Mind in practice : a pragmatic and interdisciplinary account of intersubjectivity

Bruin, L.C. de

3.

Beyond the Problem of the Other Mind

The essential implications of Cartesianism for the modern self might be summed up in two words: disengagement and reflexivity.

- Sass 1992

What lurks below

In the previous chapters I have pointed out some of the internal problems with TT and ST explanations of our everyday encounters with others - problems that appear when one uncritically accepts certain assumptions about social interaction. To a large extent, these assumptions are rooted in a very influential picture of intersubjectivity that was proposed by Descartes, became problematic during the rise of British empiricism, and eventually gave birth to the problem of the other mind. The aim of this chapter is to uncover and challenge this picture of intersubjectivity.

I start by introducing the historical background of three important assumptions that have become orthodoxy for contemporary TT and ST approaches (section 1). In the first place, this is the idea that our meetings with other minds are intrinsically *problematic*, since they are deeply infused by a Cartesian phenomenology of uncertainty. Secondly, by accepting the problem of the other mind at face value, TT and ST also accept a certain conception of the mind: as a self-centered, disembodied and disembedded entity. Thirdly, they assume that our doubts about other minds can be overcome by a conscious, cognitive process - a stepwise procedure initiated by a hyper-reflexive agent.

The chapter then continues by discussing what I, following Hurley (2008), call the 'sandwich' model of intersubjectivity, according to which this conscious, cognitive process necessarily intervenes between our perception of the bodily behavior of other persons and our interaction with their minds (section 2). For ST, this intervention involves some version of the argument from analogy: since I know my own mind and how it relates to my body, I am able to infer that this is also true for the other on the basis of an *analogy* between our bodies. TT rejects the analogy in the argument from analogy, but it retains the inferential element. It claims that we understand others by *inferring* the contents of their minds on the basis of a theory. I will take a closer look at three components of these action-perception interventions: introspection, inference and mental concept mastery (section 3), and claim that they are problematic insofar ST and TT try to construe them as internal capacities of the individual mind.

In the final part of this chapter, I challenge what I take to be at the core of the picture of intersubjectivity presupposed by ST and TT. This is the assumption that we are normally at a theoretical remove from other people, and have to adopt a theoretical attitude towards them for the purposes of prediction, explanation and control. Instead of taking such a third-person approach as the hallmark feature of our intersubjective engagements, I propose that our meetings with other minds are primarily rooted in *second-person interactions* (section 4).

3.1 The problem of the other mind

The Cartesian picture of the mind

In order to get a clearer view of the problems troubling TT and ST, we have to address the deeper assumptions that they have in common. Gallagher (2004) argues that both positions share two important presuppositions: the 'mentalistic supposition' and the 'supposition of universality'. I take the mentalistic supposition as the starting point for my diagnosis:

Supposition 1 (the mentalistic supposition): 'The problem of intersubjectivity is precisely the problem of other minds. That is, the problem is to explain how we can access the minds of

others. This is a problem of access because other minds are hidden away, closed in, behind the overt behavior that we can see. This is a mentalistic and clearly Cartesian supposition about the very nature of what we call the mind. The mind is conceived as an inner realm, in contrast to behavior, which is external and observable, and which borrows its intentionality from the mental states that control it. Both TT and ST set the problem as one of gaining access to other minds, and their explanations of social cognition are framed in precisely these terms' (p.200).

To start with, we have to restrict the scope of Gallagher's claim. It is true that those TT, ST and hybrid TT/ST positions that explain intersubjectivity in terms of *mindreading* (understood as the structural attribution of mental states such as beliefs and desires) are committed to the mentalistic supposition. However, as we saw in the previous chapter, there are also lightweight versions of ST that discard the traditional ingredients of mindreading (e.g., Gordon) and/or stress the embodied nature of our understanding of others (e.g., Gallese). Since these ST approaches often explicitly reject the mentalistic supposition and many of the other assumptions that will be discussed below, they are not the target of my criticism. At the same time, however, it is not always clear whether these positions are best interpreted in terms of *simulation*. This is why I will not group them under the general header of ST. Instead, I use this label primarily to refer to the classic simulation approaches that revolve around a robust notion of mindreading.

