

**UNIVERSITEIT VAN AMSTERDAM**

**UNDERSTANDING ATTENTION**

Submitted in partial fulfilment of the  
degree Masters in Philosophy

Supervisor:

**DR ELSBETH BROUWER**

Second Reader:

**PROF.DR MICHEL VAN LAMBALGEN**

By: **Nick Young**

**5751551**

**20-05-10**

# TABLE OF CONTENTS

INTRODUCTION

CHAPTER 1: WHAT IS UNDERSTANDING?

CHAPTER 2: UNDERSTANDING ATTENTION

CHAPTER 3: ATTENTION AND THE REPRESENTATIONAL THEORY OF MIND

CHAPTER 4: ATTENTION AND DYNAMICAL SYSTEMS

CHAPTER 5: ATTENTION AND ACTION ORIENTED REPRESENTATIONS

CONCLUSIONS

## Introduction

The primary goal of this thesis is to discover what an understanding of attention consists of. In everyday life, the word 'understanding' is used to mean different things. We can for example talk about linguistic understanding, whether the meaning of a word or phrase is understood. Alternatively, 'understanding' is sometimes used as a synonym for 'empathy', as in 'She showed a great deal of understanding when I told her my problem'. The type of understanding I will be concerned with here is closely linked to a person's being able to interact intelligently with people and objects in the world. We can, for example, talk about somebody understanding *how* to ride a bike or play tennis. We might also say that it was a person's understanding *of* car engines which allowed him to fix the car.

By 'attention' I simply mean the capacity possessed by human beings and other living creatures to perceive aspects of the environment, and to be able to move so as to control what is perceived. Although we can attend to smells or sounds, the type of attention that I will focus on here is visual attention: our ability to see objects in the world, and to control what is seen at by turning the head or eyes.

Given these two starting points, we might say that someone who understands attention is someone who can act intelligently when dealing with the perceptual capabilities of others. For example, being able to direct another's attention through pointing, or responding to what another is attending to by following their gaze. My principal aim here will be to determine how such behaviours are possible.

In Chapter 1 I will introduce three theories of cognition: the Representational Theory of Mind (RTM), Dynamical Systems (DS), and the Action Oriented Representations approach (AOR). We will see that the key difference between these three accounts is the value that they place on mental representations as the cause of intelligent action. Supporters of RTM hold that action is caused by states of the brain which bear descriptive information about the state of the world. Supporters of DS reject this idea and argue that the brain states involved in intelligent interaction should not be considered representational. Proponents of AOR argue that some states of the brain should be considered representational, but that these representations differ from those proposed by RTM in that they contain anticipatory information about what possibilities for action a situation affords, rather than descriptive information that the world is a particular way.

Having properly introduced the different approaches to understanding, I will turn in Chapter 2 to attention. Specifically, my focus will be on studies of infant development which show how behaviours suggestive of attention understanding (for example pointing or gaze following) become more sophisticated in the first 12 months of life. We shall see that psychologists and philosophers interested in attention disagree over whether these behaviours require the infant to possess a "concept of attention" and that parallels can be drawn between this debate and the questions raised in Chapter 1.

Chapters 3, 4 and 5 each critically assess the success of a theory in accounting for these empirical data. Chapter 3 addresses RTM, Chapter 4 discusses DS, and chapter 5 is concerned with AOR . While discussing these theories, I will focus on three aspects of attention understanding behaviour: inconsistent gaze following at 3 months, the development of directive behaviours between 8 and 12 months, and gaze checking behaviours at 12 months. These aspects will provide the basis for discussing the relation of the theories in question to issues such as: the role motor skills play in attention understanding behaviour, how behaviours develop in sophistication over time, and the necessity of information processing in attention understanding. My conclusion will be that AOR is the most successful of the three theories in accounting for the data and explaining the issues.

# Chapter 1 What Is Understanding?

## 1.1 The Representational Theory of Mind

One answer to the question of what understanding is can be put in terms of the Computational or Representational Theory of Mind (hereafter RTM) which conceives of mental activities like perceiving, thinking, reasoning and imagining as the manipulation of mental representations. RTM has a very long history in philosophy and psychology (for example Descartes, 1641/1991, 188; Locke 1690/1838, Chapter XXIX). A contemporary version of the theory is summarised by Steven Pinker:

For all its exquisite engineering, an eye is useless without a brain. Its output is not the meaningless patterns of a screen saver, but raw material for circuitry that computes a representation of the external world. That representation feeds other circuits that make sense of the world by imputing causes to events and placing them in categories that allow useful predictions. And that sense-making, in turn, works in the service of motives such as hunger, fear, love, curiosity, and the pursuit of status and esteem

(Pinker 2002, 52)

On this view, when I receive perceptual input from my senses, this input is converted into inner states that bear information about the world (mental representations), the processing of which gives rise to action. If a bus is coming towards me it will cause a distinctive pattern of sensory data (light on the eyes) which will cause representations bearing the content 'a bus is coming towards me'. This information is processed in sense that a chain of reasoning takes place in which the consequences of doing nothing (being hit by the bus) are recognised as undesirable and a calculation is made as to how best avoid this outcome, leading to my getting out of the way.

This list of features, adapted from Rowlands (2006, Chapter 1) and Gallagher (2008)<sup>1</sup>, allows us to define this type of representations more formally:

---

<sup>1</sup> A point that both Gallagher and Rowlands include which is missing here is the condition that representations should be “decoupleable” from action. Gallagher defines decoupleability as “The idea of decoupleability...is that one can take the representation ‘offline’ and, for example, imagine or remember an action or context. Representation in such cases involves a form of decoupling away from action, away from the target of action, or from the current context.” (Page 352). My main reason for this omission is to avoid adding any more jargon to the discussion than is necessary. These issues are important but will be subsumed into my discussions of information in the first section of this chapter and the relationship between representation in action and offline representation in Chapter 5.

RP1 Representation is internal (image, symbol, neural configuration)

RP2 Representation has duration (it's a discrete identifiable thing)

RP3 Representation bears content that is external to itself (it refers to or is about something other than itself)

RP4 Representation is passive (it is produced, enacted, called forth by some particular situation; or we do something with it)

(Rowlands, 2-9)

The first three of these features are made clear in the context of the bus example. Seeing the bus causes a particular pattern of activity in my brain (RP1), which lasts for a finite amount of time (RP2) and this pattern of activity is (somehow) *about* a bus coming towards me (RP3). Theories differ with regards to what these internal states consist of and how they bear content that is external to themselves. Most present day supporters of this view would consider mental representation to be realised in terms of particular neural configurations, but how the information encoded in these brain states should be thought of is a point of disagreement. Different opinions include those that argue representational information is image-like (Kosslyn 1980), language like (Fodor and Pylyshyn 1988), or a hybrid of the two (Tye 1991); in Chapter 3 we shall see that representations of other minds are sometimes considered in terms of either theoretical or simulatory information. A common thread through all of these variants is that representational information is *descriptive* information *that* the world is a particular way: the (real life) bus could cause an image-like representation showing that the world is a particular way; alternatively, it could cause a language like representation which describes the details of the scene. The sight of a person could cause a representation containing theoretical information that he has this belief or that or desire; or, information detailing that the other person has a mental state like *this* simulated mental state. A detailed discussion of what type of representational information is the most suitable for describing others' mental states will be discussed in detail in Chapter 3.

RP4 shows that representations are thought to be caused and manipulated in very many ways. A 'bus approaching' representation would typically be produced by sensory input caused by an approaching bus but a similar representation might also be caused by my imagining or remembering a bus coming towards me, or hearing somebody say "A bus is coming towards you". As well as this, both representations carry with them a similar type of logical structure which allows for rational thought. As Pinker says "cognition requires architecture equipped with logical apparatus like rules, variables, propositions, goal states, and different kinds of data structures, organized into larger systems..." (80). I can reason from one representation to another: from my representation that the bus is coming towards me I can deduce what will happen if I do nothing, or imagine possible actions I could take and their

consequences. Alternatively, some representations necessarily exclude others, I cannot simultaneously represent the bus as coming towards me and going away from me, and neither can I represent it as being all red or all blue.

Before moving on, there are three more elements of RTM that should be noted: the role that goals play in the theory, the fact that representation processing can be conscious or unconscious, and RTM's account of nuanced, complex interactions with the environment. On the RTM view, all representation crunching, and therefore all action in the world, would be driven by a hierarchy of underlying goal states: the action of tying shoelaces is one step towards achieving the goal of walking into town, walking into town is a step towards the goal state of buying food; moving one's body out of the way of the bus is a step towards the goal state of staying alive and unharmed. This seems a fairly intuitive, 'common sense' conception of action, as we shall see in the next section however, this idea is explicitly rejected in non-representational theories.

The processing of descriptive representations as a means of achieving goal states is claimed by RTM to constitute both conscious, reflective reasoning, as well as the sub-personal processes that underlie unreflective action. In some cases we are consciously aware of what our current goals are as well as the line of reasoning we employ to work out how best to achieve them; in other cases, goals and reasoning may be entirely hidden from conscious experience. It would seem that the majority of acting in the world would stem from this latter type of reasoning; we are not consciously aware of every calculation or goal we pursue, our interactions with objects or other people (at least non-verbal ones) are very often done on 'automatic pilot', without reflection or awareness.

One form of representation processing that would likely take place at the the sub-personal level, is that which controls the fine details of how we move our bodies in the world. So far, the type of representation processing I have discussed has been of quite a general type, inferring whether it is best to move or stay when confronted by an on coming bus. However, RTM is also employed to explain the more nuanced aspects of intelligent behaviour in which the perceived details of a situation have a bearing on the exact action made. For example, if I see a cup on the table in front of me, and go to pick it up, I do not typically over or under extend my arm. RTM accounts for this by saying that my my perception of the scene gives rise to detailed representations of (for example) the location and shape of objects in front of me. A cup on the table is represented as being 10 cm away from my hand and so I can compute the angle, direction and force needed to move my hand so that it is next to the cup. I shall adapt a term of Wheeler's (2005, 207) and refer to representations that bear detailed information about the world for the purpose of guiding the intricacies of intelligent interaction as “instructionalist” representations.

## **The Representational Mind and Understanding**

On the RTM view, understanding can be thought of as possessing the appropriate representations, which give rise to the appropriate computational structure, which will typically lead to appropriate actions being made. For example, a key component of understanding what a dog is would be for perceptions of dogs to cause a 'dog' representation. If the 'dog' representation is accurate it will interact with other representations in an appropriate way and thereby cause appropriate dog interaction behaviour like stroking, taking for walks and putting on a lead. If a person's perceiving of a dog gives rise to a representation that does not interact correctly with other representations and so does not give rise to the appropriate behaviour (offering a bowl of milk, or putting in a fish tank) then we would say that the person in question does not understand what a dog is. This conception ties understanding very closely to appropriate action, but at the same time does not identify my understanding as my behaving appropriately. Rather, my understanding is my possession of the appropriate representation which causes my dog interaction behaviour.

## **A “Degenerating Research Project” Verses Embedded, Embodied Cognition**

The account so far is a very general one and many theories of mind can be described as representational. The details as to what counts as a representation and how representations interact with one another depends on discipline and historical context. As mentioned, more modern versions of the theory consider representations to be neural configurations of the brain. Some schools of thought in Artificial Intelligence have thought that representations can be realised electronically and from this starting point attempted to build robots that can understand aspects of the world and interact with them in the same way as humans.

However, the RTM view is not unchallenged. In the first half of the 20th century the theory was criticised by philosophers such as Heidegger (1962), Wittgenstein (1953/1967) and Merleau-Ponty (1945/2002), but despite this it remained the orthodox conception of mind in much of philosophy as well as disciplines such as Cognitive Science and Artificial Intelligence. It was not until the 90s that criticisms of RTM and non-representational accounts of mind began to gain traction.

One of the main criticisms made against the theory is its lack of success in Artificial Intelligence (Clark 1998 Chapter 1; Dreyfus 1992, 2002, 2007), Dreyfus, for example, labels the use of RTM as a basis for AI as a “degenerating research project” (Dreyfus 2007, 2) . Since the 60s there has been attempts to build robots with human like intelligence by equipping them with the capacity to construct and manipulate instructional representations (for example Dennett 1994; Lenat and Feigenbaum 1992) but, despite huge advances in robotics and computing, this form of AI has failed to create intelligent robots. Two accusations often levelled at instructional robots are that the intelligence they display is 'brittle' (they are unable to adapt their behaviour so as to accommodate novel changes in their

environment)(Clark 1998, 3) and that they are unable to adequately deal with the frame problem (Dreyfus 1991, 1992; Wheeler 2005).

In response to these criticisms and others, alternative views of cognition have emerged in cognitive science. I will focus on two such alternative here, both of which can be considered varieties of Embodied, Embedded Cognition (EEC). EEC theories have two main ideas in common: The first is a rejection of the view that intelligent action is caused by descriptive representations being processed. The second is the claim that the brain is not the sole source of intelligent actions; instead, the brain, the body it belongs to and the environment in which that body is situated (embedded in), all play a significant role. Where opinions differ is the question of whether action can be explained without invoking representations at all, or whether some stripped-down, 'minimal' form of representation is still required.

## **1.2 Dynamical Systems and Radical Embodied Cognition**

A complete rejection of representations as a means of explaining understanding is what Clark calls “the thesis of Radical Embodied Cognition” (1998, 148). Supporters of this view include Gallagher (2008), Dreyfus, and van Gelder (1997). Instead of intelligent action being thought of in terms of representations we are invited to think of brain, body and environment as composing a “Dynamical System” in which "the actions of the parts cause the overall behaviour and the overall behaviour guides the actions of the parts" (Clark 1998, 107). An example of a system such as this would be:

...a radio receiver, the input signal to which is best treated as a continuous modulator of the radio's " behavior " (its sound output ). Now imagine... that the radio 's output is also a continuous modulator of the external device (the transmitter ) delivering the input signal. In such a case, we observe a truly complex and temporally dense interplay between the two system components one which could lead to different overall dynamics (e.g. of positive feedback or stable equilibria ) depending on the precise details of the interplay.

(Clark, 163)

Here it is not possible to isolate either the receiver or the transmitter as the cause of behaviour, each feeds back to the other so that a circle of interaction is formed. Proponents of DS would argue that this type of “dynamic coupling” (Dreyfus 2007, 25) is present when we perform intelligent actions. When we are driving a car, we receive visual input from the environment, for example, we see the straightness of the road or the sharpness of the turn ahead of us. This input affects bodily movement,

the degree to which the steering wheel is turned and/or an increase or decrease of pressure on the accelerator, and this will in turn affect the subsequent visual input we receive. This differs from the instructional explanation because at no time is there a detailed representation of the world created in the driver's mind; no part of the brain is thought of as containing information that the road is a particular way. Neither is there any implicit, sub-personal calculation being performed as to how far the steering wheel should be turned to navigate a curve. Instead the driver's responses are 'tuned' to respond to the subtle changes in input that he is receiving.

A more detailed description of how body, brain and environment interact, and how a system is tuned so as to respond appropriately, can be found in Dreyfus (2007):

...acting is experienced as a steady flow of skilful activity in response to one's sense of the situation. Part of that experience is a sense that when one's situation deviates from some optimal body-environment gestalt, one's activity takes one closer to that optimum and thereby relieves the "tension" of the deviation. One does not need to know what that optimum is in order to move towards it. One's body is simply solicited by the situation [the gradient of the situation's reward] to lower the tension. Minimum tension is correlated with achieving an optimal grip.

(16)

The studies that Dreyfus is discussing detail Freeman's research on the brain activity of rabbits when they smell food (Freeman and Grajski, 1987). According to Freeman, upon smelling the food the rabbit's olfactory bulb -the part of the rabbit's brain that deals with smells- exhibits a surge neural activity and it is this that can be thought of as the tension that the rabbit seeks to relieve. The rabbit's learned skills or instincts can be thought of as the system's preprogrammed routines with which to relieve this tension. In other words, the activity caused in the bulb by the smell of food initiates particular behaviours in the rabbit, like moving towards the food, or eating it. In terms of the driving example we might say that a turn in the road causes a degree of tension in the brain of the driver, to which his learned response is to turn the steering wheel. This approach also explains the fine-grained nature of intelligent action; a turn with a particular degree of curvature will cause a particular degree of neuronal tension, which will determine the degree to which the steering wheel is turned.

Learning is accounted for by the claim that the animal's neural connections are then "changed in a way that reflects the extent to which the result satisfied the animal's current need" (Dreyfus 2007, 20). This changing should not be thought of as a rule being constructed and stored in expectation of a similar situation but instead as a subtle tuning of a 'starting state' of the brain:

inputs [are] associated with outputs, but the hidden nodes of the network would always already be in a particular state of activation when input stimuli were received, and the output that the network produced would depend on this initial state. Thus input plus initial state would determine output. If the input corresponds to the experience of the current situation, the activation of the hidden nodes, determined by inputs leading up to the current situation, might be said to correspond to the expectations and perspective that the expert brings to the situation, in terms of which the situation solicits a specific response

(Dreyfus 2002, 374)

On this account, learning how to drive around corners is a case of an initial response to corners being slowly refined each time a corner is approached. Each experience of turning a corner alters the initial state leading to future turns being made with increased competence.

### **Dynamical Systems and Understanding**

The main consequence that DS has for understanding is that it is much more difficult to pinpoint any one neural process that could be said to constitute a person's understanding. Even if we concede that a person's understanding of something does not depend on neural configurations representing the world, there is still an inclination to think that a person's understanding *x* consists of a particular process (or capacity to perform a particular process) in the brain that is the cause of intelligent behaviour relating to *x*. However, DS emphasises that intelligent action is the product of interaction between brain, body and environment, and, as Wheeler puts it, “no part of the extended system should be explanatory privileged over any other” (243). Skilled driving behaviour would not be displayed if a particular pattern of neural activity was not in play, but neither would this behaviour be displayed if certain features of the environment (a car and a road) were not present. If intelligent behaviour depends just as much on features of the environment as features of the brain then it makes as much sense to say that the understanding is 'in the road' as it would to say that it is 'in the brain'.

In summary, the DS view presents a way of describing how intelligent action is possible without there being a neuronal configuration (representational or otherwise) that could be called a person's understanding. Alternative ways to define understanding that do not identify it with a feature of the brain will be examined when I contrast 'lean' and 'rich' theories of attention understanding in the next chapter. Before turning to attention however, I will demonstrate one final way in which understanding can be understood.

### **1.3 Action Oriented Representations**

A second type of EEC theory, that can be thought of as a compromise between DS and RTM, I shall call the Action Oriented Representation (AOR) approach. The compromise made to Dynamical Systems is the acceptance that dynamic coupling between brain, body and environment plays a large role in intelligent action, and in some cases is entirely responsible (Clark and Grush, 13; Wheeler, 250-251). The compromise made to the more orthodox view is that a stripped down form of representation is required in others. There are different types of AORs (sometimes referred to as 'minimal' representations) proposed (see Clark 1998; Clark and Grush 1999; Rowlands 2006; Wheeler 2005) but a common thread through all of them is that the information they carry is geared towards doing rather than describing:

...the brain should not be seen as primarily a locus of inner descriptions of external states of affairs; rather, it should be seen as a locus of inner structures that act as operators upon the world via their role in determining actions.

(Clark 1998, 47)

This view of representation retains the four features mentioned earlier but heavily modifies 3. RTM representations can be thought of bearing descriptive information that the world is a certain way whereas all forms of AOR would bear information detailing "possibilities for action" (Wheeler, 197). For example, an instructionalist representation might have the information "the curvature of the turn ahead is 28 degrees", with a further calculation required to determine how far to turn the steering wheel and what muscle contractions would be needed for such a turn. An AOR would bear information about what possibilities for action the current situation offers; the turn could be carefully followed, swerved around at speed, stopped dead at, or ignored completely and driven over. Supporters of the AOR approach sometimes, following Gibson (1979), describe the possibilities as "affordances" and I will follow their lead; a cup affords the possibilities of picking up, drinking from or filling with liquid; a dog of stroking or playing fetch.

#### **Action Oriented Representations Verses Dynamic Systems**

A little more work is needed to show why AOR affordances should be thought in terms of representations. As it currently stands, the notion of affordances could also be accommodated by DS; the cup affording picking up, or the dog affording stroking, could be thought of as the objects eliciting tension in the brain which is dissipated through appropriate action. What needs to be shown is that affordances as conceived by AOR detail possible actions that could be made, whereas DS would appear

to reject the idea that possibilities are presented and chosen from. In order to do this I will focus on Clark and Grush's proposal that AORs should be thought of as emulators. This will achieve two things: Firstly, it will put clear water between AOR and DS conceptions of affordances by emphasising the informational element of AOR; secondly, it will show that the presence of information leads to AOR retaining the RTM notion of goal states.

Emulators are used in real life to allow one system to mimic the behaviour of another system, Clark and Grush give the example of a chemical plant:

chemical plants that need to control an on-going reaction by adding chemicals to a mix but where waiting for feedback cues to prompt the process is impossible since, by the time the cues are received, it would be too late for the chemical infusions to work

(5)

Instead a computer program is used which “models the target system and generates a kind of mock feedback that can be used instead of the laggardly feedback from the real system”. This allows the controllers of the chemical mix to anticipate the consequences of a possible action before they make it. Here we can see a process in which information and representation play a key role. Mock input is fed into the emulator regarding a possible change to the chemical mix, and mock output is received detailing the effects of this change. Clark and Grush argue that AORs play a similar role in intelligent action. Gallagher summarises their position as follows:

The circuitry is a model...that stands in for a future state of some extra-neural aspect of the movement – a body position (or proprioceptive feedback connected with a body position) just about to be accomplished, e.g. in the action of catching a ball...the emulator anticipates (represents) an  $x$  that is not there – a future  $x$  – or a predicted motor state

(Gallagher 2008, 357)

Again we can see a mock input (a possible body position), an emulator, and the mock output it produces (the “predicted motor state”). On this view, my appreciation of the affordances that an object presents consists of my having an emulatory representation which contains information regarding the consequences of possible actions. On this view an object is represented as being a bundle of affordances, as a collection of anticipations as to what would happen if a particular bodily movements were performed: The cup on the table affords picking up because I anticipate that if I move my arm and hand in the correct way then I will lift the cup off the table. It also affords the possibility of being filled with liquid because I anticipate that if I perform bodily movements that involve picking up the

cup and holding it underneath a running tap, it will contain the liquid. Through the feeding of mock input into the AOR, a range of affordances is produced.

In contrast, the idea that objects present us with a range of possibilities plays no role in DS accounts of intelligent action. According to DS, the reason why I pick up the cup rather than ignore it, or throw it at the wall is because at that moment the cup elicits tension in my brain, which I am pre-disposed to dissipate by picking it up. Perhaps we could say that a full teacup would cause tension in the brain of a thirsty person and would be relieved in this case by them picking up the cup and drinking from it. At no point in the sequence, however, is the cup represented in terms of anticipations or anything else, it is merely the trigger for a response. As Dreyfus says “One does not need to know what the optimum is to move towards it”, and similarly according to DS, we do not need to form anticipations about the cup before we pick it up.

