol.4 pp.161–185

Goldman, A (2006) *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading* Oxford University Press, New York

Goldman, A (2009) "Mirroring, Mindreading, and Simulation" in Pineda, J. (ed)

(2009a) pp.311-330

Goldman, A (2009b) "*Jacob on Mirroring, Simulating and Mindreading*" Available at**:** http://fas-

philosophy.rutgers.edu/goldman/Mirroring,%20Simulating,%20and%20Mindreadin

g.pdf (accessed September 2010)

Gopnick, A (1996) "The Scientist as Child" *Philosophy of Science* Vol.63, 4 pp547-

551

Gordon, R (1986) "Folk Psychology as Simulation" *Mind & Language*,Vol.1

 pp.158–171 reprinted in Davies M and Stone T (eds.) (1995a) pp.60-73

Gordon, R (2009) "Folk Psychology as Mental Simulation" *The Stanford*

Encyclopedia of Philosophy (Fall 2009 Edition), Zalta, E N (ed.)

http://plato.stanford.edu/archives/fall2009/entries/folkpsych-simulation/ (accessed

September 2010)

Greenwood, J (1991a) *The Future of Folk Psychology: Intentionality and Cognitive*

Science Cambridge University Press, Cambridge

Greenwood, J (1991b) "Folk psychology and scientific psychology" in Greenwood

(ed.) (1991a) pp-1-21

Greenwood, J (1991c) "Reasons to believe" in Greenwood (ed.) (1991a) pp-70-92

Heal, J (1986) "Replication and Functionalism" in Davies, M and Stone, T (eds.) (1995a) pp.45-59

Heal, J (1994) "Simulation vs. theory theory: What is at issue?" In Peacocke, C (Ed.) (1994) pp. 129–144

Heal, J (1996) "Simulation, Theory, and Content" in Carruthers, P & Smith, P.K (eds.) (1996) pp. 75–89

Heidegger, M (1962) *Being and Time* trans. by John Macquarrie & Edward Robinson: SCM Press, London (first published in 1927)

Hintikka, J (ed.) (1991) *Wittgenstein in Florida Proceedings of the Colloquium on the Philosophy of Ludwig Wittgenstein, Florida State University, 7–8 August 1989 Springer* Kluwer, Netherlands

Hornsby, J. (1992) 'Physics, Biology and Commonsense Psychology', in Charles D and Lennon K (Eds.) (1992) pp.155-178

Hornsby, J (2000) "Personal and Sub-personal: A defense of Dennett's early distinction" *Philosophical Explorations*: Vol.3 pp.6-24 Available at: 'http://eprints.bbk.ac.uk/102/1/hornsby2.pdf' (accessed November 2013)

Hurley, S (2006) *Active perception and perceiving action: The Shared Circuits Hypothesis* Available at: http://www.bristol.ac.uk/philosophy/hurley/papers/ap_and_pa_shared_circuits.pdf (accessed September 2010)

Hutto, D (2008) "Lessons from Wittgenstein: Elucidating folk psychology" *New Ideas in Psychology* vol.27, 2 pp.197-212

Iacoboni, M. (2009) "The Problem of Other Minds Is Not a Problem: Mirror Neurons and Intersubjectivity" in Pineda, J. (ed) (2009) pp.121-133

Jeannerod, M & Pacherie, E (2004) "Simulation and Self-Identification" *Mind and Language* Vol.19, 2 pp.113-146

Leslie, A, Roth, D (1993) "What autism teaches us about metarepresentation" in Baron-Cohen, S. Tager-Flusberg, H. Cohen, DJ. (eds.) 1993 pp.83–111

Leslie, A (1987) "Pretense and Representation: The Origins of "Theory of Mind"" *Psychological Review* Vol.94, 4 pp.412-426

Lewis, D (1972) "Psychophysical and Theoretical Identifications" *Australasian Journal of Philosophy* Vol.50, 3 pp.249-258

Malcolm, N (1954) "Wittgenstein's Philosophical Investigations" *The Philosophical Review* Vol.63 4 pp.530-559

McDowell, J (1982) "Criteria, defeasibility and knowledge" *Proceedings of the British Academy* vol.68 pp.455–479

McGinn, M (1996) *Wittgenstein's Philosophical Investigations* (Routledge Guides to philosophy) Routledge, London

McGinn, M (1998) "The Real Problem of Others: Cavell, Merleau-Ponty and Wittgenstein on Scepticism about Other Minds" *European Journal of Philosophy* Vol.6 pp.45–58

Merleau-Ponty, M (1962) *Phenomenology of perception* Routledge & Kegan Paul, London (first published in 1945)

Merleau-Ponty, M (1964b) The child's relation to others in Edie, J M (ed.) (1964a) pp.96-155

Mill, J (1867) *An Examination of Sir William Hamilton's Philosophy and of the principal philosophical questions discussed in his writings* Longmans and Co, London

Morton, A (1980) *Frames of Mind: Constraints on the common sense conception of the mental* Clarendon Press, Oxford

Peacocke, C (Ed.) (1994) *Objectivity, simulation, and the unity of consciousness: Proceedings of the British academy vol. 83* Oxford University Press, Oxford

Perner, J (1996) "Simulation as explanation of predication-implicit knowledge about the mind: Arguments for a simulation-theory mix" in Carruthers, P & Smith, P.K (eds.) (1996) pp.90–104

Pineda, J (ed) (2009) *Mirror Neuron Systems: The Role of Mirroring Processes In Social Cognition* Humana Press, USA

Premack, D. G & Woodruff, G (1978). "Does the chimpanzee have a theory of mind?" *Behavioural and Brain Sciences* Vol.1 pp.515-526.

Read, R (2010) "Wittgenstein's Philosophical Investigations as a war book" *New Literary History* Vol.41 pp.593-612

Robbins, P (2010) "Modularity of Mind", *The Stanford Encyclopedia of Philosophy (Summer 2010 Edition)* Zalta E N(ed.) Available at:

http://plato.stanford.edu/archives/sum2010/entries/modularity-mind/ (accessed July 2010)

Sartre, JP (1943) *Being and Nothingness: An Essay on Phenomenological Ontology* (English 1958 Translation by Hazel Barnes) Routledge, London

Sinigaglia & Sparaci (2010) "Emotions in action through the looking glass" *Journal of Analytical Psychology* Vol.55 pp.3–29

Stich, S & Nichols, S (1992) "Folk Psychology: Simulation or Tacit Theory" *Mind and Language,* Vol.7 pp.35-71 reprinted in M. Davies and T. Stone (eds.), *Folk Psychology: The Theory of Mind* pp.123-158

Stich, S & Nichols, S (1995) "Second thoughts on simulation" In Davies and Stone (eds), (1995) pp.87–108

Stich, S & Nichols, S (1998) "Theory Theory to the max" *Mind and Language* Vol.7 (1-2) pp.145-71

Stich, S & Nichols, S (2003) *Mindreading An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds* Oxford University Press, Oxford

Ter Hark, M R M (1991) "The development of wittgenstein's views about the other minds problem" in Hintikka, J (ed.) 1991 pp.227-253

Umiltà, MA Kohler, E Gallese, V Fogassi, L Fadiga, L Keysers, C. and Rizzolatti, G (2001) " I know what you are doing. A neurophysiological study" *Neuron* vol.31 pp.155–165

Wimmer, H & Perner, J (1983) "Beliefs about beliefs representation and constraining function of wrong beliefs in young children's understanding of deception" *Cognition*, Vol.13 pp.103-128

Wittgenstein, L (1953). *Philosophical Investigations* Blackwell, Oxford