Returning to the topic at hand, Gallagher is right that the mentalistic supposition is a Cartesian supposition. But this is certainly not the whole story. For Descartes was not yet troubled by worries about the minds of his fellow human beings. It was only against the background of British empiricism that the problem of the other mind was recognized as an 'official' philosophical problem. The genealogy of this problem warrants more detailed investigation, since it might give us a clue as to where we should look for a solution. Let us therefore briefly consider the Cartesian picture of the mind.

Descartes is well-known for his quest for certainty. What is remarkable about this quest is that it begins with a method of radical *doubt*. Descartes writes that this method imitates that of the architect. 'When an architect wants to build a house which is stable on ground where there is a sandy topsoil over underlying rock, or clay, or some other firm base, he begins by digging out a set of trenches from which he removes the sand, and anything resting on or mixed in with the sand, so that he can lay his foundations on firm

soil. In the same way, I began by taking everything that was doubtful and throwing it out, like sand' (Replies 7, AT VII 537).[32]

The Cartesian method of doubt requires a highly reflexive attitude of *disengagement*, since we can only avoid mistakes and achieve certainty if we suspend our judgment and 'hold back', assenting only to that which we can clearly and distinctly perceive to be true. This leads Descartes to the conclusion that sky and earth, colours and sounds, and in fact all external things are nothing better than the illusions of dreams, and he even comes to consider himself as without hands, eyes, or any of the senses, and as falsely believing that he is in possession of these. Eventually, however, he manages to find something that lies beyond all doubt. This is the famous 'cogito ergo sum'.[33] It is important to notice that the cogito (the 'I think') is not the result of an *inference*.[34] Instead, it is recognized by an inner awareness - a simple and immediate act of clear and distinct perception. The cogito is the unifier of all modes of thinking (doubting, dreaming, understanding, willing etc.) and it provides Descartes with a foundation upon which to build further: 'Archimedes used to demand just one firm and immovable point in order to shift the entire earth; so I too can hope for great things if I manage to find just one thing, however slight, that is certain and unshakable' (AT VII 24, CSM II 16). This, however, comes at a steep price. For the cogito is identified as an immaterial and timeless substance, and radically cut off from body and world. It becomes a passive spectator, separated from its natural and social context and no longer situated in culture or language.

---

[32] References to Descartes' work are abbreviated as follows: AT: *Oeuvres de Descartes*. 1904. Adam C. and Tannery T. (eds.) Paris: Vrin; CSM: *The Philosophical Writings of Descartes Volumes I and II*. 1984. Cottingham J., Stoothoff R. and Murdoch D. (eds.) Cambridge: Cambridge University Press; CSMK: *The Philosophical Writings of Descartes Volume III.* 1984. Cottingham J., Stoothoff R., Murdoch D. and Kenny A. (eds.) Cambridge: Cambridge University Press. In citations of AT, CSM, and CSMK, Roman numerals refer to volume and Arabic numerals to page.

[33] Descartes writes that 'I have convinced myself that there is absolutely nothing in the world, no sky, no earth, no minds, no bodies. Does it now follow that I too do not exist? No: if I convinced myself of something then I certainly existed. But there is a deceiver of supreme power and cunning who is deliberately and constantly deceiving me. In that case I too undoubtedly exist, if he is deceiving me; and let him deceive me as much as he can, he will never bring it about that I am nothing so long as I think that I am something. So after considering everything very thoroughly, I must finally conclude that this proposition, *I am*, *I exist*, is necessarily true whenever it is put forward by me or conceived in my mind' (Med. 2, AT VII 25).

[34] Descartes remarks that: 'When someone says "I am thinking, therefore I am, or I exist," he does not deduce existence from thought by means of a syllogism, but recognizes it as something self-evident by a simple intuition of the mind' (Replies 2, AT VII 140).

The Cartesian conception of the mind as self-founded and locked into itself has become established as the official doctrine of the modern mind. This idea, which Damasio calls 'Descartes' error', fuels the mentalistic supposition and motivates contemporary notions of the mind as a disembodied entity, hidden away and closed in behind overt behavior. It also raises the question of how we can have *access* to such a mind.