This brings us to the second important difference between DS and AOR, the presence of goals. As mentioned in the previous two sections, on the RTM view, goal states are what drives intelligent action; representations are processed so that the best course of action to achieve a goal state can be realised. In contrast, DS rejects the idea that intelligent interaction is caused by the pursuit of conscious or unconscious goals. Here AOR is more similar to the former than the latter, because AOR shares with RTM the idea that information processing is involved in intelligent action, it also has a shared reliance on goals. In particular, goals would seem to be required when explaining why one affordance of an object was taken up and not another: the reason for my choosing to pick up and cup and bring it to my lips and not pick it up and throw it at the wall would be because I have the goal state of quenching my thirst and only the first possible action allows me to achieve this.

### **Understanding and AORs**

Because AOR accounts for intelligent action in representational terms, the theory's account of understanding could be thought to bear some similarity to the RTM approach: whether I understand something or not is a matter of whether I have the appropriate (anticipatory) representations. This would allow the AOR approach to share another characteristic with RTM: understanding can be identified with a particular neural configuration. Intelligent action is caused by anticipatory knowledge, and this information is not present in the environment but is instead represented in the brain. If this were true, on the AOR account, the brain might in fact be “explanatorily privileged” over the environment.

However, an AOR conception of understanding is not quite as clear cut as this. Firstly it is worth emphasising that supporters of AOR concede quite a lot to the DS approach: AORs are only employed *as required*, with a great deal of intelligent behaviour being ceded to representation-less, goal-less, dynamic coupling. Clark and Grush argue that AORs are employed only when "using real world feedback is impractical, absent, or when it “could improve real-world, real-time responsiveness" (Clark and

Grush, 8). In the example of the chemical plant, interaction with the real chemical mix is possible but impractical because feedback from the system is too “laggardly”.

The fact that AOR provides two possible ways that intelligent action can take place, raises questions about how understanding would be defined by the theory. Should we, for example, attribute understanding to someone who can dynamically couple with an object but is unable to form an appropriate AOR if it is required? This would mean that as long as the object in question is present and able to provide practical real-time feedback, intelligent interaction is possible. If however, the situation becomes “representation hungry” (Clark 1998, 147), intelligent interaction ceases. To get a clear answer here we need to examine how the ability to dynamically couple with an object relates to the ability to represent it in terms of affordances. I will postpone such a discussion until Chapter 5 where I will present an AOR conception of attention understanding.

## **Conclusions**

In this chapter I have summarised three theories of how the mind functions and shown how understanding would be accounted for by each. We have seen that DS and RTM can be thought of as opposite ends of a continuum, with the tools that RTM uses to explain intelligent action -representation, description and problem solving- being explicitly rejected by DS and replaced by complex response to stimuli and optimal gestalts. AOR can be thought of as a halfway position, embracing representation as a means of explaining action in some cases and dynamic coupling as a means in others.

The aim of this thesis is to determine which theory best explains our understanding that other people are attending beings. Having provided an overview of different ways that understanding could be understood, I will turn in the next chapter empirical research into how infants come to understand attention in others. We shall see that there is a split amongst philosophers and psychologists as to what attention understanding consists of, and that this has parallels with the divide between representational and non-representational approaches to cognition. Once the empirical data, and the debates which surround it, have been established, I will go on in Chapters 3, 4 and 5 to examine how well they are accounted for by RTM, DS, and AOR.

## Chapter 2 Understanding Attention

### Introduction

In the previous chapter I introduced three accounts of cognition and showed how understanding would be defined by each. The aim of this thesis is to determine which of these theories best explains our understanding of attention in other people. What I mean by an “understanding of attention” is simply the capacity to appreciate that other people can perceive objects in the environment: when I see you looking at a computer screen I understand that you are attending to what is on that screen, if you direct your eyes towards me I understand that I am the object of your attention, if you look out of the window I understand that you are paying attention to whatever it is that is out there.

A principle reason for focusing on the understanding of attention rather than understanding of anything else is that it allows the theoretical issues discussed in the previous chapter to be brought down to earth and applied to current research in developmental psychology. We saw that understanding is linked to the ability to interact intelligently with the environment, it therefore seems appropriate to apply questions of representation and understanding to actual empirical data on intelligent behaviour, rather than discussing them only in the abstract. There is a wealth of empirical research on the development of behaviours which indicate an understanding of attention and, as we shall see, disagreement amongst philosophers and psychologists interested in these studies over whether such behaviours are best explained by representational or non-representational theories.

I focus on the behaviours associated with attention understanding, rather than any other empirically researched set of behaviours, because they are good example of understanding that emerges very early in life, and yet appears fairly sophisticated. We shall see in this chapter that before infants are one year old they begin to display behaviours which I will, tentatively<sup>2</sup>, refer to as Attention Understanding Behaviours (AUBs). These include: responding when attention is paid to the self, for example returning a caregiver's smile; following the gaze of another when they turn their head or eyes; and directing another's attention by pointing to items of desire or interest. Behaviours like these suggest that infant has some understanding of attention in others early in life and, importantly, prior to language development.

Focusing on a pre-linguistic understanding is useful because once language enters the picture the boundaries between representational and non-representational become harder to define. We have

---

<sup>2</sup> As we shall see in 2.1, there are those who argue that only some of these behaviours display a true understanding of attention. However, 'Behaviour Suggestive of an Understanding of Attention' is rather clumsy, so 'Attention Understanding Behaviour' will be used instead.

already seen in 1.1 that some variants of RTM conceive of representations to be language like, some such as Pinker and Fodor, even go as far as describing representational cognition in terms of a “language of thought” (Fodor 1975; Pinker 1999, Chapter 2). Moreover, even anti-representationalists such as Gallagher accept that representational thought is possible once language emerges (Gallagher and Hutto 2008; Hutto 2008). Language then, causes a blurring of the line between representational and non-representational cognition. As my aim here is to determine the role of representation in intelligent action, rather than linguistic understanding, it will be useful to avoid having to differentiate between language and representation, or to examine the relationship between the two.

A secondary reason for connecting attention and understanding is that I hope applying questions of representation and action to the data on attention will also serve to illuminate issues within developmental psychology. In this field, understanding that others can attend is considered just one aspect of an infant's “theory of mind”<sup>3</sup>. As well as attention, cognitive scientists are also interested in how infants understand that other people have beliefs, desire, emotions etc. In the last 20 years or so, ideas about theory of mind and how it develops, have followed a similar path to that of the theories of cognition detailed in the previous chapter. At the end of the 1980s and beginning of the 1990s, infants' understanding of other minds was considered to be caused by either the ability to simulate or theorise what was going on in another's mind; we shall see in the next chapter that both of these positions can be considered representational. However, in recent years there has been a growing interest in 'Intersubjectivist' accounts of the mind, which reject the idea that understanding others is grounded in representations. This theory will be examined in more detail in Chapter 5.

The question of representation's role in intelligent is also present specifically in research into attention understanding. Amongst philosophers and cognitive scientists who are interested in attention there is disagreement as to whether behaviours such as pointing or gaze following are caused by representational or non-representational cognition. Proponents of what is known as a 'rich' conception of attention, argue that some or all behaviours associated with attention are caused by an “understanding of the concept of attention” (Eilan 2005, 5). We shall see in this chapter that this “concept” can be thought a variety of descriptive representation and so rich theories can be affiliated with the RTM account of understanding. Those who think that attention understanding is 'lean' deny that there is any such concept and so can be linked the DS account.

Despite these divisions, the theoretical claims of RTM, DS and AOR have never been applied explicitly to attention understanding. We have seen in DS and RTM two very different ways that action can be explained and understanding defined, so it will be instructive to examine in detail whether either of them can fully account for the empirical data. Furthermore, the question of whether AORs could be useful in our understanding of attention understanding has, as far as I am aware, never been raised.

---

3 This term is used with various degrees of specificity. It is sometimes used to describe only representational theories of mind such as Theory Theory or Simulation Theory and not non-representational theories such as Intersubjectivity. I intend it in its broadest, most general use: the capacity that adults have and infants develop that allows for intelligent interaction with other people.

This seems to me to be an omission: there is a good deal of discussion in the field of cognitive science as to the possible role AORs might play in intelligent action and I see no reason why this should not be expanded to include intelligent interactions with other people also.

In this chapter I will present an overview into the development of AUBs in the first year of life, summarise the rich/lean controversy and show that these issues can be related to representationalist and anti-representationalist views of cognition. In 2.1 I will introduce the various AUBs displayed by infants in the first year before going on in 2.2 to summarise Tomasello and Eilan's claim that infants do not possess an “understanding of the concept of attention” until they are 12 months old. In 2.3 I will present Reddy's counter arguments to this claim. 2.4 will make explicit the connections between rich and lean theories of attention and representational and anti-representational theories of cognition. All of this will lay the groundwork for Chapters 3, 4 and 5 in which each theory of understanding will be assessed as to how well it accounts for particular aspects of development.

**2.1: The Development of Attention Understanding Behaviours**

There are various behaviours displayed by infants in the first year of life that might prompt us to say that they understand attention: Infants as young as two months smile more when they are being looked at by another person, at 4 months they have been shown to follow the direction of another's gaze, as they approach one year old, they begin to point and follow the points of other people. The development of these AUBs, is charted in Table 1.

Table 1: The development of Attention Understanding Behaviours in the first year of life

Age	Responsive Behaviour	Directive Behaviour
2-4 months	Smiling more at an adult who is looking at it than one who is not (Hains and Muir,1996)  Apparent coyness such as raising an arm so as to cover the face (Muir and Hains 1999; Reddy 2000)  Attempting to make eye contact	Calling vocalisations so as to draw the attention of a caregiver (Reddy 1999).

	with another but becoming distressed if unable to make or break that eye contact (Brazelton 1986; Hobson and Hobson 2008)	
3-8 months	Following the direction of eyes or head turns (Hood, Willen, Driver 1998)	/
7-10 months	/	Clowning: The repetition of acts that elicit laughter e.g. 'shaking the head vigorously, making odd facial expressions, odd movements of the body, putting odd things such as pyjamas on the head (Reddy, 1991; 2001)  Showing off: The repetition of acts that elicit praise e.g. clapping, waving, or saying a word (Reddy 2001)
10-14 months (Joint Attention Behaviours)	Following the points of others (Liszkowski et al. 2004 )  'Gaze Checking': Looking at both the point and the gaze of an other to verify the correct target has been located (Hobson 2005; Hobson and Hobson, 2008)	Imperative Pointing: Pointing to a desired object (Butterworth and Jarret 1991; Carpenter et al. 1998)  Declarative Pointing: Pointing to an interesting or novel object (Liszkowski et al. 2004; Carpenter et al. 1998)  Pointing (of either type) which is met with indifference is not repeated (Liszkowski et al. 2007)

Loosely based on Reddy (2005)

Table 1 shows how AUBs develop in range and sophistication over the first year of an infant's life. "Responsive" behaviour refers to an infant's reacting to (normally visual) stimuli whereas "Directive" behaviours should be understood as deliberate, self-produced movements by the infant. For example, returning the smile of another, or following their head turn would be considered

responsive, whereas pointing, or showing off are acts made so as to elicit a response from another and so are directive.

The earliest AUBs which emerge are sometimes referred to as “dyadic” interactions in that they involve relations only between the infant and another, with no third object being attended to by either. A 2 - 4 month old infant will, for example, reciprocate the affectionate behaviour of a caregiver, returning a parent's smile for example, and make “calling” noises when it is not being attended to. Such behaviour *suggests* that a 2 month old has at least some understanding of attention in others; it knows when it is or is not being attended to and responds accordingly.

At around three months, a primitive form of what is known as “gaze following” emerges: if the infant witnesses another turn their head in one direction then the infant is statistically more likely to turn its head in the same direction, rather than the opposite. Although we shall see in a moment that such a claim is not uncontroversial, such a behaviour at least hints that the infant is aware that other people can pay attention to objects other than the self, and so can be thought of as the first “triadic” behaviour displayed by the infant.

Between 7 – 10 months the infants understanding of attention to self would appear to become more sophisticated. For example infants show an appreciation that others can attend not just to the self in general, but also to particular acts of the self, such as clapping or shaking the head, as well as particular parts of the body such as a hat or pyjamas on the head (Reddy 2001). Behaviours such as these should be classed as “directive” behaviours because the infant is deliberately drawing another's attention to an act or part of the body.

At about 12 months, AUBs become more sophisticated still: the infant becomes able to direct attention to more distant objects by pointing towards them; it begins to follow the points of others rather than just head or eye redirection; and, on occasion, will look back to the face of the pointer as if to verify that the point has been followed correctly. Behaviours such as these are sometimes referred to as instances of “Joint Attention” because they involve infant and other taking part in “periods of sustained attending together to objects in the environment” (Eilan, 5).

## **2.2 Joint Attention and Understanding**

In the summary just given, some care was taken in describing earlier AUBs. I referred to behaviours such as gaze following or showing off as only “suggestive of”, or “hinting at” attention understanding, rather than demonstrating or showing it. This because there are some philosophers and psychologists who claim that it is only when an infant is able to participate in Joint Attention interactions that they can be truly said to understand attention in others (see for example Eilan 2005, Liszkowski 2005;

Tomasello et al. 2007). Interactions involving pointing and gaze checking are said to be the first in which the infant is “aware of the object or event as the focus of the other person’s attention” (Hobson 2005, 185).

The claim is then, that joint attention behaviours are the first AUBs which are driven by the infant realising that the other is a mental agent of some sort, capable of directing their attention towards aspects of the environment. Proponents of this view would argue that all earlier AUBs come about through a mix of automatic response to stimuli and what might be called a 'causal understanding', rather than an awareness that other people are beings capable of attending or perceiving. This claim can be made more explicit by turning to a list provided by Eilan (5) of the four conditions required for a "joint attention triangle":

- a. There is an object that each subject is attending to, where this implies (i) a causal connection between the object and each subject, and (ii) awareness of the object by each subject.
- b. There is a causal connection of some kind between the two subjects’ acts of attending to the object.
- c. The two subjects’ experiences exploit their understanding of the concept of attention.
- d. Each subject is aware, in some sense, of the object as an object that is present to both subjects.

(Eilan 2005, 5)

Supporters of the idea that real understanding emerges only with joint attention behaviours would argue that instances of pointing or gaze checking meet all four of these conditions, whereas behaviour displayed prior to nine months consists of only a and b. Thus responsive behaviours prior to one year are claimed to demonstrate only unreflective reactions to visual stimuli, which could be either an innate capacity or the result of positive re-enforcement. For example a 3 month old who turns in the same direction as another's head turn might be driven by an innate tendency to react that way, similar to a propensity to be drawn to movement or brightness. Alternatively, it might, as Corkum and Moore contest (1998), be the product of positive re-enforcement whereby the infant learns that following head turns is often rewarded by seeing interesting things.

Directive behaviours on the other hand could be explained in terms of causal understanding: It could be argued that making calling noises at 2 months, or clowning at 7, are merely attempts to cause a particular effect, without any understanding of the other as an attention capable, minded person. Put another way, if an infant has only a causal understanding of other people it considers them simply as 'part of the furniture', objects that will respond in a particular way if a particular bodily movement is made. A 6 month old understands that if a held object is let go of, it will fall to the floor and if it cries it is likely to be fed. Similarly, a 2 month old's "calling" to its caregivers may show that it understands a particular action (the calling) is likely to bring about a particular observable effect (the perception of a face), but does not necessarily show that it appreciates that calling will lead to it being *attended to*.

Why then is there the claim that it is only when Joint Attention behaviours emerge that infants go beyond causal understanding and "exploit their understanding of the concept of attention"? One reason is that JAB behaviours show a sophistication that is arguably missing in the earlier months. We have just seen that infants of about 12 months will follow another's point and then look back at the direction of the pointer's gaze. It could be argued that by turning to the eyes after seeing the point, the infant is showing an awareness that the two are connected, an awareness that both point and gaze direction are determined by the other attending to something. It is not obvious how such a behaviour can be accounted for just in terms of causal understanding: the infant's act of looking back at the pointer looks less like an attempt to cause a particular effect and much more like it is checking the relation between the gaze direction, pointing direction and target object.

A directive behaviour used to support the idea of a breakthrough at 12 months is pointing. At 12 months, infants begin to point at objects that they want (so called 'imperative' pointing) and also at novel or interesting objects ('declarative' pointing). Both of these are purported to demonstrate the claim that infants at 12 months are able to take part in "periods of sustained attending together to objects in the environment" (Eilan 2005, 5) and so show that an infants understanding of other minds has progressed from the merely causal. However, it could be argued that imperative pointing can in fact be explained purely in terms of causal understanding: the infant learns the rule that 'If I raise my arm towards an object, then I am more likely to get it'.

The case for a mental understanding of others is more convincing if we consider declarative pointing instead. Tomasello claims that directive behaviour of this type demonstrates that infants are "attempting in their pre-linguistic communication to influence the intentional/mental states of others" (2007, 4). We can see this in the following examples:

As dad prepares to leave, J points to the door (J is 11 months, 18 days old)

J watches quietly as Dad arranges the Christmas tree. When Grandpa enters the room J points to the tree and says 'Oh'. (J is 13 months, 3 days)

When A hears an airplane from in the house, he points through the window to the sky (the airplane is not visible) (A is 12 months, 10 days old).

(46)

All of these examples are said to resist a 'causal understanding' explanation because it is unclear what effect the infant would be trying to cause. In none of the examples is there an obvious target object or state of affairs that the infant is trying to obtain. It does not seem correct to say that, in its pointing, the infant is trying to get its hands on the door, or Christmas tree, or the unseen plane and nor is it clear that the infant is trying to compel others to do something with the objects. Instead, the pointing in these examples is "declarative" in that the infant appears to be either demonstrating its knowledge (that there is a plane in the sky or that Dad is about to go through the door), or directing another's attention towards a novel, interesting object (the Christmas tree). However, we should be wary of saying that the infant is not trying to achieve anything whatsoever. Although a declarative point is not done to obtain something, or direct someone, it is still done for a reason as, presumably, the infant would not be pointing at interesting objects if there wasn't a person there to respond in some way.

One explanation might be that the infant is demonstrating its knowledge that there is a plane, or that Dad is about to leave, so as to elicit praise from those observing. This would fit with claims made by Trevarthen (1978) that infants have an innate motivation to share their experiences with others and also Tomasello who believes that infants' possess "social intentions" such as the desire to share feelings or information (2007, 17). These issues will be returned to when we look at clowning and showing off behaviours in the next section.

### **2.3 Against A Breakthrough**

We have just seen that supporters of 'joint attention breakthrough' at 12 months argue that the behaviours displayed at this age demonstrate an appreciation of other people that goes beyond mere causal understanding. Instead they show a recognition that objects in the world which can be attended to by the self, can also be attended to by other people. However, these claims have not gone unchallenged. Reddy in particular argues that much younger infants have an understanding of attention in others that goes beyond response and causal understanding. I will here focus on two sets of evidence she cites to back up her claim: the first concerns how 2 - 4 month-olds react when they are the object of another's attention, which Reddy argues shows more than mere response to stimuli; the second regards clowning and showing off behaviours, which puts pressure on the claim that "periods of sustained attending together to objects in the environment" only arrive at one year old.

## **Recognising the Self as an Object of Attention**

Reddy (2005) argues that the dyadic interactions that the infant takes part in at two months are not consistent enough to be the result of mere response to stimuli:

Interest in others' attention by two to three months...is not a constant. It is not an automatic response even in alert and happy infants. It fluctuates predictably with mood and with other factors. Further, there is no single response to the attention of others even when there is interest; the responses can range from: intense smiles, to sober gaze, to shy or coy ambivalent smiles, to distressed avoidance. It would seem, therefore, that the argument that these emotional reactions are 'merely' biological, or are so innately specified that there is no room for agency and awareness on the infant's part, cannot be maintained. that the interactions between infant and other (as opposed to the infant other object interactions of JAB)are more than mere causal understanding or response to stimuli and in fact are the result of a primitive understanding of attention.

(97)

According to Reddy, if AUB behaviour displayed at two months is the result of stimuli response then it could be induced in a reasonably uniform manner, but this is not the case. To take a (very) simple example, we might ask why the elicitation of smiling in an infant does not have more in common with a 'knee jerk response': every time an area below the knee is tapped, the lower leg jerks. Smiling has no such consistency, that there is no one pattern of stimuli that causes smiling behaviour, stimuli that causes smiling in one instance, might not do so the next. Neither is there only one response to being the object of another's attention; as we have seen, the attention of another can elicit, affection, distress or shyness.

## **Showing off and Clowning**

A further difficulty for claims of a breakthrough at 12 months is the emergence of "showing off" and "clowning" behaviours at seven months. At this age infants are able to direct attention not just towards the self in general, but also to acts by the self (clapping hands, shaking head) and to individual parts of the body (a hat placed on the head, a "revealed tummy")(Reddy 2005, 100). This is damaging because it blurs the line between (mere) understanding of attention to the self, which supporters of a Joint Attention breakthrough would argue is mere response, and understanding attention directed towards other objects, which supporters of a breakthrough argue does not emerge until 12 months. It is unclear

why the infant's drawing attention to its exposed tummy should be considered a dyadic and not a triadic interaction. Although the infant only directs attention towards objects that are part of, or in contact with, the self, and not to clearly separate objects, it still suggests an appreciation of attention more sophisticated than straightforward dyadic interactions. There is still a distinction to be made here, at one year the infant is able to cope with objects that are distant whereas at seven months only objects on the head, in the hand (2005, 102), or part of the body are able to be appreciated fully. However, the claim that triadic interactions (and attention along with them) only emerge at 12 months would appear not to stand up.

Furthermore, it is not entirely obvious that directive behaviours such as these should be considered clean cut examples of causal understanding. We saw at the end of the previous section that one reason to think that declarative pointing demonstrates a true understanding of attention is because it is not made with solely with the intention of causing a particular effect like obtaining an object. However we also saw that declarative pointing would still be performed with a "social intention", most likely receiving some type of response from an observer. A similar explanation could be given for clowning or pointing. Presumably an infant directs attention to a hat on its head not to cause an effect in the world (it already has the hat so it cannot be gesturing so as to obtain it) but to elicit a response, perhaps praise or laughter, from the observer. If this is the case, then clowning or showing off exhibit just as much of a mental understanding as declarative pointing.

## **2.4 Rich Verses Lean Attention Understanding**

We are faced then, with a dilemma: Tomasello et al. claim that the joint attention behaviours displayed at 12 months cannot be explained in terms of stimuli response and causal understanding and instead require an appreciation that the other person is a mental agent of some type. Reddy, on the other hand, argues that it is not possible to make a clear distinction between joint attention behaviours and the AUBs that precede them: the features used to distinguish joint attention from other AUBs -sophistication, understanding beyond the causal, and triadic interactions- are in fact present in earlier behaviours also. There are two possible paths to be taken here. One is to maintain that an infant possesses a concept of attention which is the cause of intelligent behaviour, but claim that it appears earlier in life, or at birth and so is responsible for more, or all, AUBs displayed in the first year. The other is to deny that there is such a thing as a concept of attention, and argue that all AUBs, including joint attention behaviours are the product of causal understanding and stimulus response.