On the Cartesian view, this works as follows. When it comes to my own mind, I have a kind of so-called 'privileged access': an immediate and intuitive awareness of my inner life. Although I may start out being in a state of confusion or error, I have the ability to turn inwards and perceive the contents of my mind with utter clarity, reflecting in a methodological manner upon my stream of consciousness. Such a clear and distinct introspection, guided by the 'great light in the intellect', is illuminating and provides me with intimate knowledge of the mind's ideas. These ideas are innate and universal - they represent 'true, immutable and eternal essences' (CSMK 183, AT III 383), and Descartes writes that they have 'a seat in our mind' (CSMK 23, AT I 145). My access to the minds of others, however, is always mediated by their bodily behavior. And my perception of this behavior, like sense perception in general, is potentially *misleading*. Descartes observes that we 'misuse them [the senses] by treating them as reliable touchstones for immediate judgements about the essential nature of the bodies located outside us; yet this is an area where they provide only very obscure information' (CSM II 57-58, AT VII 83). What we perceive through our external senses results at best in a 'spontaneous impulse' to believe something.

This, of course, presses the question how we are able to access *other* minds. Since these are not 'presented' to me in the way my own mind is, they have to be 'represented'. This, however, is not really a problem for Descartes. We have privileged access to the ideas in our own mind, and it is through our knowledge of these ideas that we are in touch with the minds of our fellow human beings. Self-knowledge provides a secure basis for our knowledge of others. Importantly, we do not have to infer the existence of their minds on the basis of an analogy. Instead, Descartes short-circuits the problem of the other mind with an argument from faith. He claims that God has created man in such a way that our ideas truthfully represent what is out there in the external world, including the other's mind.

The above picture of the mind has been decisive for contemporary views on intersubjectivity, and we can find many Cartesian elements in both TT and ST approaches. One of them is the notion of a disembodied mind - hidden away and closed in behind the

overt behavior that we can see. But this notion is inevitably the result of an attitude of *disengagement* that eventually puts one at a distance from practice, where the usual clues on which we rely to orient ourselves and make sense of things are no longer available. And this attitude, in turn, has to be seen in the context of a *phenomenology of uncertainty*, in which we constantly doubt everything that occurs around us – including the intentions and behaviors of others. In such a context, it is very tempting to propose that the resulting gap between doubt and certainty has to be bridged by a *theoretical intervention.*

The phenomenology of uncertainty, our disengaged stance towards others, and the theoretical attitude by which we are to overcome our doubts are all part of the picture of intersubjectivity that is presupposed by both TT and ST. So is the idea that this attitude is universally acquired by all human beings. This supposition bears similarities to the Cartesian postulate of innate ideas that have a universal status. Gallagher (2004) calls it the supposition of universality:

Supposition 2 (the supposition of universality): 'Our reliance on theory (or our reliance on simulation or some combination of theory and simulation) is close to universal. That is, this folk-psychological way of understanding and interacting with others is pervasive in our everyday life' (p.200).

However, there is still another aspect of Cartesianism that has been very influential in shaping the intersubjectivity debate. According to Descartes, certain knowledge requires that we clearly and distinctly perceive with the mind's eye. He claims that 'doubtless, there is nothing that gives me assurance of [...] truth except the clear and distinct perception of what I affirm, which would not indeed be sufficient to give me the assurance that what I say is true, if it could ever happen that anything I thus clearly and distinctly perceived should prove false; and accordingly it seems to me that I may now take as a general rule, that all that is very clearly and distinctly apprehended (conceived) is true' (Med. 3, AT VII 35).[35] On the Cartesian view, true knowledge is modelled on a clear and distinct perception of the individual mind.

This results in what Dewey (1960, p.23) calls a 'spectator' theory of knowledge: 'the theory of knowing is modeled after what was supposed to take place in the act of vision.

[35] Somewhere else Descartes writes that 'My nature is such that so long as I perceive something very clearly and distinctly I cannot but believe it to be true' (Med. 5, AT VII 69).

The object refracts light and is seen; it makes a difference to the eye and to the person having an optical apparatus, but none to the thing seen. The real object is the object so fixed in its regal aloofness that it is a king to any beholding mind that may gaze upon it. A spectator theory of knowledge is the inevitable outcome.' According to Descartes, to know is to clearly and distinctly perceive the immediate and the intuitive (the ideas), and to suppress spontaneous impulses to believe something merely on the basis of external sense perception. 'Real' perception, in the Cartesian sense, is not a kind of action. It merely aims to reflect ideas without altering them. It is the *passive* recognition of something that is already there. To perceive with the mind's eye is to remain still and impartial – not actively engaged in the process of perceiving.