The claim that there is an underlying understanding, or concept, of attention that causes at least some AUBs is known as a 'rich' theory of attention (see for example Liszkowski 2005; Tomasello et al. 2007 ). The advantage of this approach is that, now the idea of breakthrough at 12 months has been disproved, any AUB performed in the first year of life that appears to exhibit more than a causal

understanding, can be accounted for by its possessing a concept of attention. This would mean that when an infant returns another's smile, or directs attention to a part of the body we are not forced to explain this as the product of a response or causal understanding only.

One obligation of this approach would be to make clear what a “concept of attention” consists of and what role it plays in the performance of intelligent action. How, for example, does possession of the concept cause the infant to follow gaze, or begin to point? A second would be to show the relation between a concept of attention and the development of AUBs. As we have seen, AUBs do develop in sophistication even if we disregard claims of a breakthrough at 12 months. What role, for example, would a concept of attention have in explaining why infants follow gaze at 3 months only sporadically, but check the gaze of a pointer at 12?

The other option is to push from the opposite direction and argue that innate or learned stimulus response is responsible for all AUBs up to and including joint attention behaviours. To deny that there is any single idea, concept, or representation acquired (at least before the infant is linguistically competent) that counts as an understanding of attention is to endorse a 'lean' theory of attention understanding (see Moore and D'Entremont 2001; Muller and Carpendale 2004; Racine and Carpendale 2007). An advantage of this approach is by grounding intelligent interaction in terms of response and causal understanding, it avoids the obligations of a rich approach; there is no need to account for what a concept of attention consists of, or how it relates to action or development.

The main challenge for a lean theorist is to show that all AUBs, can be reduced to stimuli response and causal understanding. At first this might seem like quite a tall order as we have already seen that behaviours such as gaze checking would appear to clearly demonstrate an understanding beyond the causal. Nevertheless, we shall see in Chapter 4 how DS approaches might avoid this difficulty.

## **Conclusion: A Return to Theories of Understanding**

Parallels can be drawn between the debate over rich and lean conceptions of attention and the RTM and DS accounts of cognition introduced in the last chapter. Rich accounts of attention understanding can be considered a variety of the RTM approach: the "concept of attention" proposed by Eilan is thought of as a information bearing state of the brain which is the cause of intelligent behaviour, which it makes it very similar to the descriptive representations that RTM postulates to explain action. If we return to the list of representational characteristics presented at the beginning of Chapter 1, we can see that all a concept of attention meets all four: the concept is internal as presumably rich theorists would identify it with a particular neural configuration (RP1); it is elicited by a feature of the world and lasts

for a discreet duration, a concept of attention would only come into play when dealing with an object that possesses attention (RP4 and RP2) The infant's "awareness of the object as present to the other" (d) can be thought of as the infant representing *that* the other person as attending to the object (RP3).

In the next Chapter I will assess how well RTM meets the obligations of a rich approach to attention. After examining two different ways that a descriptive representation of attention might be understood, I will go on to focus on two issues: whether a descriptive representation of attention provides a clear account of how responsive and directive AUBs are achieved, and whether development can be explained in representational terms.

Lean theories of attention and DS approaches to cognition are united in their denial that there is any concept or representation involved in intelligent action. As we saw in Chapter 1, DS theorists deny that understanding can be identified with descriptive information represented in the mind and argue that environment, body and brain are equal partners in causing intelligent behaviour. An advantage of this approach is that the DS notions of 'tuned brain states' and moving towards an "optimal gestalt" can be utilised to explain complex behaviours such as gaze checking, which on the face it defy an explanation in terms of response and causal understanding. In Chapter 4, I will assess how successful DS is in providing a lean account of AUBs and their development. I will focus in particular on whether complex behaviours can be explained in terms of starting states and whether such states can truly be called non-representational.

As mentioned in the introduction to this chapter, the notion of AORs is not something that has been examined in the context of Theory of Mind, or the development of attention understanding. As it is positioned between representational and non-representational accounts of cognition, so it is between rich and lean conceptions of attention understanding. An AOR conception would be rich in that a concept of attention would be cited as the cause of some AUBs, but lean in that other behaviours are explained entirely in terms of response to the environment and causal understanding. In Chapter 5 I will examine whether this halfway position stands up when applied to the empirical evidence.

## Chapter 3 Attention and the Representational Theory of Mind

### Introduction

In this chapter I will examine how successfully RTM explains AUBs in the first year of life. In 3.1 I will present a first pass at how responsive and directive behaviours would be accounted for in terms of the RTM framework presented in the previous chapter. In 3.2 I will add detail to this picture by exploring two ways that descriptive information about attention can be encoded. 3.3 will assess how well the theory copes with two specific areas of AUB development: inconsistent gaze following displayed at 3 months and the development of directive behaviours between seven and ten months. We shall see that to account for development, RTM has to rely at least in part on the infant's motor skills, in 3.4 I will examine what effect this has on the theory.

### 3.1 An RTM Account of AUBs

**Responsive Behaviour:** We saw in the previous chapter how RTM would explain responsive behaviour made in relation to an oncoming bus. The object possesses certain perceptible features (being red, bus shaped, having wheels) and from these we infer that it possess certain properties such as being heavy and being able to cause harm if it collides with us at speed. Such calculations allow the person in question to respond to the bus in an intelligent manner. An analogous story can be told for an infant's responsive AUBs: particular features or movements of a person, lead the infant to infer that it has the property of attention which causes the infant to respond to it intelligently.

We can see how this would work if we turn to an AUB such as gaze following. The infant learns that when an object possesses certain perceptible features, like eyes or a face, it should be represented as having the property 'attention'. This representation of attention would include the information that attention 'comes out of the eyes' and so the infant is able to infer that a redirection of eyes is also a redirection of attention. As well as this, it would contain information along the lines of 'people attend to interesting, desirable, or novel things'. The final ingredient here is that the infant has a goal state, it is motivated to see interesting and desirable things. Therefore, when the infant sees a head turn it infers from the stimulus and its representation of attention that here is an opportunity to achieve a goal state, and so turns in the same direction.

**Directive Behaviour:** We can also see how this style of explanation can be used to account for directive behaviours such as pointing. As before, an infant would infer from features of an object to represent that object as having attention but this information would be coupled with the knowledge that pointing leads to a redirection of gaze, and therefore attention. Again the infant's goals are an important factor, the other's attention will be redirected so that a particular goal state can be achieved. As discussed in 2.2. and 2.3 these goal states could either be simply trying to cause an effect in the world (such as obtaining an object) or a social intention (such as sharing feelings or information).

### **3.2 Two Ways to Think of Descriptive Representations**

To better determine what role a descriptive representation of attention plays in earlier AUBs, it will be useful to examine exactly what information would be contained in an attention representation and how this information is encoded. In 1.1, we saw that there are different varieties of RTM with different conceptions of how descriptive representations should be conceived (as images, a “language of thought” etc.), but so far we have not examined which conception would be most suitable for understanding attention. I will address this issue by turning to two theories of 'mindreading' popular in contemporary cognitive science: Theory Theory and Simulation Theory. These competing approaches have been put forward as explanations of how we understand *all* aspects of others' mental lives with attention being just one type of mental state among many; we understand that others attend, but we also understand that they desire, intend, believe etc. Despite differences in exactly what information is represented, we will see that both theories can be thought of as variants of the RTM approach to cognition and therefore candidates for a rich account of attention understanding. Both theories envisage descriptive representations as being the cause of intelligent action and as we shall see in 3.3, there are also similarities as to how each theory accounts for the development of other understanding.

#### **Theory Theory**

Theory Theory (hereafter TT) is the view that our understanding of the minds of others chiefly involves the application of a theory (see for example, Baron-Cohen 1997; Carruthers 1996; Gopnik & Meltzoff 1997)<sup>4</sup>. Carruthers summarises this view as follows:

---

<sup>4</sup>As we shall see there is a split amongst Theory Theorist as to whether infants gain theoretical knowledge of others through a process of observation and hypothesis revision (The 'Child Scientist' theory), or from the maturation of innate 'mental modules' (The 'Modular Maturation' theory). However, while these two varieties differ in how theoretical knowledge develops, both conceive of representations in terms of theoretical knowledge. As my primary concern here is to show different ways that the information contained in RTM representations can be encoded, rather than its development, this split in Theory Theory will be left until a little later.

...our understanding of mentalistic notions -of belief, desire, perception, intention and the rest is largely given by the positions those notions occupy within a folk-psychological theory of the structure and functioning of the mind

(1996, 22)

On this view we possess a theoretical knowledge of how mental states interact with one another and how they relate to observable events, for example we have knowledge of:

...the relationship between line of vision, attention, and perception; between perception, background knowledge, and belief; between belief, desire, and intention; and between perception, intention, and action; that one is able to predict and explain the actions of others...

(24)

This understanding is called *folk*-psychological to emphasise its purported closeness to the 'common sense' attribution of mental states that people (folk) make in everyday life. It is deemed theoretical because it involves "unobservable theoretical postulates" (Astington 1996) being used as an explanation for observable behaviour in a similar manner to scientific theories. Gravity, for example, is an unobservable force postulated to explain the observable phenomena like objects falling to the ground. Similarly, the argument goes, we explain somebody's drinking a glass of water as being caused by unobservables such as a desire to quench thirst, and a belief that drinking water will achieve this.

A further similarity between folk psychological and scientific theory use is that they both allow for sophisticated inference. As Leslie and German put it, theories involves "inferencing that is sensitive to the structure of representations" (1995, 128). A scientist's theoretical notion of gravity allows her to calculate exactly the strength of gravitational force an object of a particular mass will create (inferring the properties of the theoretical object 'gravity' from the state of the world). She can then go on to determine what effects that that particular force will have on other objects nearby (inferring the effects that a theoretical object will have on the world).

The theoretical postulates that we attribute to other people are similarly sensitive to the state of the world. We know what 'inner states' follow from particular observable circumstances: If I see your eyes fix on a mouse I infer that you then have an (inner, unobservable) percept of a mouse. Knowing what I do about the relationship between perception and belief formation, I can deduce further that you now have the belief that a mouse is in the house. From here I can calculate what other mental states will follow from this (the feeling of fear, the desire to rid the house of mice) as well as behaviour (startling, putting down traps or poison). Just as the scientist's calculations are sensitive to the exact quantity of

mass, the precise details of what mental states are attributed to another person are sensitive to fine details of the situation: If you see only something that *looks like a mouse*, perhaps it is dark, the mental state attributed will be similarly fuzzy; perhaps I would attribute you with the *suspicion* that there is a mouse in the house, rather than a full-blooded belief.

If we turn to representations of attention specifically, we can see that they would be accounted for in a similar manner. An infant who understands attention would be able to infer what perceptual/attentional states would be formed in another person by the state of the world (the direction of eyes, the presence of an object) as well as knowledge of what other mental states and actions would follow from someone's attending to a particular object.

### **Two Types of Simulation Theory**

An alternative way that mindreading can be accounted for by RTM is through 'Simulation Theory' (ST) (see, for example, Gallese 2005; Goldman 2006; Gordon 1999). As its name suggests, the key difference between Theory Theory and Simulation Theory is the claim that our understanding of others is not governed by implicit theory but by implicit simulation: I observe the circumstances and/or behaviour of another person and put myself in their 'mental shoes' to determine how things would be for me if I was in such circumstance and/or behaving in such a way. I will here exam two varieties of ST: one involving what Goldman calls 'pretend states', the other involving mirroring or resonance mechanisms. Although this latter type is sometimes written of as though it is non-representational, I will argue here that both should be considered varieties of RTM.

### **Pretend State Simulation**

The chief proponent of what I shall call Pretend State Simulation is Alvin Goldman. He argues that when we witness another's behaviour we implicitly imagine how we would be feeling, believing, paying attention to etc. We then make "an introspective identification of [our] own...pretend state and "transfer" it, or project it, onto the target" (2006, 187). When I see you bringing a glass of water to your lips I unconsciously simulate what I would feel, desire and believe if I were doing the same. I then identify these states and attribute them to you and use them to explain or predict your behaviour.

Pretend states can be put in terms of RTM representations fairly easily. They are inferred from visual cues, contain information (that the person has a particular mental state), and are the cause of intelligent action. A pretend state explanation of AUBs would run therefore as follows: an infant who understands attention would be one who observes the attention behaviour of another, like their eyes pointing towards an object, and implicitly simulates/imagines what would be perceived/experienced if their own eyes were pointed at that object from that point of view. The information contained in this pretend state is then used to calculate what the best course of action would be in the current situation, leading to gaze of point following.

## Resonance Simulation

The second variety of mental simulation is on the face of it, less obviously compatible with RTM theories. Although Goldman emphasises the importance of pretend states in some of our interactions with other people, he, and other ST theorists, argue that on other occasions a second "automatic, unconscious, and reflexive" (Gallese 2005, 41) form of simulation comes into play. To varying degrees, Goldman (Chapter 6), Gordon (2008), and in particular Gallese (Gallese et al. 1996; Gallese & Goldman 1998; Gallese 2005) claim that at least some simulation depends instead upon 'resonance' or 'mirror' systems in the brain. I shall refer to this second form of simulation as 'Resonance Simulation'.

Two items of research are employed in support of Resonance Simulation. Firstly, there are various studies describing brain-damaged patients who have lost both the ability to feel a particular emotion and the ability to recognise that same emotion in others. Damasio (1999, 66) for example, describes a subject who after suffering damage to amygdalae could neither feel fear herself, nor identify the same emotion in other people. Similarly, Calder et al. (2000) discovered that a patient with a damaged basal ganglia was unable to identify or express disgust.

The second area of research concerns the discovery in the 1990s of 'mirror neurons' in macaque monkeys (Rizzolatti et al. 1996) and later humans (Gallese et al. 2002) which discharge either when a "goal-related hand action" is performed or when a similar action is observed in another (Gallese 2005). For example, neurons in the F5 area of a macaque's brain have been shown to discharge when the monkey grasps an object and when it observes the object being grasped by another monkey or human. Furthermore, the mirror neurons are not simply sensitive to the visual stimuli of a hand grasping an object but also to the intentions behind the action: a study by Umiltà et al. (2001) shows that mirror neurons are activated even when the target object of an observed grasp has been hidden so that the monkey knows it is there, but cannot see it.

Both Gallese and Goldman *imply* that this type of simulation should also be considered representational; Gallese describes resonance as generating "representational content" which has the function of "modelling objects" (2005, 41) and Goldman considers it a "motor representation of an action performed by others" (2006, 135). However, questions of why Resonance Simulation should be considered representational and whether it is a similar type of representation to Pretend State Simulation are not examined by either writer in any detail. This is understandable as neither are concerned with the representational statues of mirror neurons *per se*; the primary aim of both is instead to persuade that the discharging of mirror neurons should be considered a variety of simulation. Gordon, on the other hand, claims:

...mirroring processes may directly influence our efforts to anticipate and to understand another's behavior; that is, without first issuing in judgements about the other's mental states, without even requiring possession of a repertoire of mental state classifications.

(Gordon 2008, 4)

Such a description would seem to move us away from descriptive representation as it suggests that resonance processes play a causal role in intelligent action, rather than a representational one. If mirroring processes do not involve "judgement" or "a repertoire of mental state classifications" and so "directly influence" our interactions with others, then it is doubtful whether they meet the requirement of RP3 that a representation should bear information about something other than itself. We can see this problem more clearly if we turn to what could be considered a parallel case. The "On" button of a computer plays a key role in its operations but we would not say that the button contains information, or represents anything to the system. Instead, it plays an entirely causal, non-content bearing role. Resonance Simulation could be considered analogous: perceiving fear in another pushes our own 'fear button', perceiving anger pushes our 'anger button'.

Regarding attention understanding however, these concerns should be disregarded. Although there is a case for non-representational Resonance Simulation when it comes to emotion, the same arguments do not apply when we consider mirror neuron involvement in action or attention. Recognising a face as displaying an emotion can be thought of as binary: anger is either recognised in a face, or not; the appropriate neural substrate is either activated or not. Recognising actions however, is more complex; when somebody acts they do so in relation to another object. Furthermore, it has been shown that it is not simply the movement that is being responded to but the intention behind the movement. If the monkey is shown the same movement without any obvious purpose (a hand mimicking a grasping motion but with no target object present) then there is no mirror neuron response (Buccino et al. 2004). What can be drawn from this is that mirror neurons 'appreciate', as it were, that other living things can act *in relation to* other objects; they are not simply detectors for particular bodily movements.

The fact that mirror neurons are sensitive to relations between other and object should lessen the fear that Resonance Simulation is non-representational, at least when it comes to attention understanding. Although there are no studies regarding attention understanding and mirror neurons, if attention understanding were based on resonance/mirror simulation then it would appear to be closer to action recognition than emotion recognition. We do not recognise another as simply being in an attentional state, rather, we recognise them as attending *to* something -be that distal object, proximal object or self. If mirror neurons are able to detect unseen relations between other and object then this type of simulation is in some sense 'about' the world rather than simply a causal response to it. If a mirror response can be about a particular relation *x* then it follows that that response contains information

about  $x$ , and it would therefore appear that Gallese and Goldman are right in thinking of Resonance Simulation as representational.

Although questions as to what features a state of the brain needs to have for it to count as information bearing will be discussed further in the next chapter, there is no reason to think that mirror responses are any less about the state of the world than pretend states or theoretical postulates. Therefore, Resonance Simulation should be considered another variety of RTM.

### **3.3 RTM and the Development of AUBs**

ST and TT are often presented as competing theories of how we understand the minds' of others. However, we have just seen good reasons to think of both as variants of the RTM approach. Both involve internal neural configurations (RP1), which have a discrete duration (RP2), and allow for intelligent behaviour in a particular situation (RP4). These neural configurations also bear information about something other than themselves (RP3), with the chief difference between the two theories being how this information is encoded: In TT descriptive information of another's mental state is represented in terms of theoretical knowledge of how mental states relate to the world and one another. In ST this information consists of pretend states of what would be experienced if the observer was in the position of the observed, or mirror neuron response to a perceived action or emotion.

Now that RTM conceptions of mind have been fleshed out we are in a better position to see how well it accounts for the empirical data introduced in Chapter 2. Here, I will focus on two aspects of attention understanding behaviour, inconsistent gaze checking at 3 months and the development of directive behaviours from between 8 and 12 months.

#### **Inconsistent Gaze Following**

In Chapter 2, we saw that from from three months onwards infants display a primitive form of gaze following. At this age, infants are *statistically more likely* to look at an object if a pair of computerised eyes 'looks' at it first. Such a behaviour is in need of further explanation from RTM because it is not obvious how it fits with the account of responsive behaviours given at the beginning of this chapter. There we saw that representations of attention would be elicited by features of people, such as eyes or faces, and that these representations would cause intelligent pursuit of a goal. This story however, does not explain why 3 month-olds would follow the gaze of others only some of the time; if a proponent of RTM wants to explain the times when the infant does follow the other's gaze as the product of a representation, they need to provide an explanation of what happens in the cases where gaze is not followed.

One possibility might be to argue that this very early AUB is a case of mere stimuli response and is thus not representational at all. We saw in 1.3 that some rich theorists, like Tomasello or Elian, claim that a concept of attention does not appear until about 12 months, but after examining arguments made by Reddy we saw that there was good reason to reject this. However, the question was left open as to whether infants are born with a concept of attention, which therefore causes all AUBs, or whether it emerges at some point before 12 months, and so causes only some of them. A proponent of RTM could argue for the latter option and then go on to say that because inconsistent gaze following appears so early in life, it is not anything to do with a representation of attention.

To take this route, however, would weaken the theory. In 1.3 we saw that there is little reason to say that behaviours emerging at 12 months are qualitatively different from those that arrive earlier: dyadic and triadic interactions display a similar degree of sophistication and directive behaviours at 8 months suggest a similar 'mental understanding' to that of pointing behaviours at 12. If this is the case then any dividing line drawn as to where a 'true' understanding of attention emerges runs the risk of looking arbitrary. RTM would be much stronger if it could account for all AUBs in the first year.

There are three possible strategies that the theory could take in order to explain the instances where a 3 month old does not follow gaze:

- 1.** Deny that infant has the requisite goal state. The infant consistently forms a representation of attention and so consistently recognises head turns as a redirection of attention, but its desire to follow gaze fluctuates. Failure to follow gaze is caused by the absence of this goal state.
- 2.** Deny that a representation is formed. The infant consistently possesses the requisite goal state but its ability to form a representation of attention fluctuates. Failure to follow gaze is caused by an absence of a representation.
- 3.** Deny that the infant has sufficient motor skills. Both goal state and representation formation are constant, but the infant lacks sufficient motor skills to consistently turn its head in the right direction. Failure to follow gaze is caused by an inability to control movement.

### **Inconsistent Goal States**

In 2.3 we saw that Reddy argued against infant-other interactions in the first three months of life as being a case of simple stimulus response on the part of the infant. The infant does not respond to its carer in a uniform manner but instead its behaviour "fluctuates predictably with mood". A similar line of argument might be taken here: gaze following does not happen in response to every head turn, because the infant does not always have the goal of seeing what the other is attending to. However, a closer look at how the experiment is set up appears to rule this out. The trials were only carried out while infants were in a state of "alert inactivity" and the computerised eyes did not turn one way or the other until the infant had fixated upon them. If the infant is alert enough and interested enough to

focus on the eyes, then one would also expect there to be interest in what goals they can be used to achieve.

### **Inconsistent Representation**

Another option would be to argue that redirections of gaze only causes a representation of attention (theoretical or stimulatory) on the occasions that the infant does follow the eyes. In this case the infant would consistently have the goal state of seeing interesting things, but stimuli of a head turn would only sporadically cause the representation required for realising the goal.

However, this strategy is also unsatisfactory because there is no obvious explanation why the same visual stimulus only sometimes calls forth a representation. We have seen that TT theories rely on the infant's knowledge about what mental states will be formed by particular other-involving situations, an ability to infer an unobservable state from an observable state of affairs. If the infant knows (is able to represent) that an eye redirection indicates an attentional redirection then it seems odd that such knowledge would flicker on and off from moment to moment. The equivalent in the gravity analogy would be of a scientist who infers on some occasions that a particular mass will cause a particular gravitational force, but when observing the exact same mass at other times, is unable to make any inferences at all. Such a situation seems nonsensical, if the scientist understands gravity at all he should be able to make the inferences ever single time. ST faces very similar problems. If a pretend state is invoked to explain the times when eye direction is followed, then it is puzzling why the same state should not be elicited every time the eyes change direction. This question is even more pronounced if we turn to Resonance Simulation. Here, mirror response is considered to be "automatic, unconscious, and reflexive" so we would expect to see a simulation produced every time the stimulus is presented.

### **Inconsistent Motor Control**

The most promising strategy available to RTM is to argue that infant who does not follow a redirection of gaze possesses the requisite representation and goal state, but at 3 months lacks sufficient motor skills for consistent head turning. On this account, the 3 month old possesses the general goal of seeing interesting or desirable things. It is also able to consistently represent others as attending beings and therefore recognises all (or almost all) head turns as opportunities for goal fulfilment . What is missing is the full control over bodily movements enjoyed by adults.