Dewey argues that a spectator theory of knowledge inevitably leads to a strict separation between perception, thinking and action. In the case of Descartes, it results in a distinction between (i) sensory perceptions (or bodily sensations), (ii) 'thinking', or a clear and distinct perception of mental ideas, and (iii) behavioral responses. In his discussion of the concept of the reflex arc in psychology, Dewey (1896) complains that these old distinctions are still firmly in place: 'instead of interpreting the character of sensation, idea and action from their place and function in the sensory-motor circuit, we still incline to interpret the latter from our preconceived and preformulated ideas of rigid distinctions between sensations, thoughts and acts. The sensory stimulus is one thing, the central activity, standing for [representing] the idea, and the motor discharge, standing for [representing] the act proper, is a third. As a result, the reflex arc is not a comprehensive, or organic unity, but a patchwork of disjointed parts, a mechanical conjunction of unallied processes' (p.358).

In contemporary discussions about intersubjectivity, the boundaries between perception, thinking and action are often still in place as well. And although there are signs that they are slowly dissolving, many proponents of TT and ST still maintain that our understanding of others somehow has to follows a Cartesian perception-thinking-action route. Of course, they have different ideas as to how the specific steps of this route should be explicated. This brings us to the next stage in the development of the problem of the other mind.

3.2 Empiricism and the argument from analogy

Although Descartes argues that clear and distinct perception is by far the best candidate for knowledge, he realizes it still falls short of absolute certainty. Descartes points out that, in order to achieve absolute certainty, he has to overcome the Evil Genius doubt and prove the existence of a non-deceiving God. 'But, that I may be able wholly to remove it, I must inquire whether there is a God, as soon as an opportunity of doing so shall present itself; and if I find that there is a God, I must examine likewise whether he can be a deceiver; for, without the knowledge of these two truths, I do not see that I can ever be certain of anything' (Med. 3, AT VII 36). Unsurprisingly, Descartes eventually comes to the conclusion that there is a God and that He is no deceiver. This enables him to evade the solipsistic consequences of his method of doubt and neutralize the problem of the other mind.

The problem of the other mind did not come to the fore until the rise of British empiricism, when the appeal to a benevolent God was no longer taken for granted and a number of other Cartesian commitments became unacceptable as well. Locke, for example, still accepted the essentials of the Cartesian picture of the mind, but rejected the claim that some truths must be innate because they are universally understood. He pointed out that the universality of a certain truth does not imply that it is therefore necessarily innate, for it could have been *learned* by all people. Moreover, the fact that infants and the mentally impaired do not understand them testifies against the plausibility of universal innate ideas. Contrary to Descartes, Locke believed that the mind of a person at birth is a tabula rasa, a blank slate upon which knowledge is imprinted through experience. He argued that ideas are derived from experience either by sensation (the affection of the senses through the observation of external bodies) or reflection (the perception of the operations of our own mind). By rejecting innateness, Locke had all the ingredients to conjure up the problem of the other mind. However, he seems not to have recognized this.

It is generally thought that, as such, the problem of the other mind was not recognized until John Stuart Mill (1878) explicitly articulated it as a prominent philosophical issue by asking: 'By what evidence do I know, or by what considerations am I led to believe, that there exist other sentient creatures; that the walking and speaking figures which I see and hear, have sensations and thoughts, or in other words, possess Minds?' (p.243). This is

not correct, however, since it was actually Thomas Reid who was the first to identify the problem of the other mind (cf. Avramides 2001).[36] It is also generally accepted, correctly this time, that Mill was the first to propose the infamous argument from analogy as a solution to this problem.[37]

Mill argued that, by observing that the bodies of other human beings behave as my body does in similar circumstances, I am able to infer that the mind I know to accompany my bodily behavior is also present in the case of others. 'Other human beings have feelings like me, because, first, they have bodies like me, which I know, in my own case, to be the antecedent condition of feelings; and because, secondly, they exhibit the acts, and other outward signs, which in my own case I know by experience to be caused by feelings. I am conscious in myself of a series of facts connected by a uniform sequence, of which the beginning is modifications of my body, the middle is feelings, the end is outward demeanor. In the case of other human beings I have the evidence of my senses for the first and last links of the series, but not for the intermediate link. I find, however, that the sequence between the first and last is as regular and constant in those other cases as it is in mine […] I must either believe them to be alive, or to be automatons: and by believing them to be alive, that is, by supposing the link to be of the same nature as in the case of which I have experience, and which is in all other respects similar, I bring other human beings, as phenomena, under the same generalizations which I know by experience to be the true theory of my own existence' (1878, p.243).