In the first three months of an infant's life a great deal of its movements are either responsive: pressure on the infant's palm causes it to grasp; or spontaneous: kicking the legs without any apparent stimulation or goal. (Vereijken 2005, 218; Haywood & Getchell 2008, 93). At the age when inconsistent gaze following emerges, the infant is just beginning to curb its spontaneous and responsive movements, full motor control is still a long way away. We can therefore hypothesise that

the 3 month old who does not follow the gaze, is attempting to realise a goal but fails to move its body in a way necessary to achieve it.

This option appears the most plausible RTM account of failure to follow gaze. Its strength is that it explains inconsistent behaviour by appealing to a characteristic of the infant, its motor skills, that we know to be inconsistent also. Options 1 and 2 fail because both are attempts to explain inconsistent behaviour as the result of inconsistency in unobservable, and at this point hypothetical, mental entities. The more qualifications that need to be added to them so that they fit the empirical data, the less convincing they become. To argue therefore, that inconsistent behaviour is the result of a fluctuating goal, or a fluctuating ability to form the representations appropriate for action, we require a clear explanation as to why there are fluctuations, and it would seem there is none forthcoming. An infant's motor skills, on the other hand, we know to be poor, and so when looking to explain a failure to act would seem a likely candidate. However, as we shall see in 1.4, relying on non-representational factors to account for AUBs threatens the whole RTM project.

### **Development of Directive Behaviours**

We have seen that around 8 months the infant is capable of directing attention to parts of its own body (clowning and showing off) but cannot direct attention to objects in the world (aspects of the environment) through pointing until about 12 months. The challenge this presents to RTM is what role representations play in this progression. We saw in 2.2 different ways that information can be encoded by RTM representations and how it could give rise to intelligent action. What has not yet been examined is the role that descriptive representations would play in the development of AUBs. There are two strategies that RTM could adopt here:

- 1 Development of AUBs is a product of the infant's representation of attention developing.
  
- 2 Development of AUBs is a product of the infant's increasing competently motor skills

### **Development of Representation**

Both Simulatory and Theoretical versions of RTM explain development in terms of information in a representation being revised or added to. For example, the emergence of an “understanding of belief” at about 4 years old is explained by both as being caused by the development of representation. Such an approach is understandable, RTM conceives of descriptive representations as being the primary cause of intelligent action and so accounting for development of behaviour in terms of development of representations as opposed to changes in (for example) motor control, adds explanatory weight to the theory. Before showing why this approach is inadequate for explaining the move from clowning to

pointing, I will show that all varieties of RTM can be thought of as accounting for development in terms of information revision.

### **Theory Theory**

There is a split amongst Theory Theorists differ over how best to explain development. Supporters of what is known as “Child Scientist” (Gopnik & Meltzoff 1997; Gopnik 2004) TT argue that folk-psychological theory shares one more characteristic with scientific theories: the method with which theories are created. Scientific notions such as gravity have been revised in the face of new empirical evidence throughout their history, and so too, according to this view do an infant's ideas of mental states. As the infant grows up in the world, data are gathered through observation, hypotheses are then formed, tested, and subsequently revised if necessary. In the early years, an infant might possess a theoretical understanding of attention, belief or desire that is flawed in some way, leading to less than competent behaviour in response to certain situations. Over time, as more observations of other people are made, these ideas are revised and the infant comes to have an understanding of mental states not unlike that of an adult.

The alternative approach is to argue that an infant's theoretical knowledge of mental states is innate, and that increasing competency in dealing with other people is the result of the maturation of mental modules. Baron-Cohen, for example argues that infants are born with mental “mechanisms” such as an “Intentionality Detector”, which enables an infant to distinguish between movements made by animate and non-animate objects, and an “Eye Direction Detector”, which is fairly self-explanatory. Over time these modules mature, and the interaction between them becomes more sophisticated, leading to increasingly sophisticated interactions between infant and other (See Baron-Cohen 1997, Chapter 4; Scholl & Leslie 1999).

Despite their differences, both Child Scientist and Modular Theory Theories consider development to be the product of information contained in a representation being revised or added to. Both would explain the difference between the behaviour of an 8 month old and that of a 12 month old as a case of the latter having a greater theoretical knowledge of other minds than the former. This would lead to "theoretical postulates" with a richer inferential structure leading to more sophisticated interactions being performed.

### **Simulation Theory**

The ST account of development can also be put in terms of an increase in the amount of the information represented. Two varieties of mental simulation were presented in the previous section, and development can be explained in a similar way by both. Turning first to Pretend State Simulation, on this view infants are able to simulate how it is for another in situation X only if they themselves have experienced situation X. This can be seen in Goldman's comments on research presented by

Meltzoff and Brooks (2003) in which 12 month-olds were tested to see if they followed the 'gaze' of people who were blindfolded. The group of infants who had had first-hand experience with blindfolds (they had been blindfolded themselves) did not follow the head turns of blindfolded people, whereas a group with no experience of being blindfolded did. Goldman's ST interpretation of this result is as follows:

When a “trained” infant observes an open-eyed, nonblindfolded person turn and gaze in a certain direction, the infant can simulate that person as having some visual experience of an object in the direction of gaze and attending to the object. So the infant turns its own head to obtain a visual experience of the object. When the infant observes a closed-eyed person or a blindfolded person turn her head, the infant can simulate the model as having no visual experience (or black visual experience). So the infant does not follow the model’s head turning.

(Goldman, 194-195)

This suggests that an infant's understanding of attention depends upon its having already had the relevant attention experiences itself. In the first year of life the infant experiences itself paying attention to aspects of the world and also the actions that these attentive states cause; as these experiences accumulate, the infant is able to project them onto others leading to increasingly competent interactions.

Research into mirror neurons is at a fairly early stage and so it is difficult to say with certainty if and how Resonance Simulation develops. One possibility suggested by Goldman (142 – 144) and Catmur et al. (2007) is that mirror neurons can 'learn' to respond to a particular visual stimuli displayed by another. Over time, the infant observes its own actions and comes to associate a particular movement with particular stimuli. Goldman thinks that this association also takes place on the neural level:

Action potentials in STS neurons responding to the sight of this type of grasping overlap in time with activity in the PF and F5 neurons that cause the infant to grasp in that way, for example, precision grip. This creates the prerequisites for Hebbian associations: Neurons that fire together, wire together.

Once these two sets of neurons are suitably "wired together", they will fire not just when the infant observes its own precision grip, but also when it observes that same action being performed by another. This allows for Resonance Simulation:

After “learning” the association between F5 motor commands and visual STS descriptions of the monkey’s own actions, the observation of someone else performing a similar action will also activate the neurons in F5, and mirror properties will have emerged.

(143)

If it is correct that mirror response can be learned in this way, the development of Resonance Simulation is similar to that of Pretend State Simulation. We saw that the infants in the blindfold example were unable to simulate the experience of another until they had had the relevant experience themselves. Similarly, in this case the infant must be able to perform an action itself, thereby allowing the neurons involved in moving the body and seeing the action to “wire together”, before it can achieve a mirror response when observing the action performed by others. We can see then that both versions of ST can also be thought of in terms of information addition. Whereas TT representations develop by the addition of theoretical knowledge and the accompanying inferential structure, ST representations develop as the infant gains experience of the world and is therefore able to simulate a greater number of mental states.

### **Information Revision Fails to Explain the Development of AUBs**

We have now established that all varieties of TT and ST account for the development of intelligent behaviour in terms of representations becoming more information rich. We shall see now however, that such an approach fails to explain the progression from clowning to pointing. If this development is to be accounted for in terms of information revision then we should be able to pinpoint exactly what information is acquired so as to cause the change, but this is problematic.

TT couches development in terms of the inferential structure of a representation being altered or added to, but the infant's representation at 8 months would appear to already have sufficient inferential structure for pointing behaviour to emerge. By this age infants know to some degree that other people are capable of attending to distal objects, as we have seen, gaze following is already possible at this age. They also know that the attention of others can be directed, as demonstrated by their clowning and showing off behaviours. If the 8 month old's representation of attention already contains these two items of theoretical knowledge we might ask what more needs to be added for pointing to take place. We can see this if we return once again to the gravity analogy; the equivalent here would be of a scientist who is able to infer that proximal objects can be affected by a particular gravitational force, but unable to see that distal objects will be affected as well.

The problem remains if theoretical information is switched for stimulatory. At 8 months the infant is able to simulate other minds competently enough to allow for gaze following and directing attention to aspects of the self. If the explanations as to how mental simulation develops are correct, then these

abilities depend upon the infant having had the relevant experiences itself: it knows that if it turns its own head, it will receive different perceptual information; it has had experience of directing its own perception at proximal objects, or parts of others' bodies. Again though, if the 8 month-old has already had the experiences necessary for gaze following and directing towards proximal objects, it is unclear what else needs to be added (what further experience is necessary) to make pointing behaviours possible. The fact that an 8 month old can gaze follow would show that it is able to simulate others paying attention to distal objects, the fact that it can draw attention to proximal objects shows it has had experience of its own attention being directed. Once again we have a case where it would seem that the infant's representation of attention contains all the descriptive information required for pointing behaviours, but does not perform the action.

### **Development of Motor Skills**

If development cannot be accounted in terms of information revision then it would appear that RTM is best chance of explaining development is by once again appealing to the infants' motor skills. We have seen that the most plausible way to explain inconsistent gaze following under RTM is to argue that the infant's representation of attention, and its goal to see an interesting object, remained constant across cases where gaze was and was not followed. Whether gaze was followed or not depended on whether the infant was able to control its bodily movements with sufficient competence. Similarly, we might argue that the information contained in the representation of an 8 month old is the same as that contained by a 12 month old, but that the older infant has a greater degree of motor control and so is able to perform more sophisticated behaviours. In other words, at 8 months there is sufficient motor competence for the infant to be able to consistently move so as to draw attention to proximal objects, but not yet enough for it to be able to produce pointing actions.

It might be asked why the infant gains the motor skills involved in directing attention to parts of the self before the motor skills involved in directing attention to distal objects. Why do clowning behaviours always precede joint attention behaviours? A response to this might be to compare directive behaviours with the learning of other types of practical skill. In a similar way to the adage that you can't run before you can walk, when we observe the development of practical skills we can see that certain abilities necessarily precede others. Babies start to crawl at around 6 months, and do not start to walk until after 12, playing a two finger chord on the guitar is a necessary step before being able to play more complex four finger chords. We might argue that AUBs progress in a similar manner and say that the ability to point at 12 months is dependent on the earlier ability to gesture towards proximal objects.

Comparing AUBs with skills on a musical instrument also provides a clear example of motor skills being solely responsible for development without changes being made to the content of a

representation. It does not seem plausible that the development of guitar skills depends on my representation of the guitar being altered; I may know very clearly where to place my fingers on a guitar to form a four finger chord, but still not be able to do it. Similarly, an infant might know what movements are required to point at an object, but be unable to make them.

### **3.4 Motor Skills and Representations**

We can now see that if RTM it is to have a chance of accounting for the development of AUBs, it must rely to a large extent on the development of infants' motor skills. Such an approach however, causes the theory serious difficulties.

First of all, it puts pressure on the claim that the main cause of intelligent action is the processing of descriptive representations in the brain. The three theories of understanding introduced in Chapter 1 are put forward as ways of explaining how intelligent interaction is possible and it would seem that one requirement of a successful account would be to explain how intelligent action develops. We have just seen, however, that when it comes to the development of gaze following or directive behaviours, the RTM toolbox of descriptive information, inference, and goal states is of little use. That the theory has so little to say about how AUBs like these develop, certainly weakens the claim that intelligent action should be accounted for in RTM in terms. Put another way, by having to rely on motor skills, RTM moves a step closer to the DS claim that the brain should not be explanatorily privileged over, body or environment.

A weaker position that RTM might adopt in the face of these criticisms is to accept that motor skills are responsible for development in the two areas I have examined, but that other instances of AUB development are caused by developments in representational ability. For this claim to hold water, however, RTM would need to point out an example of development that cannot be explained in terms of motor skills and therefore requires an information based explanation, but this does not appear easy. We saw in our discussion of Reddy's arguments in **2.3** that is difficult to find any qualitative difference between AUBs either at 12 months or any other stage in the first year. If this is the case then there is no reason why motor skills explanations cannot be invoked to explain, for example, the progression in the first three months of life from infants simply responding to the attentions of another to (sometimes) following their gaze..

If these arguments hold then RTM is forced to adopt the even weaker position of saying that infants are born with a fully fledged representation of attention (already containing the information that others can direct their attention towards self, proximal, and distal objects) but are unable to

demonstrate it due to a lack of bodily co-ordination. With this, however, little of the theory's explanatory value remains. Aside from the fact that it seems intuitively unappealing to think that infants are born immediately and completely understanding that others have the capacity to attend, it is difficult to see what work such a representation would do. We can see this by returning to the guitar example I introduced in the previous section. I suggested that learning how to play the guitar provided a good example of an increase in motor skills being the sole cause of increased competence, without the descriptive representation of the guitar having to be altered. We could take this argument one step further and ask whether there needs to be a descriptive representation of the guitar underlying the development of guitar skills at all.

One way we can conceive of an underlying representation in this case might go as follows: despite not contributing to the development of guitar skilfulness, the player possesses a representation containing information about the instrument's size, location and features (that it has 6 strings, a thin neck and a thicker body etc.). It could be argued that such facts about the object need to be known to the player before competent interaction is possible; that guitar skills can develop at all depends upon this information being represented in the brain. However, if we think about these features, they too can be replaced by motor-control based explanations. Rather than the player knowing *that* the instrument has six evenly spaced strings, we can instead say he has knowledge of how to move his fingers so that they consistently press down on the guitar neck without any string being accidentally muted. Rather than the player knowing *that* the object is located in the corner of the room we could instead argue that he has knowledge of how to traverse the room so that he is nearer to the object. Similarly, we might argue that no informational foundation is required for AUBs and their development.

The merits of explaining intelligent action in terms of practical knowledge how will be examined in more detail in the next two chapters. What can be concluded here though, is that the notion of descriptive information does not appear to contribute anything to the explanation of intelligent action that cannot be explained in terms of motor skills. Furthermore, motor skills explanations would appear to be clearer and more informative than RTM style explanations: motor skills explanations allow for an account of how development is possible, descriptive representations do not; it is clear that motor competence is required to perform intelligent actions, it is not clear what role descriptive information plays that could not be accounted for in motor terms.

## **Conclusions**

It would seem then, that RTM fails to explain attention understanding. The empirical evidence surveyed in this chapter demonstrates that descriptive representations are unable to fully explain attention understanding without also appealing to the infant's motor-control. This approach however, undermines the RTM claims that neural configurations bearing descriptive information are the most important factor in explaining intelligent action. For these reasons, the theory should be disregarded as a means of explaining AUBs.

We have seen also, that an infant's motor capabilities would appear to play a key role in attention understanding and its development. The following two chapters will explore these issues further. In the next chapter I shall examine DS explanations of attention understanding, which reduce understanding to tuned motor response to the environment. In Chapter 5 I shall discuss AOR accounts of attention, which shall return us to the relationship between motor knowledge and representational understanding.

## **Chapter 4: Attention and Dynamical Systems**

### **Introduction**

At the end of Chapter 2 we saw that DS accounts of cognition can be affiliated with lean accounts of attention understanding because both reject the claim that there is a single representation, concept, or neural configuration that is the cause of AUBs. In this chapter I will explore whether DS provides a better explanation of AUBs than the RTM approach. One reason for thinking that it might is that it is better able to accommodate motor skills as an explanation of intelligent action. We have just seen that RTM's claim that brain activity is the principle cause of intelligent interaction leads the theory into difficulties. As we saw in 1.2, however, embodied accounts place much more importance on the body (and the environment also) and so would appear to be more amenable to motor skills explanations.

The first two sections of this chapter will follow a similar pattern to the Chapter 3. 4.1 will present a general outline of how DS would explain AUBs before going on in 4.2 to add more detail to the account. Whilst RTM required further details as to what descriptive information of attention could consist of, more detail needs to be given to the DS claim that states of the brain are entirely non-informational. As we shall see, even though descriptive representations fail to explain intelligent action, it is difficult not to think of the states of the brain involved in action as being some how 'about' the objects or people that are being interacted with. To combat this, proponents of DS often invoke a form of intentionality based on that of phenomenologists like Heidegger and Merleau-Ponty which conceives of brain states 'directing' behaviour rather than 'about' the world. I shall focus on two aspects of embodied intentionality: the "intentional arc" and "maximal grip".

In the last two sections I will return to the empirical evidence and assess how effective this freshly enriched form of DS is in explaining it. In 4.3 I will show that the intentional arc and maximal grip provide an account of inconsistent gaze following and the development of directive behaviours that is much more compelling than the RTM account. In 4.4 however, we will see that these explanatory tools fair less well when explaining a third aspect of attention understanding: gaze checking.

### **4.1 A first pass at DS accounts of AUBs**

In the last chapter, we saw that responsive and directive behaviours are accounted for by RTM in terms of goal states. I showed that even responses to the actions of another could be thought of as the infant's recognising an opportunity to achieve a goal. In contrast, DS grounds both responsive and

directive behaviour in terms of complex response to stimuli. A first pass at how DS would account for responsive and directive behaviours might run as follows:

**Responsive:** On the DS account, intelligent action "is experienced as a steady flow of skilful activity in response to one's sense of the situation". Over time, the brain receives positive and negative feedback on how successful a particular response to a stimulus is in achieving an "optimal gestalt" and the brain's starting state is "changed in a way that reflects the extent to which the result satisfied the animal's current need" (Dreyfus 2007, 20) . Responsive AUBs would therefore be accounted for in the following manner: observable features of the world, such as head turns, elicit tension in the infant's brain and the infant acts so as to try and relieve this tension. How successful the infant's response is depends upon how much feedback has been received up until that point: in the first few months of life turns of the head are met with various responses by the infant, with the ones that dissipated tension most successfully being more likely to be repeated in the future. Eventually, enough feedback from successful motor responses leads to the starting state of the brain being sufficiently tuned so that turns of the head are consistently followed.

**Directive:** AUBs such as imperative pointing would be explained by a slightly more complex version of the same strategy. Desirable objects can be thought of as eliciting tension in the infant; the optimal gestalt for a situation in which there is a desirable object would be the infant possessing that object. Over time (the first few months of life) various actions are made in response to desired objects, the results of which, positive and negative, feed back into the system and modify the starting state of the brain. Actions that are successful in achieving the optimal gestalt configure the starting state so that they are more likely to be performed. Over time the system hits upon pointing at the object as a successful way of reaching optimal gestalt, causing this particular response to be positively reinforced. Considered in these terms, directive behaviours are no different from responsive behaviours: both rely on a starting state which has been sufficiently tuned so that the infant consistently responds to stimuli in a way that moves it towards the optimal gestalt.

## 4.2 Dynamical Systems and Information

We have seen, in 1.2, that one of the principle ways DS theorists distinguish their account of intelligent action from RTM is by rejecting RP3: states of the brain involved in action should not be thought of as bearing content or information about something external to themselves. From our discussions of motor skills and RTM we have already seen good reason to reject the idea that the brain states involved

in action contain *descriptive* information. This was because information describing that a person is attending to an object was shown to do little work when explaining how AUBs develop. Proponents of DS make a stronger claim, however, and argue that information processing of any type, descriptive or otherwise, has nothing to do with an individual's intelligent action. Whether it is correct to argue this has so far not been established. As we saw in Chapter 1, descriptive representations are not the only variety of mental representation on offer: supporters of AOR hold that the states of the brain involved in action should be thought of in terms of anticipatory information about the affordances a person or object provides. Before I go on in Chapter 5 to investigate how well the claims of AOR stand up to scrutiny, this chapter will explore whether intelligent action really can be accounted for in entirely non-informational terms

The task of DS is not helped by the intuition that, whether or not they can be thought of as descriptive, the brain states involved in action must somehow be 'about' the world. A compelling way of thinking about the brain states involved in action might run as follows: If we could find some configuration of neurons common to all instances where the infant sees a pair of eyes -this does not seem unreasonable to suppose- then that neuronal configuration bears information about, represents, attention. Similarly, if we could find a pattern of neuronal activity common to all instances when a head turn is followed we could call that the processing of representations aimed at achieving a goal state. The task of the DS theorist is to persuade us that these intuitions are incorrect and that these brain states are not content bearing and are in fact better thought of in terms of starting states, response, and tension dissipation etc.

Put another way, DS needs to show that the language it uses to describe brain processes is not simply disguised representational terminology. We have seen that the theory explains intelligent action in terms of the system attempting to move towards the optimal gestalt of any given situation. As it currently stands however, the notion of an optimal gestalt is open to the charge of being simply a disguised goal state. We can see this by noting that optimal gestalts and goals play very similar roles in explaining intelligent action. For example, if we were explaining gaze following, or any other behaviour, in terms of DS we might say "The infant turns its head in the same direction because the system is trying to move towards the optimal gestalt". In terms of a representational theory we would say "The infant turns its head in the same direction because the system was trying to bring about a goal state".

A similar difficulty can be seen in pre-tuned starting state strategies. Once a starting state has been sufficiently 'set' to respond to head turns we might expect that the perception of head turns would consistently elicit brain processing of a particular type. Such a process would play a similar role in DS theories as representations do in informational theories: In both, stimuli from the environment causes a particular neural configuration, which in turn causes intelligent action. The difference between the two theories is that one claims that this process is informational whereas the other denies it. As with

starting states then, a supporter of DS needs to provide good reasons to believe that this process is truly non-representational.

### **Embodied Intentionality**

In the previous chapter RTM was supplemented with two 'theories of mind' in order to make better sense of the notion of descriptive information. Here, DS accounts will be 'filled in' in a similar manner, except this time the aim will be to make better sense of *non*-informational cognition. To do this, I will turn to a philosophical theory often employed by DS supporters which I will refer to as 'Embodied Intentionality'.

Often, the term 'intentionality' has been used by philosophers to refer to the ideas just discussed -that mental states are about things in the world. The term was originally used in this way by Brentano who claimed that:

Every mental phenomenon includes something as object within itself, although they do not do so in the same way. In presentation, something is presented, in judgement something is affirmed or denied, in love loved, in hate hated, in desire desired and so on.

(Brentano 1874, 88)

Echoes of these ideas were seen in the discussion of RTM in Chapter 1. There is a common something in my imagining, seeing, or fearing a bus; although my attitude in each case is different, all three are about the same object. We saw that RTM conceives of representations as being reasoned with or processed in some way so as to cause action. From this point of view, it could be argued that part of what links my deduction that a speeding bus will do me harm, my imagining that a speeding bus will do me harm and my fearing that a speeding bus will do me harm is that all three contain the same intentional object. This fact could be used to explain why my reasoning about a real bus and my reasoning about a hypothetical bus follow broadly the same 'rules' (have the same inferential structure).

In rejecting information as a means of explanation, proponents of DS reject this conception of intentionality and embrace instead a theory derived from the phenomenology of the philosophers Heidegger and Merleau-Ponty (see Heidegger 1967; Merleau-Ponty 1945/2002 and Dreyfus 1991, 1992; Freeman 2001; Gallagher & Zahavi 2008; Thompson 2007,<sup>5</sup>). On this view, “intentional experiences are conceptualized not as states having content but as acts having directedness” (Thompson, 25). What is meant by “directedness” can be seen if we consider an example from the

---

5 It should be mentioned that there are other phenomenologists who were not anti-representationalists. For example Husserl (2001).

natural world. This description, taken from a website providing information about wild flowers, explains how a sunflower turns towards the sun:

There is a collection of specialized cells at the base of the flower bud or leaf called a pulvinus that carry the "motor" cells that enable the plant leaf or flower to track the sun. These cells enlarge or shrink according the turgor pressure from the water inside them. In response to blue wavelength light, potassium ion concentration increases in the "motor" cells on the shadow side of the pulvinus. With the increase of potassium ions the osmotic potential in the cells becomes more negative and the cells absorb more water and elongate, turning the face of the flower to the sun.

(“Why do sunflowers turn towards the sun?”)

Here we can see an example of a process that directs 'behaviour' towards an end state (facing the sun), which at no point contains anything we would be tempted to call a representation. It does not seem correct to describe the increase of potassium ions as representing or containing information about the sun, despite its being caused by the sun's blue wavelength light, and subsequently causing the the face of the flower to turn. Rather, the behaviour of the flower is directed by the sun through complex causal interactions between its parts. Supporters of a non-representational form of intentionality consider the brain states that influence human behaviour to be similarly directed by and towards the world.

Of course, this issue cannot be just left like this. This type of directedness maybe clearly seen in the behaviour and structure of a plant, but human beings' brains and bodies are much more complex, as is the behaviour they display. Again, the intuition emerges that sophisticated behaviour cannot be the product of anything other than information processing. To combat this idea I will examine two ideas from Dreyfus, which he himself has taken from Merleau-Ponty: the “intentional arc” and “maximal grip”.

### **The Intentional Arc and Perception**

The intentional arc is used by Dreyfus to show how non-informational learning can take place:

What one has learned appears in the way the world shows up; it is not represented in the mind and added on to the present experience. That is, according to Merleau-Ponty, what the learner acquires through experience is not represented in the mind at all but is presented to the learner as a more and more finely discriminated situation, which then solicits a more and more refined response. In so far as the situation does not clearly solicit a single response or the response does not produce a satisfactory result, the learner is

led to further refine his discriminations. Merleau-Ponty calls this feed-back loop the intentional arc.

(Dreyfus, 2002, 373)

This idea is clearly compatible with the idea first introduced in 1.2, of a brain's starting state being tweaked by positive and negative feedback so that stimuli are met with appropriate responses. What the intentional arc adds is a way of seeing more clearly how such a process is possible without information or representations. Although embodied approaches to cognition typically blur the divide between perception and action, here Dreyfus separates the intentional arc into two components: a perceptual part in which the learner is able to discriminate to an increasingly fine degree what opportunities a situation offers, and an active part in which the learner's responses to the stimuli become more and more refined. The latter I will examine more closely in a moment, when I turn my attention to the notion of maximal grip.

The perceptual component of the intentional arc allows us to see how features of the environment can elicit intelligent action from a starting state, while at the same time avoiding any reference to representation or information. We can see more clearly what is meant by the world “showing up” and situations being “more and more finely discriminated” by taking the example of a digital camera. The amount of detail included in a photo taken camera depends to a large extent on the number of image sensor elements (“megapixels”) that the device has. If I take a photo of my sofa on a 5 megapixel camera, the picture will show quite a detailed scene, I will be able to see the sofa's colour, shape and size relative to surrounding objects. If I take a photo with a 12 megapixel camera all of these things would be in the photo, but also extra details, like the pattern or texture of the sofa cover and the stitches around the cushions. That the 12 megapixel model can detect these extra elements does not depend on the camera *deriving more information* from the environment, but instead is due to an increased sensitivity to, an increased ability to “pick up” features of the environment.

How finely we can discriminate a situation can be thought of as analogous. The opportunities presented to us depend upon how well we can detect features of the environment, not upon those features causing representations in the brain. Depending on how well developed an infant's perceptual capacities are, it will be able to distinguish objects, like pairs of eyes, and events, like head turns, from other less relevant features of a situation. They are distinguished not because they trigger some information bearing state in the infant but because the infant has been primed through previous experience to respond to them. We could say that this gradual increase in sensitivity is similar to slowly increasing the resolution of the camera, but here we reach the limit of the analogy. Increasing the number of megapixels in a camera could only be done by adding extra image sensor elements to the device, whereas in the case of the intentional arc development is the product of the brain's starting state being gradually altered by positive and negative feedback. As Dreyfus says, if the stimuli detected

in a situation does not clearly direct the subject towards maximal grip, then further perceptual discriminations will be made, and more finely detailed opportunities for action will be picked up.

We are now some way towards seeing more clearly why DS theorists think that starting states can be thought of as causing intelligent action while avoiding representations. A starting state is not just a bundle of accumulated information, waiting to be elicited by the appropriate visual stimuli. It is a mechanism that is able to pick up increasingly fine-grained features of the environment. I will now move away from the perceptual element of embodied intentionality and turn instead to action. We shall see that the notion of maximal grip can help us to properly separate optimal gestalts from representational goal states.

### **Action and Maximal Grip**

The notion of maximal grip is used by Dreyfus to explain how intelligent action is caused by features of the environment directing our behaviour:

When we are looking at something, we tend, without thinking about it, to find the best distance for taking in both the thing as a whole and its different parts. When grasping something, we tend to grab it in such a way as to get the best grip on it

(Dreyfus, 2002, 378)

The idea here is that instead of our movements being the result of information processing (having a goal state and working out which movements will best achieve that goal) our starting states are tuned so that perceived features of the environment direct our intelligent behaviour. One half of intelligent action on Dreyfus's account is the ability to discriminate aspects of the environment, the other is the ability to intelligently move one's body in response to them.

To see how the notion of maximal grip can better differentiate optimal gestalts from goal states it will be useful to extend the metaphor of gripping an object. If we pick up an object, perhaps something slightly awkward like a bulky television set, for the first few moments that we have the object in our hands we fumble with it, finding the best, most comfortable way to hold it. The optimum way to hold the object will depend on features of the object, like its size and shape, how heavy it is and how that weight is distributed, as well as on features of our body, like our size and strength, posture etc. When we begin to handle an object it does not seem that we have a particular goal state in mind. It is not normally the case that I see an object, decide upon exactly the best way to hold it and then move so as to immediately bring it about. Rather, the object itself guides my hands as I move them: I move my hand along the object until I feel that the heaviest part of it is supported by my palm, I turn the object so that no sharp corner is pressing against me. In my fumbling I find my way towards the optimal way

of carrying the object but no clear, planned trajectory towards a goal state is present before I pick it up. Dreyfus and Merleau-Ponty want to say that our acting in the world is similar to this type of handling of objects. Just as the curvatures and corners of an object push us towards to optimal way of holding it, the various aspects of a situation -what features of the environment we detect- push us towards the optimal gestalt.

With maximal grip, we have an approach to action that is much more compatible with the motor skills explanations of AUBs than that provided by RTM. We saw in the previous chapter that infant motor skills are RTM's Achilles heel, the theory seems forced to rely on them to explain intelligent action but doing so erodes much of the theory's explanatory power. We can see with maximal grip however, that motor skills are able to be much more closely tied with DS accounts. How easily I can obtain a maximal grip depends upon the state of my motor skills, which in turn depend on how much previous experience I have of similar situations: If I frequently picked up bulky, heavy objects, I would likely develop a particular technique of how to carry them. A removal man would likely be more adept at picking up television sets than someone who doesn't do it very often. Similarly, the more experience I have of the world the more fluidly and automatically I am able to respond to the environment.

This idea returns us to the intentional arc. Just as our perceptual capacities are refined through experience, so too are our motor skills. Becoming more adept at picking up bulky objects (being able to do so with growing speed and efficiency) is a case of the features of an object, its corners and curves, being detected with increasing accuracy but also with their detection being met with increasingly sophisticated motor responses. An experienced removal man would automatically meet a particular curve with a particular grip, or feel where the heaviest point of the object is and quickly react so that his hand best supports that weight. An inexperienced lifter might struggle to find a comfortable way to hold the object. This would not depend on the removal man having learned more 'rules' about lifting objects, but instead upon how well tuned his body is to react to the features of the object.

### **Information and Explanatory Leverage**

We can now see how embodied approaches to intentionality can be used to supplement DS. With the intentional arc and maximal grip, we can see more clearly how an entirely non-informational account of cognition could be possible and that starting states and optimal gestalts are not simply pseudo-representations. In 4.3 I will assess whether this form of DS can provide a better account than RTM of the two aspects of attention understanding introduced in the previous chapter: inconsistent gaze following and the development of directive behaviours. Before doing this however, it will be useful to clarify exactly what would determine whether a process involves information processing or not. When considering intelligent behaviour, how can we say with certainty whether an action was the result of information processing or an attempt to get a maximal grip on the situation?

When answering this question it does not seem useful to try and identify information with a particular physical property that groups of neurons can either have or lack. This is because, when looked at closely, real life examples of information processing do not possess a feature that distinguishes them from mere causal processes, and can, if examined in sufficient detail, be reduced to mere causal interactions themselves. A clear, and I assume uncontroversial, case of information processing in the real world is that desktop computer: information is received (through presses on the keyboard, or mouse clicks), processed by various internal mechanisms (a video card, a CPU), and used to generate output (an image on a screen). But of course, when we look at the same process more closely, all we really have is a great many causal interactions: presses on the keyboard affect the patterns of protons sent to microchips. In a similar fashion, the workings of the brain, complex as they are, can be boiled down to brute causality, interactions of activation between individual neurons. Therefore, the idea that informational processing can be distinguished from the merely causal by some physical characteristic alone, seems implausible.

A more fruitful approach to deciding if brains states should be considered informational, would be to ask whether or not it is useful to describe them as such. Does thinking about a process in terms of information being sent, received, retained etc. allow for a clear, practical description which highlights the system's most important characteristics? Or is a more useful picture of the system provided by the intentional arc and maximal grip? We can find this idea in Clark (1998), who proposes that “explanatory leverage” should play a role in determining whether a process is informational or not:

... it will be sufficient if the system is such that items which are to be treated similarly become represented by encodings (such as patterns of activation in a population of neurons or an artificial neural net) which are close together in some suitable high-dimensional state space...It seems reasonably clear *that by glossing states of the neuronal population as codings... we gain useful explanatory leverage* . Such glosses help us understand the flow of information within the system, when, for example, we find other neuronal groups (such as motor-control populations) that consume the information encoded in the target population.

(145, my emphasis)

Clark, being one of the originators of AOR, believes that explanatory leverage is gained by thinking of populations of neurons as encoding information which is then consumed by other groups. He believes that by identifying particular neural configurations as information bearing we are better able to see how the brain functions.

The description of how a PC functions has a similar informational gloss, allowing for a similarly practical description. The workings of the PC could be described in a non-informational way, but at the

cost of sacrificing explanation as to how the computer does what it does. If you are having difficulty getting the printer to function, it would normally be much more useful to think of the printer as a device that consumes information (thereby allowing questions of whether the information is being received by the printer, or whether the appropriate information is being sent), than in terms of electronic charges activating microchips.

There are plenty of other causal processes, on the other hand, that do not benefit from such a representational gloss. We have just seen that once we know a little about a sunflower's internal workings there is little reason to try and shoehorn them into a representational framework. Again, we could think of blue wavelength light as the stimulus and the increase of potassium ions as the representation it elicits, but it would not add to our understanding of the sunflower. How the sunflower turns towards the sun can be explained perfectly easily without recourse to informational language.

I will use the idea of explanatory leverage to assess whether AUBs are better accounted for by representational or anti-representational theories. In a moment, I shall turn to DS explanations of particular AUBs and assess whether a clear and informative account can be given in terms of maximal grip and the intentional arc. This is not to say, however, that this method is perfect, as it is not always the case that people will agree on what the most practical way of describing a system is. Although most people would agree that a sunflower's functioning is best explained in non-informational terms, in more complex cases, such as human behaviour, there is a greater chance of disagreement over whether a system is better described in terms of information or not. Furthermore, it may be the case that different explanatory glosses are required for the same system depending on context. It might help, as Clark says, to think of sections of the brain in terms of information when one is concerned about how intelligent behaviour is caused, but on the other hand, there could be cases, perhaps in cases of brain surgery, where it would better to think of brain processes in causal terms.

Nevertheless, explanatory leverage can still help to determine the value of a theory. In the previous chapter, we saw that RTM does not provide a satisfactory account of attention understanding. This was not done not by determining that a physical property of information was missing from the neuronal states involved in action, but instead by showing that the tools of RTM (inference, description etc.) were not sufficient to account for the empirical data in question. My strategy here will be similar, to determine whether the theoretical framework of DS can accommodate all aspects of AUBs and their development. The question of whether the brain processes involved in AUBs are informational or not can therefore be re-framed as: Is more explanatory leverage gained by thinking of AUBs as the product of information processing or non-informational directedness? To answer this question, I now return to the two aspects of attention understanding that were so problematic for RTM accounts: inconsistent gaze checking and the development of directive behaviours.

## **4.3 Embodied Intentionality and Attention Understanding Behaviours**

We saw in the previous chapter that RTM accounts of these behaviours were forced to appeal to motor-control to provide a full explanation and this caused the theory serious difficulties. I will show here that DS does not have such problems as the intentional arc and maximal grip provide sufficient explanatory leverage when accounting for inconsistent gaze following and the development of directive behaviours

### **Inconsistent Gaze Following**

One possible way that DS could explain this behaviour is to appeal to the perceptual component of embodied intentionality and argue that the infant does not follow gaze consistently because it is not yet able to consistently detect head turns. We saw in the previous section that the further along the intentional arc an infant is, the greater its ability to detect the fine details of a situation. It could therefore be argued that the starting state of a three month old needs to be tweaked further by positive and negative feedback before the infant will be able to consistently distinguish gaze redirection.

However, this approach is not entirely satisfactory. In the previous chapter we saw that it was not possible for RTM to explain inconsistent gaze following as the product of an inconsistent ability to form a representation. This was because there appeared to be no clear reason why a stimulus would only fleetingly cause a representation. A similar criticism could be made of this approach: Why would a starting state detect eye movement in some instances and not others? There would be no equivalent to this in the camera analogy, if an object was photographed from the same angle in the same lighting conditions then the same features of the environment would be detected every single time.

A better way that this behaviour can be explained in terms of embodied intentionality is by appealing to maximal grip. The instances where the infant does not follow gaze are not due to a failure to detect the event but a failure to get a grip on the situation. In the previous section, we saw that maximal grip and motor response are closely intertwined; how easily the optimal gestalt can be obtained depends on how honed a person's motor responses are. The removal man, through years of practice, will be able to find the optimal way of holding a heavy bulky object quicker than someone with less experience. Someone with little or no experience of heavy lifting might have trouble picking up the object at all, or often drop if they do. A 3 month-old's inconsistent gaze following could be thought equivalent to this last case. Despite being able to detect head turns, the infant's lack of experience, and subsequent lack of motor-control, means that it frequently fails to respond in an appropriate manner.

### **Development of Directive Behaviours**

A similar explanation could be put forward to account for the development of directive behaviours. We saw in the previous chapter that motor skills are a plausible explanation of how the infant develops from gesturing towards parts of its own body to full blown pointing. If this is the case then we can claim that an 8 month-old is able to obtain a maximal grip on situations involving proximal objects but has not yet learned to 'handle' situations involving distal ones. However, such an approach only goes so far: the infant's failure to point towards distal objects is not so much a failed attempt at achieving maximal grip as a failure to respond at all. Distal objects elicit no response, not an imperfect one. The equivalent in the removal man example would be not of an object being handled but of an object, or a feature of an object being ignored.

A better approach is to appeal to the perceptual component of the intentional arc and argue that a failure to point stems from a failure to discriminate distal objects. We saw that this strategy could not account for inconsistent gaze following because the infant is presented with an identical stimulus in each case and so it is difficult to see why it would be detected only some of the time. With directive behaviours, however, there are two different types of stimulus, proximal and distal, and so it seems more reasonable to argue that one can be detected and the other not.

Once again we can see this with the metaphor of the camera. Depending on its megapixel count, there will be a limit to how far away the camera has to be from an object before it ceases to pick it up. If I take an outdoor shot, the camera will be able to detect the individual leaves on the trees closer to me, but those of a tree 100 metres away will be a unified blob of colour. Similarly, the lack of pointing in an 8 month old could be attributed to its inability to differentiate distal objects from the background. It is important to notice here that the camera still detects *something*, the distant tree is still in the picture, it simply lacks fine detail. With the infant it is the same, it is not blind to distal objects, something is seen, but there is a lack of clarity and therefore a failure to respond. As its powers of perceptual discrimination improve, distal objects are better differentiated and so better responded to.

At this point it appears that DS offers an account of AUB development which has much greater explanatory leverage than RTM. We saw in the previous chapter that motor skills were an important factor in explaining how AUBs develop and we have seen that this can be easily incorporated into accounts based on embodied intentionality. However, as we shall see in this chapter's final section we should not be too quick to dismiss representations and information as a means of accounting for action.

#### **4.4 Gaze Checking and Dynamical Systems**

In this section I shall introduce one behaviour in which it appears more explanatory leverage would be gained by thinking of brain activity in informational terms: the emergence of gaze checking at 12

months. This AUB can be considered a responsive behaviour, as the infant is caused to act by visual stimuli: another's pointing gesture is seen by the infant which causes it to redirect its own gaze. However, whereas other responsive behaviours like ordinary gaze following have one step -stimuli and action- instances of gaze checking have (at least) a second; the infant will either look at the object and then look back at the pointer's face, or will look first at the pointer's face and then the object.

Hobson & Hobson provide two examples, the first of a child aged 13 months and the second aged 12 months:

**GC1:** Initially, the child's gaze lingered on the tester's outstretched hand. Next she looked back to the tester's face as if to ascertain what the tester might be trying to communicate. For a moment she seemed to dwell on her face and then suddenly she shifted her gaze to the target of the tester's still outstretched finger.

**GC2:** ... a 12- month-old girl swiftly followed the tester's gaze and outstretched finger to locate a poster on the wall and then, giving herself barely enough time to take in the contents of the poster, quickly looked back to the tester's face with an engaging smile. The tester spontaneously commented on this sharing, with a playful and affirming "Did you see Big Bird?" and the child immediately turned, still smiling, to look back at the poster again

(Hobson & Hobson 2008, 78)

DS is committed to explaining these behaviours in the same way that it has explained inconsistent gaze following and the development of directive behaviours: in terms of the intentional arc and maximal grip. However, it is difficult to see how the perceptual element of embodied intentionality would be of use here, as neither behaviour suggests that the infant is attempting to, or failing to, distinguish a feature of the pointer. Both the point, the target object, and the pointer's face appear to be picked up by the infant without difficulty.

The best chance DS has of accounting for gaze checking would appear then to be in terms of action and maximal grip. The infant's looking at the pointer's face before or after looking at the target object could be thought of as the infant attempting to move its body so that the tension caused by the stimulus is fully dissipated and maximal grip obtained. The success of this style of explanation depends upon how easily each movement made by the infant can be thought of as its being *directed* towards the optimum by the features of the situation (the hand, the face, the target object). The maximal grip explanation of inconsistent gaze following had features of the environment guiding or

pushing the movements of the infant rather than informing the infant, we should expect a similar pattern of interaction in this case also.

However, it would appear that gaze checking is not so easily accommodated in these terms. As with gaze following behaviours we can presume that the optimal gestalt for the infant in GC2 is for it to be looking towards the target object. If this is true, however, we might ask why the infant finds the target object, looks away, and then returns to it. Why was the tension not dissipated (maximal grip not achieved) the first time that the infant turns to the poster? Put another way, in the movement towards the optimal gestalt, why do the features of the situation direct the infant's behaviour so that it 'doubles back' and is guided to the same stimulus twice (first the point directs to the target, then the face directs to the same target)? The equivalent in the removal man example would be of a person feeling corner A of the object, then corner B, before returning to corner A again. In both cases, the actor returns to a location from where it has already been directed, suggesting that there is more going on than them simply being guided towards the optimal gestalt.

In GC1 the problem is slightly different. If we again presume that the optimal gestalt of this situation for the infant to be looking at the target object, then it is the case that features of the pointer (the outstretched hand, the face) direct it there eventually. However, maximal grip does not explain the pauses in the infants behaviour. Why does the infant's gaze "linger" on the pointer's finger and then stare at the face before the target object is found? If embodied intentionality is correct then the infant at 12 months has sufficiently honed its powers of discrimination and maximal grip for it to be able to pick out features like points and gaze direction, and react to them accordingly. There does not seem a good reason why this type of stimulus response would not be immediate. If my aim is to find the best way of holding a television set and I am using the shape of the object to guide me towards the optimal way of holding it, there seems no reason why my hand should "linger" on any one feature of the object.

### **Information and Gaze Checking**

These difficulties are highlighted when we consider the ease with which gaze checking can be explained in terms of information processing. Even if we disregard Hobson and Hobson's slightly evocative description ("as if to ascertain what the tester might be trying to communicate") GC1 invites an informational explanation: the infant looks at the hand first, but is unable to recognise its significance; it is only when it sees that the pointer's eyes are looking in the same direction as the outstretched finger that it realises that it is being invited to look in a particular direction. The movement from hand to face is not a case of the infant's gaze being guided by the first stimulus to the second, but of extra information being sought in the face of ambiguity. Once information has been gathered from the finger and the face, the infant is able to infer that the pointer is attempting to draw attention towards some target object.

GC2 is also explained fairly easily in terms of information. In fact, there are two information-based explanations as to why the infant turns back to the pointer's face. Firstly the infant could be 'double-checking' that it has in fact found the correct object. It does this by checking that the direction of the pointer's eyes is the same as that of the pointer's hand. If the two are the same then the infant can be confident that it has found the correct object, if they do not then the infant can look to see if there is another target object to attend to. Again, this style of explanation implies that information is being gathered. The infant turns to the face in order to see whether the information it presents correlates with the information presented by the hand. The second possible explanation would be that the infant is gauging the pointer's attitude towards the object. Having inferred that the picture of big bird is the target of the pointer's point, the infant seeks additional information as to what value or relevance the pointer places on the object. Such an approach would fit well with the fact that, due to the infant's "engaging smile" and the pointer's friendly response, that there is some emotional interaction place here.

As both of these explanations involve the processing of information they are much better placed to account for the pauses in the infants' behaviours. If the infant is calculating, inferring or comparing items of information then these processes would take time, and so it would make sense that there are gaps in the infant's actions. It would appear that the embodied intentionality is at its strongest when explaining fluent, immediate responses to the environment, like following gaze, but that when behaviour is not so fluid, information processing is the better option.

### **Gaze Checking and Explanatory Leverage**

In response to these difficulties a proponent of DS might argue that I am presenting an over-simplified version of the theory and that I have failed to appreciate how nuanced the interaction between body, brain and environment can be. DS distinguishes itself from behaviourism with the claim that the relation between the starting state of the brain and the stimuli that acts upon it, is very fine-grained:

...a sensory stimulus from an object does indeed induce the formation of a pattern in the brain, but when it is given repeatedly it does not induce precisely the same pattern in the same brain... its meaning for the same person is continually shifting.