The argument from analogy crucially depends on the ability to make inferences such as 'if there is a modification of my body of kind B, then usually an experience of kind E is occurring as well', or 'if there is an experience in my mind of kind E, then usually this causes a bodily reaction of kind R.' In our own case, according to Mill, these psycho-behavioral generalizations are available because we are 'conscious' of the proper connections between (a) the modifications of my body, (b) my feelings, and (c) my outward

[36] Indeed, it seems that the first frequent use of the words 'other minds' is to be credited to him (Somerville 1989, p.249).

[37] The idea that we understand others by means of *inference* was already introduced by David Hume, who wrote that 'no passion of another discovers itself immediately to mind. We are only sensible of its causes or effects. From these we infer the passion: And consequently these give rise to our sympathy' (2003, p.410). But this answer only postpones the difficulty: by what *sort* of inference do we understand other minds? Hume is not of much help here, though it is clear that he thinks of the reasoning in terms of causes and effects: whatever inferences they are, they are based on laws or regularities which we have learned through experience hold in experience.

demeanor. In case of other minds (a) and (c) are present, but (b) is missing. However, if the connection between (a) and (c) is of the same nature as in my own case, then *by analogy* we have reason to expect them to be just as regular and constant.

Although the argument from analogy still retained the Cartesian appeal to introspection and the Cartesian primacy of self-knowledge, British empiricism rejected not only the existence of innate ideas but also abandoned the Cartesian search for absolute certainty. This, however, led to the following question: how sure can we actually be of the existence of the other mind? Consider Bertrand Russell's formulation of the argument from analogy, for example. Russell (1948) initially proposed that 'from subjective observation I know that A, which is a thought or feeling, causes B, which is a bodily act, e.g., a statement. I know also that, whenever B is an act of my own body, A is its cause. I now observe an act of the kind B in a body not my own, and I am having no thought or feeling of the kind A. But I still believe, on the basis of self-observation, that only A can cause B; I therefore infer that there was an A which caused B, though it was not an A that I could observe. On this ground I infer that other people's bodies are associated with minds, which resemble mine in proportion as their bodily behavior resembles my own' (p.486).

However, Russell soon realized that the argument from analogy, thus formulated, is only applicable in *idealized* circumstances. In practice, 'the exactness and certainty of the above statement must be softened', because even in our own case we cannot be sure that A is the only cause of B. It is possible that, although we experience A to be the cause of B, there are other causes of B 'outside our experience'. Therefore, Russell also offered a 'common sense' version of the argument of analogy: 'If, whenever we can observe whether A and B are present or absent, we find that every case of B has an A as a causal antecedent, then it is *probable* that most B's have A's as causal antecedents, even in cases where observation does not enable us to know whether A is present or not' (ibid., italics added). In other words, Russell weakened the conclusion of the argument from analogy because he doubted the human capacity for self-observation. But he still thought the argument itself was basically correct as a solution to the problem of the other mind.

The same is true for Theodor Lipps, who also remained loyal to the argument from analogy. According to Lipps, however, our understanding of others is not based on a conscious, inferential process that begins with the clear perception of our own mind. Instead, it depends on an unconscious process of *empathy*, in which we project ourselves

into the physical manifestations evinced by others.[38] Lipps (1993) suggested that such a process involves an element of 'inner imitation', and it is driven by our 'natural instinct' When watching an acrobat on a tightrope, for example, the perceived movements and affective expressions of the acrobat are 'instinctively' and simultaneously mirrored by kinesthetic 'strivings' and experiences of corresponding feelings in the observer.

Lipps used the notion of empathy in order to stress the affective, bodily and experiential dimension of how we understand others. Although empathic understanding is still based on an analogy, it does not necessarily require that we are always *aware* of how our own mind relates to our body, or that we are continuously busy *inferring* that the same is also true for other persons. In this respect, it clearly gives us a more parsimonious phenomenological account of everyday intersubjectivity.