(Freeman 2001, 29)

The precise details of the brain state that an infant possesses at a given moment are obviously very complex, as are the fine details of a given situation. Therefore, the exact manner in which the brain will seek to find optimal environment-body gestalt will be equally nuanced, depending on any number of environmental and neurological interactions. It could be argued that we should not expect DS accounts of behaviour to be obvious or easy to find, the complicated interactions between neurons and

the nuances of which details are picked up from the environment might lead to maximal grip being obtained by less intuitively obvious routes than we might think. These might include doubling back to a stimulus that has already been responded to, or pausing before responding.

Clearly, dynamic coupling with the environment involves complex interaction and so this claim holds some weight. However, there is a danger here that 'complex interaction' can be invoked as an explanation for any behaviour that the DS has difficulty explaining. We saw in 4.2 that a useful way of deciding if a system is informational or not is to ask which theory provides the most explanatory leverage. To argue that gaze causing is caused by unknown interactions between the starting state of the brain and the environment weakens the explanatory value of DS.

As well as this, there is the worry that this strategy moves DS too far away from maximal grip style explanations. Part of the explanatory power of maximal grip is that by comparing interaction with the environment to the contours of an object guiding how one holds it, we have a clear benchmark for how intelligent action is possible without the use of representations or goals. If, however, DS is forced to disregard this metaphor and allow for the type of doubling back or pausing just mentioned then it casts doubt on whether handling an object is a suitable means of explaining behaviour at all.

A response to this criticism might be to argue that the maximal grip metaphor has its limits. Although obtaining a maximal grip on a television set might not require movements to be repeated, there are other activities that do. For example, if we are operating a computer with a mouse we frequently need to 'double click' on an icon to launch a program. Such a behaviour involves the same stimulus being doubled back to (the button is clicked twice), but it does not invite an informational explanation. It is quite easy to think that once we are sufficiently used to operating a computer, tapping on the mouse becomes an automatic behaviour of the type employed by maximal grip explanations. In this case it does not seem plausible that any information gathering is going on, or any information processing between my first click and my second.

This same type of reasoning might be applied to GC2 by arguing that the infant's double check of the pointer's face is an automatic response to the exact configuration of stimuli in the situation. Perhaps the infant has developed (through experience, like all maximal grip competencies) this behaviour as a response to situations which involve a certain type of target object, or a particular pointer, or a particular style of point. To take this last example, it could be the case that when the infant is confronted with a slightly unclear point (the arm not being raised fully, for example), its behaviour has developed so that it follows the point to the target object, but is then directed by the target object to the pointer's face. It is subsequently directed *a second time* to the face. Such a behaviour might have developed to guarantee that the infant is looking at the relevant object.

It is true to say that maximal grip should not be taken too far as a metaphor, behaviours which involve a return to the same stimuli can be thought of as non-informational. However I do not think the

above argument the above argument can be applied to GC2, the infant's looking back at the speaker's face cannot be thought of as a pre-configured response.

Firstly, it is difficult to see how the doubling back to the pointer's face could be thought of as an automatic. If a brain state is tuned so that a particular stimulus elicits a particular motor response it should be possible to isolate that stimulus and produce the same response again and again. If a person is a competent computer user then it is fairly easy to initiate the 'double click' response: whenever there is an icon on the screen that opens a required program or file, that user will automatically, unreflectively double click on it. This is not true of gaze checking. Infants do not double back to a pointer's face every time they are directed to a target object. In GC1 for example, the infant is (eventually) directed to the target object but upon seeing it does not respond in this case by checking the pointer's line of sight.

Secondly, once the infant has turned back to face the pointer, the emotional interplay between the two participants still suggests that there is more going on than the infant being simply 'redirected' back to the target object. We see that the infant pauses to smile at the pointer and returns to the target object, still smiling, after it has been "commented on". What we could take from this is that the infant is not just being guided back to the target object by the direction of the pointer's eyes, but has, through emotional interaction, changed its attitude towards the target object.

## **Conclusion**

The DS account of AUBs is more successful than that offered by RTM. However, as we have just seen, there are limits as to how much explanatory leverage the theory can provide for a behaviour like gaze checking. The main advantage that DS has is that, in embodied intentionality we have a means of explaining gaze following and directive behaviours which includes the infant's possession or lack of motor skills. Being an EEC approach DS defines an infant's understanding not just as activity in the brain, but also in terms of its being able to competently move its body in response to features of the world. As maximal grip and the intentional arc place much more emphasis on the infant's bodily capacities, what it can perceive and how well it can respond to these perceptions, the role of motor control in AUBs helps the theory rather than, as was the case with RTM, hinders it.

The theory's main disadvantage is that there are some AUBs that would appear to resist explanation solely in terms of directive intentionality and are instead much better suited to information based accounts. With gaze checking we can see that not all intelligent behaviour is easily thought of in terms of the environment directing the infant towards a maximal grip, and attempts to do so risk looking overly obscure and so weaken the theory's explanatory value.

With these problems in mind I shall turn now to AOR, an embodied but representational account of understanding.

## **Chapter 5: Attention and Action Oriented Representations**

### **Introduction**

In this Chapter, I shall turn to the final theory of understanding, AOR. We shall see in 5.1 that, given the problems experienced by RTM and DS, AOR is well placed to provide the best explanation of AUBs. However, the account as it stands is incomplete. Just as RTM required ST or TT to properly explain how descriptive information should be understood, further detail needs to be given as to how the 'anticipatory information' of AORs should be understood.

In 5.2, I will suggest that AOR can be supplemented by a theory of understanding other minds known as Intersubjectivity. This theory can be distinguished from RTM based accounts by its rejection of the idea that others' mental states are hidden and need to be inferred through simulation or theorising. Instead, it is claimed that our understanding of other minds is based on perceptual and action-oriented knowledge *how* to interact with others. I will focus in particular on Shaun Gallagher's claim that we are able to perceive the mental states of others "directly" without the need for theorising or simulation, and argue that this theory of Intersubjectivist Perception (IP) can be used to add the necessary detail to emulatory accounts of representation.

In 5.3 I will return to gaze checking and critically assess whether the newly this newly filled out form of AOR can provide more explanatory leverage than DS. We shall see that Gallagher's notion of 'perceptual affordances' allows AOR to provide a more complete account of AUBs than those offered by RTM and DS.

### **5.1 AORs and Gaze Checking**

Chapters 3 and 4 I have shown that neither RTM and nor DS are able to adequately explain the empirical data presented in Chapter 2. In Chapter 3 we saw that RTM's main obstacle is that in order to account for inconsistent behaviour, or the development of behaviour, it has to rely on the infant's motor skills to do much of the explanatory work. The result of this is that descriptive representations begin to look rather redundant. In contrast, we saw in Chapter 4 that DS embraces motor skills as a means of explaining development and so is able to account for inconsistent gaze following and the progression from clowning to pointing with ease. However, a third behaviour, gaze checking, puts pressure on the DS claim that information processing never plays a role in intelligent interaction.

We saw in 1. 3 that AOR can in some ways be thought of as an augmented version of the DS approach: AOR theorists agree that "fluent, coupled real-world action-taking is a necessary component of

cognition" (Clark and Grush, 18), but hold that in certain situations a type of representation is required "whose functional role really does involve playing the part of environmental features that need not be in close, constant interaction with the agent at the time "(16). In the face of the type of difficulties suffered by RTM and DS accounts of AUBs this would seem a potentially fruitful approach. Because AOR accepts action can sometimes be caused entirely by dynamic coupling the theory is able to utilise the (successful) DS explanation of inconsistent gaze checking and development, but invoke AORs to account for the more "representation hungry" (Clark 1998, 169), information based, behaviour of gaze checking.

Gaze checking resisted a maximal grip style explanation because the infant's behaviour was difficult to account for just in terms of the perception of and response to features of objects. The lack of a fluid response in GC1 and the doubling back of GC2 instead suggested that an explanation grounded in information processing was required. We saw in 1.3, that according to Clark and Grush a representation hungry situation would elicit an AOR which "takes as input information about the starting (or current) state of a system...and about the control commands that are being issued" and then "gives as output a prediction of the next state of the system" (5). A successful AOR explanation would demonstrate how this type of emulatory processing causes the infant's gaze to linger on the pointer in GC1, and return to previously perceived stimuli in GC2.

Before it can be determined whether AORs really can account for the infant's lingering gaze, or the returning to a previously perceived stimuli, further detail needs to be added to the theory. Both the informational processing of RTM and the non-informational processing of DS required further elucidation and so too does AOR. Specifically, a fuller account of how anticipatory information can be represented. We saw in Chapter 3 that RTM representations of attention could be thought of as containing either theoretical or simulatory information, but have not yet seen the equivalent for AOR. Describing an AOR as an "emulator" is not sufficient as real life emulators do not represent simply by virtue of being emulators, the information is represented in a particular way. In the chemical plant example, the emulation has three components: mock input, the emulator itself, and mock output. Mock input consisting (presumably) of keyboard presses and mouse clicks, is fed into the emulator, which, through a series of computations, produces a "set of values" (Clark & Grush, 5) as mock output.. Depending on how the emulator is set up, these values would be (re)presented to the user of the emulator in some way; a predicted state of the chemical mix could be shown as a diagram, graph or set of numbers displayed on the screen.

If AOR is to be successful in explaining gaze checking then it needs to provide an account of all three of these components. What would be an AOR equivalent of keyboard input? How would an emulatory representation process it? How would mock output be represented? These questions need to be answered in sufficient detail before it can be determined whether AOR provides a more satisfactory account of gaze checking than DS. In order to achieve this, I shall now turn to the Intersubjectivist theory of understanding others.

## 4.2 Intersubjectivity and Perception

Although Intersubjectivity as an explanation of how we understand others was originally put forward by Trevarthyn in the late seventies (1978, 1979), it has enjoyed a resurgence in popularity in recent years (for example Gallagher 2008b, 2008c, Gallagher & Zahavi, 2007, Johnson Chapter 2 2008, Zlatev et al. 2008). In a similar fashion to embodied accounts of cognition, intersubjectivist theories are characterised by an emphasis on practical knowledge how, rather than inference and representation based knowledge that, as a means of interacting intelligently with others. Zlatev et al. (2008) provide a list of general features which Intersubjectivist theories share:

- 1** Human Beings are primordially connected in their intersubjectivity, rather than functioning as monads that need to “infer” that others are also endowed with experiences and mentalities that are similar to their own.
- 2** The sharing of experiences is not only, not even primarily, on a cognitive level but also (and more basically) on the level of affect, perceptual processes and conative (action-oriented) engagements.
- 3** Such sharing and understanding is based on embodied interaction (e.g., empathic perception, imitation, gesture and practical collaboration).
- 4** Crucial cognitive capacities are initially social and interactional and are only later understood (sic) in private or representational terms.

(Page 3)

What can be seen here is a theory that argues that normally functioning children are born with particular capacities that lead to their interactions with other humans being fundamentally different from their interactions with other objects. Experiences are “shared” between infant and caregiver in that the infant can see (for example) the actions or expressions of her caregiver as meaningful. Experimental data employed to argue for this primordial connection includes studies that show: infants being able to distinguish human faces from other objects neonatal imitation of faces (Johnson 2000; Johnson et al. 1998) and emotional contagion (Hobson & Hobson 2008). Points 3 and 4 emphasise that these capacities are non-cognitive in that the infant does not attend to an object and reflectively make a judgement as to whether an object is a face or not, or what facial muscles would need to be employed so as to imitate another's expression.

From this list we can see common ground between Intersubjectivity and the dynamic coupling explanations of intelligent action employed by both embodied accounts. Firstly, both hold that interaction with others is not based on representations and inference, but rests instead on non-cognitive knowledge of *how* to act. intersubjectivists claim that infants' interactions with others are based not on simulation or theorising but instead on representation-less action-oriented abilities which are “fast, automatic, irresistible and highly stimulus-driven”(Scholl and Tremoulet 2000, 299). Such a view fits well with the type of reaction based behaviours that DS theories claim are necessary to obtain maximal grip.

As well as emphasising a non-representational approach, intersubjectivist theories and embodied accounts of cognition also share a similar approach to the development of intelligent behaviour. We saw in the previous chapter that one way DS theories account for the development of directive behaviours is in terms of motor development, with the motor skills required for direction towards proximal objects acting as the foundation for the later ability of pointing towards distal objects. In a similar fashion, intersubjectivists argue that early skills such as imitation or distinguishing faces “underpin those developmentally later” (Gallagher & Zahavi, 189).

There are then, some broad similarities between understanding others in terms of Intersubjectivity and the action oriented approach of DS and AOR. As we shall now see however, it is intersubjectivist ideas about perception which are most helpful to AOR accounts.

### **Intersubjectivist Perception and Affordances**

Some intersubjectivists claim that the skills displayed by infants rest in part on an innate capacity to “directly perceive” the mental states of others. Trevarthen and Aitken, for example, write of infants being born "with *awareness* specifically receptive to subjective states in other persons" (2001, Page 4 my emphasis), and Gallagher and Zahavi claim:

The capabilities involved in primary intersubjectivity suggest that before we are in a position to wonder what the other person believes or desires, we already have specific *perceptual understandings* of what they feel, whether they are attending to us or not, whether their intentions are friendly or not, and so forth

(187, my emphasis)

The main proponent of what I shall call Intersubjectivist Perception (IP) is Shaun Gallagher (2008b, 2008c). He argues that instead of theorising or simulating to infer the mental state of another from their perceived behaviour, in most everyday cases we “have a direct perceptual grasp of the other person’s intentions, feelings, etc.”(Gallagher 2008a, 535). From a phenomenological point of view at

least, such an account is plausible. The cases where I observe your behaviour and consciously theorise ('He is doing X so is likely to believe Y and desire Z') or simulate ('If I were doing that then my mental state would be...'), are few and far between. In the majority of cases, what you are doing, wanting or attending to, are as immediately obvious to an observer as the colour of the clothes you are wearing. Although we should of course be wary of drawing definitive conclusions from phenomenological reflection, Gallagher's observations do raise the following question: Why should the processes involved in seeing colour be so different to those involved in seeing attention when from the first-person perspective they seem equally "direct"?

Our direct perception of colour can be explained (at least in part) by appealing to the physiology of the eye (different photoreceptor cones are sensitive to different segments of the visual spectrum), but in the case of mental states this option is not available -there are no 'sensors' in the eye that detect attentional or intentional states in others. Gallagher argues instead that IP depends upon an idea which we first saw in 1.3: the ability to recognise objects as presenting particular "affordances" for action:

I see the car not just as some object among others, but as an object that I can use —that I can climb into and drive... It affords certain kinds of action, and these affordances inform —are part and parcel— of my perceptual process. I see my car as drivable. This does not mean that I see my car, and then judge that it is drivable. A separate judgement is not required, unless, of course, the situation is unusual

(Gallagher, 537)

Gallagher proposes that our perceptual understanding of people can also be put in terms of affordances. In the same way that the kinds of action that a car affords are "part and parcel" of seeing it, the kinds of actions that another person affords are part and parcel of my seeing them. A friendly face, for example, might provide the possibilities of asking for assistance or comfort, which a frowning or scowling face would not. What affordances a particular face or object offers will depend on the details of its perceived features: how aerodynamic the body of the car appears will affect whether the vehicle offers affordances for driving fast or slow, small differences in the type of smile we see (the particular curve of the lips, whether the the eyes are 'smiling' as well as the mouth) so that it appears friendly or unfriendly, honest or insincere, will affect what opportunities it provides.

We can see then that IP ties action and perception very closely together. Perceiving an object depends up what Noë & Regan call: "Sensorimotor knowledge": knowledge of what sensory stimulation will be experienced if a particular bodily movement is made<sup>6</sup>. Seeing the car as drivable consists of knowing

---

6 Noë and O'Regan's theory of "enactive perception" is concerned with perception in general rather than perception of other people. However, the central claim of their theory -that perceptual experience depends upon knowledge of how sensory stimulation will change with movement- can be applied to IP.

that if I move my body in such a way so that I get inside the car and start the engine, I will experience sights, sounds and sensations typical of a car being started, seeing a point as followable depends on knowing that a head turning behaviour will be met with an interesting sight.

Parallels can be drawn between knowledge of how experience will change through movement can be linked with maximal grip explanations of intelligent action. Being able to respond to a sensation so as to be directed towards the situations optimal gestalt, is not far removed from knowing that *this* bodily movement will bring about *this* sensory experience. These connections shall be returned to shortly. Before this however, it will be useful to return to the differences between DS and AOR conceptions of affordances.

### **Dynamical Systems and Perceptual Affordances**

Gallagher would very likely endorse the DS approach to affordances as he considers his Intersubjectivist Perception to be “non-representational” (footnote 2, 537) and has elsewhere rejected explicitly AORs as useful in explaining intelligent action (Gallagher 2008). As we saw in 1.3, on the DS account perceiving the affordances offered by an object or person and then acting upon them is just another way of describing dynamic coupling with the environment. Certain perceptible features of an object (the steering wheel of a car, a smiling face) elicit learned responses from the brain's starting state. Possessing sensorimotor knowledge of a car or person would not be to have the capacity to *represent* the object in any way, action-oriented or otherwise. It would instead be to have a brain tuned so that it responds to such objects in a way that is appropriate to circumstance. Put in terms of embodied intentionality we could say that the recognition and acting upon an affordance is simply a case of a feature of the environment being discriminated (the perceptual component of the intentional arc), and consequently directing the body towards the optimal gestalt of the situation.

However, understood in these anti-representational terms the notion of affordances in perception adds little to DS accounts of understanding and so gets us no further towards solving the problems outlined at the end of Chapter 4. If the enaction of sensorimotor knowledge is considered just another case of dynamic coupling then it still suffers from the fact that even complex behaviours such as gaze checking must accounted for in terms of tuned response to the features of an object. Furthermore, explaining all intelligent interaction in terms of dynamic coupling, would appear to undermine Gallagher's claim that the *various* affordances an object offers are part of our perceiving it. On the DS view, which affordance of an object is acted upon will depend entirely on the starting state of the brain and the fine details of situation at hand. But if this is the case, then what can it mean to see the different possibilities offered by an object? In the example of the car Gallagher states that “It affords certain kinds of action, and these affordances inform —are part and parcel— of my perceptual process”. Seeing what affordances are on offer implies that different opportunities are recognised but only one is acted upon. Dynamical coupling however, would appear to rule this out: there is no presentation of possibilities to a subject and no choosing between them, there is only the

environmental feature being distinguished its directing the subject towards maximal grip. It is true to say that the same environmental feature can elicit different responses from an individual depending on their starting state (we saw that this was Freeman's position in 4.4. The stimuli caused by a car might elicit getting-in-and-driving-away behaviour from the starting state of someone who is on their way to work, on a Sunday afternoon however this same individual will receive the same stimuli from the car, but because his starting state is different, car washing behaviour will be caused instead. These, however, should not be considered choices: the starting state of an individual *pre-determines* how a perceived feature is responded to.

### **AOR and Perceptual Affordances**

A clearer and more useful account of perceptual affordances can be given if we modify Gallagher's account slightly and reconceive it in terms of anticipatory information and AORs. The perceived features of the object do not simply direct our behaviour, they cause a representation containing anticipatory information of what would happen (what sensory input we would receive) if various movements were made. I see a car as drivable because it causes a representation detailing the different sensory outcomes of various possible motor commands: 'Moving your body in manner X will be met with sensory outcome Y'. An AOR of a car would contain information such as 'Moving your body so that you reach forward and grab the handle of the car door will be met with the sensory information typical of opening a car door (the feeling of the hand moving through space and then touching the door handle, the sight of the hand reaching out towards it). The reason why we see the different affordances the car offers rather than blindly responding to just one of them, is because the AOR contains a 'list' of possible movements and possible outcomes.

If we conceive of IP in this way, we can begin to see how AOR might be given a fuller account of anticipatory information. In 5.1 I identified three elements of emulation and all can be put in terms of sensorimotor knowledge: The role of 'mock input' fed into the emulator can be played by motor commands (hypothetical turns of the head or pointing gestures), the 'mock output' consists of sensory information (what sensory stimulation would be experienced) and the role of the emulator itself can be played by the subject's sensorimotor knowledge. An AOR of a car would contain information regarding various motor commands that could be made in relation to the object, along with their hypothetical outcomes.

This type of information processing is quite different from that of RTM. At no point is there any representation that the world is, or will be, a certain way. In fact, the content of anticipatory information can be linked more closely to the motor skills explanation of maximal grip. We saw in the previous chapter that being able to obtain maximal grip on a situation depends upon the ability to discriminate and respond appropriately to features of the environment. Responding appropriately to a stimulus means to act in such a way that one moves closer to the optimal gestalt. However, as we have just seen in the discussion of DS and perceptual affordances, the same stimuli will not always meet the

same uniform response, what response is appropriate varies according to the situation: the same car stimuli can cause washing behaviour or driving behaviour, depending on the subject's starting state.

I want to suggest that the anticipatory information contained in a perceptual AOR, the 'list' of possible motor commands and their outcomes, is made up of movements that could be made, and have been made, in response to a particular stimulus. In other words, an AOR bears content about the totality of different ways that this stimulus has been responded to in the pursuit of an optimal gestalt. My AOR of the car contains the information that it presents opportunities for driving, opening, washing etc. because these actions have in the past allowed me to obtain a maximal grip on a situation.

This account of the relationship between embodied intentionality and AORs is quite speculative and not all together complete. My purpose, however, is to properly differentiate anticipatory information from descriptive information by showing that the sensorimotor knowledge which informs perceptual affordances is similar to, and perhaps the same type of thing, as the knowledge which allows us to get a maximal grip on a situation. Whether it is true to say that the ability to form AORs of an object *depends upon* the ability to obtain a maximal grip on situations involving that object, or whether it is in fact the other way around, will be discussed in the final chapter. Having said this, it still seems quite plausible to think that AORs and the skills needed for embodied intentionality are linked in some way: in most normal cases I would not see that the car door provides the affordance of being opened, if I was not able to open it myself.

## 5.4 AORs and Gaze Checking

Now that a proper account of information has been added to AOR, I will here suggest how the theory can be used to explain gaze checking. I will first of all present an outline of how GC1 and GC2 can be explained in terms of AOR, then discuss reasons why we might think of this account as providing more explanatory leverage than DS.

**GC1:** *Initially, the child's gaze lingered on the tester's outstretched hand. Next she looked back to the tester's face as if to ascertain what the tester might be trying to communicate. For a moment she seemed to dwell on her face and then suddenly she shifted her gaze to the target of the tester's still outstretched finger.*

When the infant perceives the initial point, the infant fails to be directed by it and instead begins to form an AOR of the affordances that could be made in response to it. This information would be based on the infant's previous experience obtaining maximal grip with similar stimuli and detail what is likely to be perceived if certain movements were made.

After this processing has taken place, the infant then turns to the face of the pointer. Once again, the stimulus does not cause the infant to immediately act so as to obtain maximal grip, but instead adds information to the infant's AOR. Consequently, the infant is now able to appreciate the affordances offered by the combination of the pointer's outstretched hand *and* the direction of their eyes. We might think of this in terms of an emulator being modified. Prior to seeing the face, the emulator contained anticipatory data about hand stimuli only. After seeing it, however, this information is updated and the emulator's output details the consequences of bodily movement in relation to *hand and eyes*.