However, the solutions offered by Russell and Lipps do not really show how the argument from analogy is able to provide us with *reliable* knowledge of the other mind. Although many philosophers no longer care about absolute knowledge Cartesian-style, they do find it problematic that the argument is based on an inductive generalization from only *one* case. Paul Churchland (1988), for example, has argued that this makes it the weakest possible instance of an inductive argument.[39] He thinks it is possible to overcome

[38] Lipps' ideas about empathy resulted from his translation of David Hume's 'A Treatise of Human Nature' into German, although Hume actually used the term 'sympathy' to describe what Lipps was interested in. Hume suggested that 'the minds of men are mirrors to one another, not only because they reflect each other's emotions, but also because those rays of passions, sentiments and opinions may be often reverberated, and may decay away by insensible degrees' (2003, p.259). He argued that this was made possible by the sole principle of all passions: sympathy. 'No quality of human nature is more remarkable, both in itself and in its consequences, than that propensity we have to sympathize with others, and to receive by communication their inclinations and sentiments, however different from, or even contrary to our own' (p.225). Sympathy makes it possible to 'enter deeply into the sentiments of others', and their affections are 'rendered present to us by the imagination', operating as if originally our own. 'We rejoice in their pleasures, and grieve for their sorrows, merely from the force of sympathy' (p.277).

[39] Some philosophers have tried to avoid this objection by arguing that the argument from analogy should be based on the multitude of correlations between mental states and behavior that one observes in one's own case, rather than on a generalization proceeding from just one observed case. Ayer (1956) for example, suggests that 'The objection that one is generalizing from a single instance can perhaps be countered by maintaining that it is not a matter of extending to all other persons a conclusion which has been found to hold for only one, but rather of proceeding from the fact that certain properties have been found to be conjoined in various circumstances. So the question that I put is not: Am I justified in assuming that what I have found to be true only of myself is also true of others? but: Having found that in various circumstances the possession of certain properties is united with the possession of a certain feeling, does this union continue to

this problem by adopting a different standard of theoretical *justification*. Churchland points out that the problem of the other mind was first formulated at a time when our grasp of the nature of theoretical justification was still rather 'primitive'. It was believed that a general law could be justified only by an inductive generalization from a suitable number of observed instances of the elements comprehended by this law. But this only works for *observable* things and properties, while modern science is full of laws that govern the behavior of *unobservable* things and properties. These laws require a different form of empirical justification. Churchland notices that contemporary theorists postulate unobservable entities and specific laws governing them, because occasionally this produces a theory that allows them to construct predictions and explanations of observable phenomena hitherto unexplained. More specifically, they assume certain hypotheses and conjoin with them information about observable circumstances in order to deduce statements about further observable phenomena, statements which are systematically true. This is commonly called 'hypothetico-deductive' justification.

Churchland claims that it is precisely this kind of justification that allows us to solve the problem of the other mind. The idea is that we understand others by employing a folk psychological theory - a network of general laws connecting mental states with perceptions, bodily behavior and other mental states. These laws are plausible for the same reason that the laws of any theory are plausible: their explanatory and predictive power. The existence of the other mind is a *hypothesis*, which is plausible to the extent that the other's behavior can be explained and predicted in terms of desires, beliefs, perceptions, emotions and so on. If this is the best way to understand the behavior of most humans, then one is justified in believing that they are 'other minds'.

Churchland can be seen as an early adaptor of the TT approach to folk psychology. He argued that folk psychology is successful as a theory if it allows us to 'explain and predict the behavior of human beings better than any other hypothesis currently available'

---

obtain when the circumstances are still further varied. The basis of the argument is broadened by absorbing the difference of persons into the difference of the situation in which the psycho-physical connections are supposed to hold' (p.249). However, this counterargument does not work. Despite that we now have a multitude of correlations, the simple fact remains that not all instances of behavior we observe in our own case are accompanied by mental states. So the conclusion to be drawn, were we proceeding from this multitude of correlations, could only be that many instances of behavior are associated with mental states. But this is not the conclusion we need. For such a conclusion is still compatible with the idea that some of the human bodies we encounter behave just as our own body does, without being associated with mental states and thus without having a mind (cf. Hyslop and Jackson 1972).