One mock motor command (detailing a turn of the head in the same direction as the hand and eye's of the pointer) which is fed into the modified emulator yields the output that an interesting object will be seen. Having recognised this opportunity, the infant takes it.

**GC2:** ... a 12-month-old girl swiftly followed the tester's gaze and outstretched finger to locate a poster on the wall and then, giving herself barely enough time to take in the contents of the poster, quickly looked back to the tester's face with an engaging smile. The tester spontaneously commented on this sharing, with a playful and affirming "Did you see Big Bird?" and the child immediately turned, still smiling, to look back at the poster again.

The immediacy with which the initial point is responded to, suggests that this first movement is a simple dynamic coupling response. Having discriminated a point, the infant (automatically) attempts to gain a maximal grip on the situation. Upon perceiving the target object, however, an AOR is formed detailing the possible outcomes of various motor commands that could be made in response to it. The infant then turns to the pointer's face, causing more information to be added to the AOR. As with GC1, one of these affordances leads the infant back to the target object.

### **Explanatory Leverage**

These accounts are brief, and some of what follows will be mildly speculative, but here I will try to motivate the idea that this approach provides more explanatory leverage than the DS model.

The principle reason why this account might be thought compelling is that it provides an explanation of the actions suggestive of information processing (lingering and doubling back) but is at the same rooted in the infant's practical ability to obtain maximal grip. We saw in 5.3 that the sensorimotor knowledge required for perceptual affordances could be thought of as being grounded in the ability to respond intelligently to objects. Now, we can see that this approach is just informational enough to avoid the difficulties that DS suffered from when it comes to gaze checking.

In 4.4 we saw that the "lingering" on the hand and then the face in GC1 could be explained by information processing taking place on the infant's part. We now have a clearer idea of what this information processing is. On my account, the infant sees the outstretched finger and the forms an

AOR containing anticipatory information about possible movements and their responses. It seems plausible that the processing involved here -the listing of the perceptual outputs of possible inputs followed by the taking up of one affordance- would take time and so explain the gap in the infant's behaviour.

We saw also in 4.4 that the order of the infants' movements in GC1 and GC2 was better explained in terms of information than in embodied intentionality. Again, information conceived of in terms of AOR can take up this role. The reason why the infant does not turn towards the target object until it has seen the face is because until that point its AOR did not contain sufficient anticipatory information. It is only when the AOR's emulator has been suitably updated, that the infant becomes aware of the opportunity to see an interesting object.

My explanation of GC2 is slightly different. The infant returns to the target object because the speaker's friendly, interested attitude towards it adds new information to the AOR and consequently more opportunities to act. Upon seeing a poster of big bird the infant might appreciate that it affords opportunities of looking, touching or playing with. After having seen that the pointer also has a positive disposition towards the object, new affordances are available: the poster becomes not just something to play with by oneself, but also something that could be interacted with this interested person.

### **Representation Hungry Situations**

What I think is missing from the account I have presented is a reason why the infants' in both cases cease to dynamically couple and begin to represent instead. In 1.3 we saw that Clark and Grush consider AORs to be employed only some of the time, arguing that anticipatory information only comes into play in situations where "using real world feed back is impractical, absent, or when it "could improve real-world, real-time responsiveness" (8). As the relevant elements of the real world are not absent, nor is getting to them impractical, we must assume that they are invoked for the third reason: to enhance real-time responsiveness. This, to some extent, seems to be the case as we have just seen that AORs allow for a more skilled reaction to other people's points than would be provided just by dynamic coupling. However, what is not clear is how the infant comes to recognise that representations might enhance performance in a particular situation.

We can see this difficulty in both cases. In GC1, an outstretched hand would appear to be a relatively simple and unambiguous stimulus and so it is not clear why it would be met with a representation rather than an immediate attempt to gain maximal grip. After all, in GC2, the same stimulus is met with an immediate turn of the head, suggestive of a dynamic coupling explanation. The problem for GC2 is explaining why an AOR would be formed when the target object has already been found.

This is quite a serious difficulty for the theory as it stands, because if a adequate account of an AOR 'trigger' cannot be given there is a risk that AORs can be invoked any time that there is not an obvious, dynamic coupling-based, explanation of behaviour.

In the next chapter I will suggest some ways that we might go about solving this problem. For the time being however, we should note that the DS account of gaze checking would fair no better. DS would explain different responses made to the same stimulus as the product of the infant's ever changing starting state, as we saw in 4.4, this solution risks DS explanations of action becoming overly obscure and lessening the theory's explanatory value.

## **Conclusion**

We have seen in this chapter that AOR offers an account of attention understanding that, while not perfect, can be said to provide more explanatory leverage than either of the alternatives. We have seen that, through my modified version of Gallagher's theory, a suitably detailed conception of anticipatory information can be found. This can then be used to show how gaze checking behaviour can be accounted for.

Although this account is not perfect, questions can still be raised as to why AORs are initiated in some circumstances but not others, it is still able to account for lingering and doubling back behaviours better than non-informational theories.

# Conclusions

In this final chapter I shall present a summary of my findings, detail the conclusions that can be drawn from them, and provide two recommendations as to how they can be expanded upon in future research.

## Findings and Conclusions

### 1 Representational and Non-representational Cognition

In Chapter 1, I introduced RTM, DS and AOR as three different ways that our intelligent action with objects and people could be accounted for. The key difference between the three was the value that each placed on mental representation as a cause of intelligent interaction. RTM holds that a person's understanding of an object consists of their possessing a mental representation of it. This consists of descriptive information that the world is a certain way: that the bus is red and that the bus is heading towards us. This information allows us to infer what the best course of action to take would be so that we realise a goal state. Such an approach locates understanding in the brain, if a person understands what a bus is, then they will have the appropriate, representational, neural configuration which leads to appropriate, intelligent action.

In contrast, DS rejects representations as a way of explaining behaviour. Instead, a person's intelligent action is caused by fluid interaction between the brain, body and environment: perceived features of the environment, influence what movement is made, which leads to further perceptions, and further movement. The three elements of this process are considered equal partners in causing behaviour: the activity in the brain is not held to be more important than the body or environment as it is in RTM, and neither is it considered informational.

The final theory, AOR, was shown to occupy a halfway position between these two accounts. On this view, behaviour is very often caused solely by dynamic coupling with the environment, however, in cases where it would enhance intelligent action, a form of representation is employed. These representations differ from those of RTM in that they don't contain descriptive information about the state of the world but are instead "action oriented", detailing what opportunities for action a particular situation provides. Upon seeing a person or object we are able to anticipate what would happen were we to perform a particular movement.

## **2 Attention**

In Chapter 2 I surveyed various behaviours displayed by infants in their first year of life which suggest they have some understanding of attention in others. We saw that opinions differ amongst philosophers and cognitive scientists as to how these behaviours come about. Rich theories of attention understanding state that as well as possessing a 'causal' understanding -knowledge that a certain bodily movement will often elicit a response from the environment- infants also possess, or come to possess a 'mental' understanding that other people are capable of attending to aspects of the environment. Possession of this "concept of attention" is claimed to be the cause of at least some of an infant's AUBs. The opposite view is that attention understanding is 'lean': there is no concept of attention driving attention understanding behaviour. AUBs are not the product of a mental understanding of others but are caused solely by the infant's response to stimuli and causal understanding.

We saw that clear parallels could be made between these opposing conceptions of attention understanding and the split between representational and non-representational accounts of cognition. Rich conceptions of attention understanding can be understood as a variety of RTM: both theories hold that some type of informational/representational unit is the primary cause of intelligent action (and so the primary criteria for understanding). In the same way that RTM would hold that a 'bus' representation is the cause of intelligent interaction with a bus, rich conceptions of attention can be thought of as claiming that AUBs are caused by a representation of attention. Similarly, lean accounts of attention understanding can be considered a variety of DS explanation as both theories reject the idea that behaviour is caused by a unit of information. Common ground can also be seen between the lean theorist's claim that intelligent behaviour is no more than causal understanding and response, and the DS claim that action is no more than sophisticated response to stimuli.

We also saw in Chapter 2 that there is reason to doubt the claim made by some rich theorists that an infant's concept of attention only emerges at about 12 months of age. Tomasello and others argue that although behaviours suggestive of attention understanding are displayed by infants almost from birth, a concept of attention, and therefore a true understanding of attention, does not appear until the infant is able to display "joint attention behaviours" at around 12 months of age. They claim that earlier AUBs such as gaze following or returning a smile can be explained purely in terms of response and causal understanding. On the other hand, joint attention behaviours such as declarative pointing cannot be reduced to causal interactions and show a sophistication that strongly suggests a mental understanding of others. After examining arguments by Reddy however, these claims were found not to hold: The responsive AUBs displayed by younger infants were shown to be more complex than mere response to stimuli and closer examination of clowning behaviours at 8 months suggested that infants have some mental understanding earlier than 12 months

## **3 The Representational Theory of Mind**

In Chapter 3, I assessed how well the RTM framework could account for the empirical data on attention understanding. We saw that there are two possible ways that descriptive representations of attention can be understood: as “unobservable theoretical postulates”, or as simulatory information, based on our own experiences of attending to objects. Having suitably supplemented the theory, I turned to two aspects of AUB development: inconsistent gaze following at three months, and the development of directive behaviours between eight and 12 months.

The primary conclusion that drawn from the discussion was that RTM should be rejected as a means of explaining attention understanding. Cognitive scientists should not seek to define an infant's understanding of attention in terms of whether or not it possesses a descriptive representation of attention. The key difficulty the theory faced was that to explain the development of the two AUBs in question, it needed to appeal not just to the simulatory or theoretical information contained in the infant's representation of attention but also to the state of the infant's motor skills. This appeared to be a move that proponents of RTM are forced to make, but by doing so, much of the theory's explanatory value is compromised. It seems a reasonable requirement of a theory of understanding that it should be able to explain how intelligent action develops, but when faced with the questions of how gaze following becomes more consistent over time, or how an infant's ability to point to distal object emerges from its ability to gesture towards more proximal ones, appealing to inference, goal states or the processing of descriptive information was of little use. Instead, it appeared that the infant's motor abilities should be the main point of reference when seeking to explain the development of intelligent interaction, to the detriment of RTM.

Three further conclusions can be drawn from this failure. The first is that even though RTM was only discussed in terms of TT or ST, it seems likely that any theory which seeks to explain AUBs in terms of descriptive information will fail. The problems experienced by the two theories examined were not caused by their reliance on simulation or theorising *per Se*, but were instead caused by their emphasis on descriptive information in the brain rather than motor competence and the body. Descriptive information appears to be of little use when trying to explain how intelligent action develops, and we should therefore be wary of any theory that seeks to explain AUBs as the product of the infant's representation *that* the other has the property of attention, regardless of how such information is encoded.

Secondly, as well as scuppering RTM accounts of attention understanding these problems also weaken the theory more generally. First of all, there is the worry that ST and TT explanations of other aspects of an infant's developing theory of mind, such as its understanding of belief, desire or emotion, will suffer from the same problems regarding motor skills. If an infant acts in a manner that suggests it understands desire in another, for example giving someone a desirable object, worries again emerge as to how much of the action can be clearly said to be caused by the infants knowledge *that* the other desires the object, and how much by the infant's ability to control its own bodily movements.

Whether these arguments could also be made against RTM based explanations of belief understanding is more doubtful as here the situation is complicated by the presence of language. A normally developing infant's understanding of belief in others is commonly thought to emerge at around 4 years old as this is the age that most infants are able to pass the 'false belief task'. However, by this age, the child has some linguistic skills: in most versions of the experiment the child is shown a scene in which one character is and then asked "Where will Maxi look for the chocolate". Although the infant's response can be verbal or non-verbal (either saying or pointing towards where it thinks Maxi will look) it has at least sufficient language skills to understand the question. As I mentioned at the beginning of Chapter 2, language makes it more difficult to differentiate representational from non-representation cognition. Even supporters of strong anti-representational approaches to action do not deny that representational thought is possible once language arrives: as we saw in the previous chapter, Gallagher is an anti-representationalist regarding action but accepts that language allows for representational thought<sup>7</sup>. Having said this, accepting that representation is possible once language arrives does not necessarily help TT or ST specifically. If mental representation is derived from language, it is not clear that this type of thinking about people would bear any similarity to the type of processing postulated by TT and ST. Besides this, both theories are considered by their supporters to be the primordial and primary way which infants deal with others, arguing that simulation or theorising is only possible with language weakens this claim considerably.

The final point to mention is that the discussion of RTM's problems also backs up the general complaint made by supporters of EEC against RTM. As we saw in Chapter 1, philosophers like Clark, Dreyfus, Gallagher and Wheeler argue that RTM is ill-suited to explaining how any type intelligent of action is possible, the discussion in Chapter 3 shows that AUBs are no exception. On the face of it AUBs might look like prime candidates for an approach based on representations that the world is a particular way as, intuitively, we imagine that an understanding of attention consists of an understanding that there is a hidden mind, as it were, 'looking out from' the eyes of another. We have seen however that this is not the case and that an infant's knowledge of how to move its body so as to interact is much more useful in seeking explain intelligent behaviour. It would seem therefore, that my findings here strengthen the position of those that consider the RTM project to be misguided in cognitive science and a "degenerating research project" in AI.

#### **4 Dynamical Systems**

In Chapter 4, we saw that DS offers a stronger account of AUBs than RTM. Its main advantage was that whereas RTM locates understanding in the brain and so suffers when motor skills enter the picture, understanding as conceived by DS is spread over body and environment also and so explanations which rely on motor skills can be accommodated far more easily. This allowed for a more plausible account of inconsistent gaze following and the development of directive behaviours than that offered by RTM.

---

<sup>7</sup> A similar approach can be seen in Hutto 2008

We also saw in the discussion of embodied intentionality that they worries about DS's claim to be non-informational were unfounded. By explaining action in terms of the intentional arc, perception as discrimination, and maximal grip, DS is able to avoid the charge that starting states, dynamic coupling and optimal gestalts are simply disguised representational terminology.

The main conclusion to be drawn from this discussion is that DS approaches provide an entirely non-representational account of how some AUBs are possible. Those interested in how attention understanding arises should focus on the state of the infant's perceptual and motor capacities, rather than whether or not they possess a representation or concept of attention.

An infant's understanding should not be thought of as its possession of a fact about the world which allows it to make inferences; understanding that another is capable of attending is not equivalent to the scientist knowing what gravitational force is caused by a mass. Instead, attention understanding should be thought of as similar to practical bodily skills like walking, riding a bike, or playing a musical instrument. Unlike the gravity example, these abilities are not the product of a single cognitive insight, and we would instead think of them in terms of degrees of competence. A scientist's understanding of a gravitational force is binary, she either understands it or not; with practical skills however, an actor can have a higher or lower level of understanding, there can be novice, amateur and expert players of the piano.

This way of thinking about understanding can be used to by psychologists interested in AUBs. Rather than trying to divide behaviours into the merely responsive and those which show a true understanding, the degree of the infant's understanding should be assessed in terms of embodied intentionality. When asking why a 2 month old will respond when attention is directed towards the self but not when attention is directed elsewhere, we should look to the state of the infant's perceptual and motor skills. It could be that at 2 months there is an inability to easily discriminate the environment into individual objects, which prevents it from responding appropriately when someone's gaze is directed towards them. Alternatively, the infant could very well be able to discriminate adequately but is hampered by the fact that it cannot control its body sufficiently for it to be able to obtain a maximal grip on the situation.

As with my criticisms of RTM, these recommendations regarding DS and attention understanding can be applied more broadly to issues in developmental psychology and cognitive science. When seeking to explain infant behaviours that suggest an understanding of desire, emotion or belief, it would be instructive to focus on the infant's ability to obtain maximal grip on a situation. Equally, the same approach could be applied to intelligent action more generally and so again back up the more general claims of EEC. Perceptual and motor capacities of the type I have described can be used to explain intelligent action with features of the environment other than people.

However, when we turned to gaze checking behaviours we also saw that there are limits to this approach. Maximal grip explanations of action rely on the subject being directed by perceived aspects of the environment, but when behaviour involves pauses between movements, or a 'doubling back' to a previously perceived feature it is difficult to see how it can be thought of as being guided or pushed. Instead, gaze checking strongly suggests that the infant is involved in some sort of data gathering: the infant looks back at the face to check the direction of the pointers gaze, or their emotional attitude towards the object. With these problems in mind we turned in Chapter 5 to AOR.

## **5 Action Oriented Representations**

Chapter 5 began by noting that AOR is well placed to explain both the earlier AUBs that caused problems for RTM as well as the more sophisticated behaviours that weaken a purely DS approach. This is because proponents of AOR accept that some instances of intelligent behaviour are caused entirely by information-less, dynamical coupling with the environment which allows the theory to utilise the DS explanations of inconsistent gaze checking and the development of directive behaviours. On the other hand, AOR also claims that in some “representation hungry” situations information processing is required for intelligent action. As gaze checking resisted an explanation in terms of maximal grip and had features that suggested that some sort of information gathering or processing was taking place, it appeared a good candidate for an AOR account.

As with the other two theories however, the theory needed to be supplemented before we could determine how successful it would be. Just as RTM required TT or ST to properly account for descriptive information, AOR required a more detailed account of what is meant by 'anticipatory information'. This was achieved by turning to intersubjectivist accounts of other understanding. Like TT and ST, Intersubjectivity is conceived as a general theory of how we understand other minds; its supporters hope to explain the development of attention understanding, but also understanding of desire, belief or emotion as well.

As with embodied approaches to intentionality, Intersubjectivity places a good deal of importance on the infant's perceptual abilities. Gallagher argues that we directly perceive the mental states of others rather than supplementing our perceptions with theory or simulation. He justifies this by saying that our perception of others' mental states (what makes us see someone as desiring this or attending to that) consists of our appreciation of the affordances for action that they offer.

However, we saw that, although Gallagher is an anti-representationalist with regards to action, it is difficult to see how an appreciation of affordances could be explained purely in terms of feature detection and tuned response. A more promising avenue, one that allows for our perceptions to involve the recognition of the *multiple* ways that an object affords action, is to argue that in “representation hungry” situations perceiving an object causes an AOR to be formed which contains information regarding what would happen were we to make a particular bodily movement.

Thinking of IP in these terms allowed for a fuller account of how anticipatory information is encoded in AORs . Clark and Grush conceive of AORs as emulators which are fed mock input, process it, and produce mock output. With IP, we can say more clearly what fills these roles. Mock input fed into the emulator can be thought of as hypothetical combinations of muscle contractions, how the body needs to be moved so as to accomplish various tasks; with the role of mock output being played by the possible sensory states that these bodily movements would lead to.

The advantage of this approach is that it allows for information processing of the kind apparently required to explain gaze checking, while at the same time rooting AORs in the infant's non-representational, action-oriented know how. I suggested that which affordances (which hypothetical actions and perceptions) were contained in an AOR depended on how well maximal grip could be obtained with the object or person in question; it seems implausible that one could see a car as drivable without possessing the motor and perceptual skills with which to drive it. We saw in 3.4 that RTM's difficulties are rooted in the tension between representations as containing descriptive information about the world and motor skills explanations of development. Because the information of AOR is based on practical knowledge how to navigate the world, the theory does not suffer the same difficulties.

I then went on to show that this newly supplemented version of AOR offered sufficient explanatory leverage when applied to gaze checking. Although the account I gave had its drawbacks (the difficulty of saying with certainty which situations are representation hungry and which are not), we saw that the idea of representation as emulation could be used to provide an emulation based account of this AUB. The pauses between actions in GC1 could be explained as the product of the infant processing anticipatory information, discovering which motor commands will have which perceptual effects on the perceived stimulus. I also put forward a way in which the order of actions in GC1 and GC2 (why the infant turns from hand to face to target object, or from target object to face to target object) could be explained in terms of emulation and affordances.

In light of these findings I want to propose that in some cases, more explanatory leverage is gained through AOR than is through an entirely non-informational approach. Despite the success of maximal grip and the intentional arc in explaining some aspects of intelligent behaviour, philosophers and psychologists interested in AUBs, as well as action more generally, should not follow a doggedly anti-representational approach: sometimes the states of the brain involved in action should be considered information bearing.

In 1.2 I introduced the EEC claim that, in action, neither brain, body or environment should be explanatory privileged over the other two components. My findings here, while not quite showing that the brain should be 'privileged', do at least suggest that in some cases states of the brain should be explained in different -informational- terms to the body or environment. With gaze checking we can see that some states of the brain are not best thought of as being part of the dynamic 'flow' between

individual and world. Instead, pauses and doubling back behaviour are better explained in terms of neural configurations bearing anticipatory content.

### **There are two intermingled forms of attention understanding**

When the claims of a 'joint attention breakthrough' were discussed in Chapter 2, I concluded that they should be rejected and that there was no good reason to think that infants only acquire a “concept of attention” at 12 months. However, from the discussion of AOR it would seem that the claims of Tomasello and Eilan were half-right, a distinction can be made between AUBs caused entirely through dynamical coupling and those caused by the processing of anticipatory information. As I shall go on to discuss in the next section, those interested in how attention understanding in infants develops should focus on this divide in behaviour, rather than on whether some or all behaviours are caused by a descriptive representation of attention.

Having said this, we should be careful not to over emphasise the difference between action caused by representational and non-representational means. It might help here to not think of entire acts (the whole series of actions which take place in GC1, for example) as being caused entirely by either AOR or DS. Instead we should focus on the the individual movements which compose them. As we saw in GC2, the movements of one 'act' of gaze following can be caused by a mixture of dynamic coupling and AORs.

This intermingling of representational and non-representational action goes against the ideas that DS supporters have regarding the split between action and representation. Gallagher, for example, is of the view that there is a clear line between arm chair based 'offline' speculation and fluid, 'online' interaction with the world, the implication being that once we are dynamically coupled with an object no planning, imagining, thinking etc. takes place:

I also want to set aside the question of whether representation plays a role in deliberation about action, or the planning of action, or the working out of prior intentions, and so forth. The question I address is whether representation is necessary in action, as part of action itself

(Gallagher 2008)

Here we see that Gallagher considers there to be a clear distinction between action and deliberation. As mentioned, he does not reject the notion of representation in cognition absolutely, arguing instead that representation is possible once language has been acquired. What he does disagree with though, is that representation can be “part of action itself”: when we are in the midst of interacting with the environment, representation is not involved.

Our discussion of AORs and gaze checking shows gives us reason to reject this claim. We have seen that something like deliberation, planning or working out is present in even relatively simple acts. The planning of action (the choosing which affordance to take up) takes place while we are acting in the world, not just when we are in the armchair reflecting about what to do. We have already seen this mingling of dynamic coupling and representation in the pauses and 'double takes' of gaze checking, but the same could be said for AORs' role in action more generally. Even when we engage in activities like walking or playing a musical instrument, actions likely to depend mainly on our ability to dynamically couple with the environment, it would seem that our fluid interaction is still punctuated with moments of reflection. For example, I might have practised a piece on the guitar a sufficient number of times so that I can for the most part play it 'in my sleep'. However, at one point there is a particularly tricky chord change, for which I need to concentrate a little harder, to reflect upon where my fingers are on the fretboard, and where they are going. It is here that AORs could play a role similar to the one they play in gaze following: an anticipation of exactly what the results of particular movements of my hand will be.