(1988, p.71). Importantly, this does not require the examination of our own case. It is the success of our folk psychology with respect to the behavior of people *in general* that matters. Nor does it require an element of analogy, in the sense that the other is 'like me'. In fact, the other might be quite different. But this, Churchland argues, does not affect my 'theoretical access' to their 'internal states', since one could 'simply use a different psychological theory to understand their behavior, a theory different from the one that comprehends one's own inner life and outer behavior' (ibid.).[40]

Churchland frames the problem of the other mind as an *inference to the best explanation*: an inference which is guided by a folk psychological theory, bringing us from observed behavior to a hidden mental state. Although this inference does not provide us with certain knowledge of the other mind, at least it gives me more reason to believe in its existence than to deny it. But the question is whether a folk psychological theory gives us the best explanation (cf. chapter 1.5). Churchland thinks this is not the case, and he dismisses folk psychology as an empirically and conceptually degenerating research program that needs to be terminated in favor of its superior alternative: cognitive neuroscience. Other proponents of TT usually do not go as far as Churchland, and instead adjust their standards of justification. They frame the problem of the other mind in terms of *adequacy*. Although we are certainly not infallible, it is very often the case that folk psychology allows us to successfully predict what others are going to do, or explain what they have done.

3.3 Deconstructing the argument from analogy

So far I have sketched (a part of) the historical background of the problem of the other mind and the argument from analogy. We saw that the problem of the other mind encompasses more than just a notion of the mind as a disembodied and disembedded entity. At a far more profound level, it is inspired by a Cartesian anxiety, and a longing for certainty that has to be met by methodological thinking.

[40] Notice that Churchland's solution is different from the one offered by other proponents of TT, who argue that we employ the same folk psychological theory in case of self and other knowledge (cf. chapter 2).

This anxiety is still present in contemporary TT and ST explanations of intersubjectivity. It suggests that our encounters with our fellow human beings are essentially *problematic*, since we are always in the dark about their intentions, feelings and beliefs. In order to overcome our doubt in these situations, we need to take a step back and disengage from active participation. We need to adopt a theoretical, third-person stance towards others in order to figure out what they are up to, ascribing causally efficacious inner mental states to them for the purpose of prediction, explanation and control.[41] As a result, we are not actively involved but rather stand as passive observers at the margins of the situation. We do not have the slightest clue about what is going on, or how we need to *respond* to what happens, unless we call forth a theory or run a simulation routine. Proponents of both ST and TT think that we need some kind of *intervention* between our initial observation of others and our final reaction towards them.

This separation between perception and action can be seen as a consequence of the Cartesian spectator theory of knowledge, and it leads to a 'sandwich model' of intersubjectivity. Hurley (2008) argues that such a model 'regards perception as input from the world to the mind, action as output from the mind to the world, and cognition as sandwiched in between. Central cognition, on this view, is where all the conceptually structured general purpose thinking happens: perceptual information is assessed in light of standing beliefs and goals, deliberative and inferential processing occurs, action plans are formulated and sent on for execution' (p.2). According to ST, this cognitive intervention proceeds according to some version of the argument from analogy: since I know my own mind and how it relates to my body, I am able to infer that this is also true for the other on the basis of an *analogy* between our bodies. TT, by contrast, rejects the analogical element but sticks to the idea of an intervention based on *theoretical inference*. It claims that we understand others by inferring the contents of their minds on the basis of a folk psychological theory. In what follows, I will deconstruct the argument from analogy into three components: introspection, inference and mental concept mastery, and argue that these components are problematic insofar they come with serious developmental constraints and are modeled on the minds of *individual* agents.

[41] This is what Bogdan (1997) labels 'the spectatorial view of interpretation', since it portrays 'the subject as a remote object of observation and prediction' (p.104).

Introspection

In the previous chapter we already encountered a number of initial objections to the argument from analogy by philosophers such as Ryle and Scheler (cf. chapter 2.3). It pays to follow Scheler a bit further here, since he not only provides us with a whole list of direct criticisms of the argument from analogy, but also attempts to dismantle two crucial presuppositions behind it.