## **Recommendations for Future Research**

We have seen that that AOR provides the most plausible account of attention understanding. Some AUBs are caused entirely by dynamic coupling whereas others require the processing of anticipatory information. However, this account is in some ways incomplete. I will outline here two possible avenues of research that could strengthen AOR explanations of AUBs, as well as of action more generally.

The first concerns the difficulties mentioned towards the end of 4.4 regarding the difficulty in saying exactly why some situations elicit AORs and others do not. The second is concerned with the exact relation between AORs and the skills needed to acquire a maximal grip on a situation.

### **What causes AORs?**

We saw at the end of Chapter 5 that a difficulty for my account of AOR gaze following is that we are as yet unable to say why some aspects of gaze checking cause a representation to be formed. Although we have seen that some in some instances emulation does enhance intelligent interaction, I have not been able to identify the dividing line between situations that require, and so elicit AORs, and those in which intelligent behaviour can be accounted for purely in terms of dynamic coupling. The root of this worry is that AOR explanation risks being invoked on an *ad hoc* basis, used to account for any behaviour which cannot obviously be attributed to perceptual skills or motor response. There are two ways in which a more definitive answer to the question might be found.

Firstly, we could ask whether there is any distinctive perceptual features of GC1 and GC2 that switch the infant's behaviour from 'fluid interaction mode' to 'representation and reflection mode'. In the case of GC1, there could be some feature of the initial point which is detected by the infant and causes it to cease coupling and start representing. Perhaps the gesture was ambiguous, or only half seen and so an AOR is invoked to provide assistance. In GC2 there could be some feature of the target object which causes the shift in behaviour. Perhaps some feature of the picture of big bird -that it is brightly coloured or familiar- which initiates an AOR.

One way that this hypothesis could be tested for is through a careful study of how infant's interact with people as well as other features of the environment. The aim of this would be to discover whether there is a common factor (or factors) in all situations which elicit non-maximal grip behaviours like pausing or doubling back .

An alternative explanation of representation hungry situations might be found by concentrating not on aspects of the situation itself, but instead on what starting state the infant brings to a situation. This idea has some plausibility because if we consider GC1 and GC2 we can see that in one case a point initiates an AOR, but in the other the same stimulus (a pointing finger) initiates dynamic coupling instead. If we accept Freeman's claim that the starting states of our brains are in a constant state of flux, we could say that in GC1 the infant had a brain state in which a point initiates an AOR but in GC2 it has a different brain state and so responds dynamically.

The advantage of this approach would be that it would explain how similar stimuli can initiate different responses, something which the perceptual cues explanation I have just put forward cannot do. The disadvantage of this approach is that it would be very difficult to test for difficult it would be to test for. If an infant enters every new situation set to respond to to familiar stimuli in novel ways, it would be much more difficult to predict when and when not AOR would be elicited.

### **What is the link between Maximal Grip and Action Oriented Representation?**

Another interesting line of research might be to see what type of relationship there is between an infant's ability to dynamically couple with the environment and its ability to form AORs. In Chapter 5 I argued that the sensorimotor knowledge needed to appreciate the affordances offered by an object was linked to the ability to obtain a maximal grip on that same object, but stopped short, of saying exactly what this link is. It seem that there are three possibilities available:

**AOR1:** Once an infant has honed its maximal grip abilities sufficient for it to perform a particular type of dynamic coupling, it is able to form AORs relating to that action.

**AOR2:** The infant's ability to dynamically couple with the environment depends upon its first being able to form AORs. First, the infant is able to reflect on what options a person affords, then it is able to dynamically couple with that person.

**AOR3:** At some point in the first year of life the infant gains the ability to form AORs generally. All behaviours (regarding attention or otherwise) are purely dynamic until a certain age.

Which of these options is correct can be determined by further investigation into the development of infants' AUBs. In my discussion of gaze checking, I identified two ways which AOR driven behaviours might be recognised: a lack of fluid response to aspects of the situation, and the presence of movements such as 'doubling back' both of which resist explanation in terms directed intentionality. However, this does not mean that the ability to form AORs only emerges with gaze checking. It could be that earlier behaviours displayed by the infant also involve pausing or doubling back, but these features of behaviour have so far been neglected by theorists. By re-examining the development of an infant's intelligent action and using the presence or absence of these features to categorise AUBs as either dynamical or representational, we can come closer to determining the relationship between AORs and dynamical coupling.

In practice this would work as follows. If after examining the data we see that infants first achieve competence in fluidly reacting to another's action and only later see them pause before responding, we might conclude that something like AOR1 is true. We could for example, investigate whether AORs are present in early gaze following behaviours. We saw in Chapters 2 and 3 that at 3 months, infant's are statistically more likely to look at an object if a pair of eyes has looked in that direction first. Upon looking more closely, we might find that when the infant does follow the gaze, its response is immediate and fluid. If we then found that it was only later that the infant pauses before following a gaze redirection, we might conclude that the AOR is somehow derived from the infant's ability to obtain maximal grip.

If on the other hand, instances of pausing before following the gaze of another were seen to occur earlier than fluid response, we might instead conclude that AOR2 is correct. The infant's earlier behaviour is caused by its representing the affordances offered by a stimulus, and its learning how to fluently move towards the situation's optimal gestalt, depends upon this. If this was found to be the case then it would have interesting implications for EEC. On the one hand it would seem to count against the EEC claim that cognition for action precedes the ability to represent. DS holds that dynamic coupling is our default way of acting in the world, and we also saw in Chapter 5 that intersubjectivists consider the "sharing of experiences" to take place on a non-cognitive level first and foremost. On the other hand this conclusion would appear to fit with remarks made by Dreyfus (2002) that when we are learning a new skill, we initially learn it in terms of rules:

The student automobile driver learns to recognize such interpretation-free features as speed (indicated by the speedometer) and is given rules such as shift to second when the speedometer needle points to ten miles an hour.

(368)

and is only later able to competently acquire a maximal grip:

The expert not only sees what needs to be achieved: thanks to a vast repertoire of situational discriminations he sees how to achieve his goal... Among many situations, all seen as similar with respect to plan or perspective, the expert has learned to distinguish those situations requiring one action from those demanding another.

(371)

Here, Dreyfus suggests that when adults learn a new practical skill they start by inferring what the best thing to do is by thinking in terms of rules and goal states and it is from this that the ability to find a maximal grip arises. However, it does not seem likely that Dreyfus would consider AUBs to be a parallel case. Although he does not say so explicitly in this paper, Dreyfus, like other DS theorists, would hold that dynamic coupling is the default, primordial way of acting in the world and that an appreciation of rules and goal states only emerges with language. Nevertheless, if AOR2 was found to be true, and representation does precede the skills of embodied intentionality, then Dreyfus's ideas might be used as a basis for how development is possible.

The final possibility is that none of the infant's behaviours (AUBs or more generally) before a certain age are found to involve pausing or doubling back. This would imply AOR3, that there is no direct relationship between being able to dynamically couple with a person or object and being able to form an AOR detailing those possibilities. Instead, the ability to form AORs emerges as a more general skill which is then applied to the behaviours involved in attention understanding. Rather than there being a 'Joint Attention Breakthrough' as Tomasello et al. suggest, there might instead be something like an AOR breakthrough.

## **End**

In this thesis, I have attempted to show the best way that our understanding of attention can be understood. We have seen that what might be thought of as the traditional view, the RTM conception of cognition, struggles when it is applied. Embodied approaches, on the other hand, provide a better account of how AUBs develop from relatively simple responses to stimuli, to more complex interactions involving infant, other and object.

Despite the failure of RTM, we have seen that the idea of representation in action should not be given up entirely. It has been shown that there is a limit to what behaviours can be explained in terms of DS and maximal grip and, in some cases, AOR accounts can provide greater explanatory leverage than non-informational accounts.

At the beginning of Chapter 2 one of the reasons I gave for focusing on attention understanding rather than understanding of any other kind was that it allows the quite abstract theorising about action and understanding to be applied to concrete empirical data. With the two recommendations that I have made in this final chapter I hope to have shown how my findings might be used to inform future empirical research into infants' understanding of attention, as well as how intelligent action is possible more generally.

## Bibliography

- Astington, J. "What Is Theoretical About a Child's Theory of Mind?: A Vygotskian View of Its Development" in *Theories of Theories of Mind*. Ed. Peter Carruthers: Cambridge UP, 1996.
- Baron-Cohen, S. *Mindblindness: an Essay on Autism and Theory of Mind*. Cambridge, Mass.: MIT, 1995
- Brazelton, T. B. "The Development of Newborn Behavior." *Human Growth: a Comprehensive Treatise*. Ed. F. T. Falkner and J. M. Tanner. New York: Plenum, 1986. 519-40.
- Brentano, F.C., Oskar, K. and L. L. McAlister. *Psychology from an Empirical Standpoint*; . London: Routledge and Kegan Paul, (1874/1911/1973)
- Buccino G., Lui F., Canessa N., Patteri I., Lagravinese G., Benuzzi F., Porro C.A., Rizzolatti G. "Neural Circuits Involved in the Recognition of Actions Performed by Non-conspecifics: an FMRI Study." *Journal of Cognitive Neuroscience* 16.1 (2004): 114-26. Web.
- Butterworth, J., & Jarret, N. "What Minds Have in Common Is Space." *British Journal of Developmental Psychology* 9 (1991): 55-72. Web.
- Calder,A.J.,Keane,J.,Manes,F.,Antoun,N.,and Young, A.W"Impaired Recognition and Experience of Disgust following Brain Injury." *Nature Neuroscience* 3 (2000): 1077-1078.
- Carpenter, M. Nagell, K. Tomasello, M. "Social Cognition, Joint Attention, and Communicative Competence From 9 to 15 Months of Age." *MONOGRAPHS- SOCIETY FOR RESEARCH IN CHILD DEVELOPMENT* 255th ser. 63.4 (1998): 1-143.
- Carruthers, P."Simulation and Self-knowledge: A Defence of Theory-theory." *Theories of Theories of Mind*. Ed. Peter Carruthers: Cambridge UP, 1996. 22-38.
- Clark, A. *Being There Putting Brain, Body, and World Together Again*. Cambridge, Mass.: MIT, 1998
- Clark, A. and R. Grush"Towards a Cognitive Robotics." *Adaptive Behavior* 7.1 (1999).
- Corkum, V., and Moore, C."The Origin of Joint Visual Attention in Infants." *Developmental Psychology* 34.1 (1998): 28-38.
- Damasio, Antonio R. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace, 1999.
- Dennett, D. "The Practical Requirements for Making a Conscious Robot." *Royal Society Homepage Philosophical Transactions of the Royal Society Series A* 1689 (1994): 133-46.
- Descartes, R. *The Philosophical Writings of Descartes*. Cambridge [USA.: Cambridge Unix., 1991
- Dreyfus, H L. *Being-in-the-world: a Commentary on Heidelberg Being and Time, Division I*. Cambridge, Mass.: MIT, 1991.
- Dreyfus, H L. *What Computers Still Can't Do: a Critique of Artificial Reason*. Cambridge, Mass.: MIT, 1992

- Dreyfus, H. (2002) "Intelligence Without Representation: Merleau-Ponty's Critique of Mental Representation." *Phenomenology and the Cognitive Sciences* 1.4 (2002): 367-83
- Dreyfus, H. "Why Heideggerian AI Failed and How Fixing It Would Require Making It More Heideggerian." *Philosophical Psychology* 20.2 (2007): 247-68.
- Eilan, N. "Joint Attention, Communication, and Mind" *Joint Attention - Communication and Other Minds: Issues in Philosophy and Psychology*. Ed. N. Eilan, C. Hoerl, T. McCormack, and J. Roessler. Oxford: Oxford UP, 2005. 85-109
- Freeman W. J. & Grajski K.A. "Relation of Olfactory EEG to Behavior: Factor Analysis,." *Behavioral Neuroscience* 101 (1987): 766-77.
- Freeman, Walter J. *How Brains Make up Their Minds*. New York: Columbia UP, 2001.
- Fodor, J A. *The Language of Thought*. New York: Crowell, 1975
- Fodor, J A. *The Modularity of Mind: an Essay on Faculty Psychology*. Cambridge, Mass.: MIT, 1983
- Fodor, J. A. *Connections and Symbols*. By Steven Pinker. Ed. Jacques Mehler and Steven Pinker. Cambridge, Mass.: MIT, 1988
- Gallagher, S. and Hutto, D. (2008): Understanding others through primary interaction and narrative practice. In: Zlatev, Racine, Sinha and Itkonen (eds). *The Shared Mind: Perspectives on Intersubjectivity* (17-38). Amsterdam: John Benjamins
- Gallagher, S, and Zahavi, D. *The Phenomenological Mind: an Introduction to Philosophy of Mind and Cognitive Science*. London: Routledge, 2008.
- Gallagher, S. "Are Minimal Representations Still Representations." *International Journal of Philosophical Studies* 16.3 (2008a): 351-69.
- Gallagher, S. "Direct Perception in the Intersubjective Context." *Consciousness and Cognition* 17.2 (2008b).
- Gallagher, S. "Intersubjectivity in Perception." *Continental Philosophy Review* 41.2 (2008c): 163-78.
- Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. "Action Recognition in the Premotor Cortex." *Brain* 119.2 (1996): 593-609
- Gallese, V. Goldman, A. "Mirror Neurons and the Simulation Theory of Mind-Reading." *Trends in Cognitive Sciences* 2.12 (1998): 493-501
- Gallese, V. "Embodied Simulation: From Neurons to Phenomenal Experience." *Phenomenology and the Cognitive Sciences* 4.1 (2005): 23-
- Gallese, V. "Being like Me': Self-other Identity, Mirror Neurons and Empathy." *Perspectives on Imitation: from Neuroscience to Social Science*. Ed. S. L. Hurley and N. Chater. Cambridge, Mass.: MIT, 2005. 101-18.
- Van Gelder, T. "Dynamics and Cognition." *Mind Design II: Philosophy, Psychology, Artificial Intelligence*. Ed. John Haugeland. Cambridge, Mass.: MIT, 1997. 439-48.

- Gibson, James Jerome. *The Ecological Approach to Visual Perception*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1986.
- Goldman, A.I. "In Defense of the Simulation Theory." *Mind & Language* 7.1-2 (1992): 104-19.
- Goldman, Alvin I. *Simulating Minds: the Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford: Oxford UP, 2006
- Gopnik, A. "Finding Our Inner Scientist." *Daedalus* 133 (2004): 21-28.
- Gopnik, Alison, and Andrew N. Meltzoff. *Words, Thoughts, and Theories*. Cambridge, Mass.: MIT, 1997
- Gordon, R. M. *Mind and Cognition: an Anthology*. Ed. William G. Lycan. Malden, Mass.: Blackwell, 1999
- Gordon, R.M. "Beyond Mindreading." *Philosophical Explorations* 11.3 (2008): 219-22.
- Hains, S.M.J., & Muir, D.W. "Infant Sensitivity to Adult Eye Direction." *Child Development* 67.5 (1996): 1940-1951
- Haywood, Kathleen, and Nancy Getchell. *Life Span Motor Development*. Champaign, IL: Human Kinetics, 2005.
- Heidegger, Martin. *Being and Time*. London: Blackwell, 1967
- Hobson, P., and J. A. Hobson. "Introduction." *The Shared Mind: Perspectives on Intersubjectivity*. Ed. Zlatev, J., Racine, T. P. Sinha, C. and Itkonen, J. Amsterdam: John Benjamins Pub., 2008. Print.
- Hobson, R. P. "What Puts the Jointness into Joint Attention?" *Joint Attention - Communication and Other Minds: Issues in Philosophy and Psychology*. Ed. N. Eilan, C. Hoerl, T. McCormack, and J. Roessler. Oxford: Oxford UP, 2005. 185-204.
- Hood, B.M., Willen, J.D. & Driver, J. "Adult's Eyes Trigger Shifts of Visual Attention in Human Infants." *Psychological Science* 9.2 (1998): 131-34.
- Husserl, E. *Logical Investigations*, Dermot Moran Edition: 2 Routledge. (2001)
- Hutto, Daniel D. *Folk Psychological Narratives: the Sociocultural Basis of Understanding Reasons*. Cambridge, Mass.: MIT, 2008
- Johnson, M. *The Meaning of the Body: Aesthetics of Human Understanding*. Chicago: University of Chicago, 2007
- Johnson, S. C. et al. (1998). "Whose Gaze Will Infants Follow? The Elicitation of Gaze-following in 12-month-old Infants." *Developmental Science* 1.2 (2002): 233-38
- Johnson, S.C. "The Recognition of Mentalistic Agents in Infancy." *Trends in Cognitive Sciences* 4.1 (2000): 22-28.
- Kosslyn, Stephen Michael. *Image and Mind*. Cambridge, Mass.: Harvard UP, 1980.
- Lenat, D., and E. Feigenbaum. "On the Thresholds of Knowledge." *Foundations of Artificial Intelligence*. Ed. David Kirsh. Cambridge, Mass.: MIT, 1992

- Leslie, A., and T. German. "Knowledge and Ability in "theory of Mind": One Eyed Overview of a Debate." *Mental Simulation: Evaluations and Applications*. Ed. Tony Stone and Martin Davies. Oxford, UK: Blackwell, 1995. 123-50
- Liszkowski, U., Carpenter, M., Henning, A., Striano, T., & Tomasello, M. "12-month-olds Point to Share Attention and Interest." *Developmental Science* 7.3 (2004): 297-307.
- Liszkowski, U. (2005). "Human Twelve-month-olds Point Co-operatively to Share Interest with and Provide Information for a Communicative Partner." *Gesture* 5.1/2 (2005): 135-54.
- Locke, John. *An Essay concerning Human Understanding*. London: Tegg and, 1838
- Merleau-Ponty, Maurice. *Phenomenology of Perception*, Translated by Colin Smith, Edition: 2, illustrated, reprint, Routledge, 2002
- Minsky, Marvin Lee. *The Society of Mind*. New York: Simon and Schuster, 1985
- Mueller, U., & Carpendale, J. I. M. (2004). From joint activity to joint attention: A relational approach to social development in infancy. In J. I. M. Carpendale, & U. Müller, (Eds.), *Social interaction and the development of knowledge* (pp. 215-238). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Muir, D., and S. Hains. "Young Infants' Perception of Adult Intentionality." *Early Social Cognition: Understanding Others in the First Months of Life*. Ed. Philippe Rochat. Mahwah, N.J.: Lawrence Erlbaum Associates, 1999. 155
- Moore, C., and D'Entremont, B. "Developmental Changes in Pointing as a Function of Attentional Focus." *Journal of Cognition and Development* 2.2 (2001): 109-29.
- Noë, A. *Action in Perception*. Cambridge, Mass.: MIT, 2004.
- Pinker, Steven. *How the Mind Works*. London: Penguin, 1999
- Pinker, Steven. *The Blank Slate : the Modern Denial of Human Nature*. New York: Viking, 2002
- Racine, T. P., & Carpendale, J. I. M. "The Role of Shared Practice in Joint Attention." *British Journal of Developmental Psychology* 25.1 (2007): 3
- Reddy, V. "Playing with Others' Expectations: Teasing, Joking and Mucking about in the first Year." *Natural Theories of Mind: Evolution, Development, and Simulation of Everyday Mindreading*. Ed. Andrew Whiten. Oxford, UK: B. Blackwell, 1991. 143-58.
- Reddy, V. Barrett, Martyn D. "Pre-linguistic Communication. In Language Development." *The Development of Language*. Hove, East Sussex, UK: Psychology, 1999. 25-50.
- Reddy, V. "Coyness in Early Infancy." *Developmental Science* 3.2 (2001): 186-92.
- Reddy, V. "Infant Clowns: the Interpersonal Creation of Humour in Infancy." *Enfance Paris* 3 (2001): 247-56.
- Reddy, V. "Before the "third Element": Understanding Attention to the Self." *Joint Attention - Communication and Other Minds: Issues in Philosophy and Psychology*. Ed. N. Eilan, C. Hoerl, T. McCormack, and J. Roessler. Oxford: Oxford UP, 2005. 85-109

- Rizzolatti G, Fadiga L, Fogassi L, Gallese V. "Premotor Cortex and the Recognition of Motor Actions." *Cognitive Brain Research* 3.2 (1996): 131-41.
- Rowlands, Mark. *Body Language: Representation in Action*. Cambridge, Mass.: MIT, 2006.
- Scholl, B. J. & Leslie, Alan M. "Modularity, Development and 'Theory of Mind'" *Mind & Language* 14.1 (1999): 131-53.
- Scholl, B.J. and Tremoulet, P.D. "Perceptual Causality and Animacy." *Trends in Cognitive Sciences* 4.8 (2000): 299-309.
- Thompson, Evan. *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge (Mass.): Harvard UP, 2007.
- Tomasello, M., Carpenter, M., & Liszkowski, U. "A New Look at Infant Pointing." *Child Development* 78.3 (2007): 705-22.
- Tomasello, Michael. *The Cultural Origins of Human Cognition*. Cambridge, Mass.: Harvard UP, 1999
- Trevarthen, C., & Aitken, K. J. "Infant Intersubjectivity: Research, Theory, and Clinical Applications." *Journal of Child Psychology and Psychiatry* 42.1 (2001): 3-4
- Trevarthen, C., and Hubley, P. "Secondary Intersubjectivity: Confidence, Confiding and Acts of Meaning in the First Year." *Action, Gesture, and Symbol: the Emergence of Language*. Ed. Andrew Lock. London: Academic, 1978. 183-229
- Trevarthen, C. "Communication and Cooperation in Early Infancy: a Description of Primary Intersubjectivity." *Before Speech: the Beginning of Interpersonal Communication*. Ed. Margaret Bullowa. Cambridge: Cambridge UP, 1979. 321-47.
- Tye, Michael. *The Imagery Debate*. Cambridge, Mass.: MIT, 1991
- Umiltà, M. A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., and Rizzolatti, G. "I Know What You Are Doing: A Neurophysiological Study." *Neuron* 31.1 (2001): 155-65.
- Vereijken, B. "Motor Development." *The Cambridge Encyclopedia of Child Development*. Ed. B. Hopkins. Cambridge, UK: Cambridge UP, 2005. 217-27.
- Wheeler, Michael. *Reconstructing the Cognitive World: the next Step*. Cambridge, Mass.: MIT, 2005.
- Wittgenstein, Ludwig. *Philosophical Investigations*. Oxford, Eng.: Basil Blackwell, 1967
- "Why do sunflowers turn towards the sun?", *Mr Smartyplants*, 27 February 2006, Web. Accessed 1 May 2010. <<http://www.wildflower.org/expert/show.php?id=589> >
- Zlatev, J., Racine, T. P. Sinha, C. and Itkonen, J. "Introduction." *The Shared Mind: Perspectives on Intersubjectivity*. Ed. Zlatev, J., Racine, T. P. Sinha, C. and Itkonen, J. Amsterdam: John Benjamins Pub., 2008. Print.