First, the argument from analogy assumes that we perceive only the bodies of others and therefore have to *infer* the existence of their minds. As a result, we are unable to experience the thoughts, feelings and emotions of others in a direct way. According to Scheler, however, this assumption is not supported by the phenomenological evidence. On the contrary, it is a 'phenomenological fact' that we perceive other minds, much like we perceive our own mind. Rather than being busy with inferring their mental states, we are able to directly perceive them. Scheler (1973) famously claims that 'we certainly believe ourselves to be directly acquainted with another person's joy in his laughter, with his sorrow and pain in his tears, with his shame in his blushing, with his entreaty in his outstretched hands, with his love in his look of affection, with his rage in the gnashing of his teeth, with his threats in the clenching of his fist, and with the tenor of his thoughts in the sound of his words. If anyone tells me that this is not 'perception' [...] I would beg him to turn aside from such questionable theories and address himself to the phenomenological facts' (p.254). This argument is directed against the traditional idea that perception and action require the intervention of cognition.

Second, the argument from analogy is grounded in the assumption that self-knowledge is 'given' to us in our first-person experience and can be used as a foundation for our knowledge of others. This is doubtful as well, according to Scheler, for 'who can say that it is our own individual self and its experiences which are "immediately given" in that mode of intuition, by which alone the mental, a self and its experiences, can possibly be apprehended, namely in inner intuition or perception? Where is the phenomenological evidence for this assertion?' (p.244). Scheler suggests that the argument from analogy 'underestimates the difficulties involved in self-experience and overestimates the difficulties involved in the experience of others' (ibid.).[42]

[42] Scheler's objection is similar to Sellars criticism of the myth of the given (cf. chapter 1.2). However, where Sellars and his TT followers maintained that self and other knowledge are equally

The idea that we need to introspect an inner mental realm before we can engage in social interaction is problematic when we consider our everyday phenomenology, as I remarked in the previous chapter. But according to Scheler, there is another problem as well. This has to do with the *unreliability* of introspection. The fact that for a long time this has been overlooked is partly due to the strong influence of the Cartesian ideal of introspection as a clear and distinct perception. The founders of psychology, Wilhelm Wundt and William James, were still convinced that introspection was of crucial importance for our knowledge of the mind. James, for example, said that 'the word introspection need hardly be defined – it means, of course, the looking into our own minds and reporting what we there discover. Everyone agrees that we there discover states of consciousness' (James 1890/1981, p.85). Back then, it was still thought that introspection, as a method, distinguished psychology from the natural sciences. Hempel (1949) describes the received view at the time as follows: 'It is impossible to deal adequately with the subject matter of psychology by means of physical methods. The subject matter of physics includes such concepts as mass, wave length, temperature, field intensity, etc. In dealing with these, physics employs its distinctive method which makes a combined use of description and causal explanation. Psychology, on the other hand, has for its subject matter notions which are, in a broad sense, mental. They are *toto genere* different from the concepts of physics, and the appropriate method for dealing with them scientifically is that of empathetic insight, called 'introspection', a method which is peculiar to psychology' (p.375).

However, with the rise of behaviorism, psychologists became increasingly doubtful about the prospect of introspection as a viable psychological method. Watson (1913), for example, published a statement of behaviorist principles that began as follows: 'psychology as the behaviorists view it is a purely objective experimental branch of natural science. Its theoretical goal is the prediction and control of behavior. Introspection forms no essential part of its methods, nor is the scientific value of its data dependent upon the readiness with which they lend themselves to interpretation in terms of consciousness' (p.158).

A more recent and very influential critique of introspection can be found in an article by Nisbett and Wilson (1977), who concluded that people have little or no introspective access to higher order cognitive processes. The authors reported evidence of subjects

---

problematic and in need of *theory,* Scheler argues that the *practice* of self and other experience is well established.

confabulating stories about the cause of the mental states they were entertaining. At a shopping mall, they mounted a display table with four pairs of identical pantyhose, labeled A, B, C and D from left to right, and asked passersby which pair they preferred and what reasons they had for doing so. In a previous version of the study, they had ascertained that there was a strong position effect: pair A was preferred by 12 percent of the participants, pair B by 17 percent, pair C by 31 percent and pair D by 40 percent. In the main study, when people where asked the reason for their choice, people pointed to some attribute of the preferred pair, such as its superior knit, sheerness, or elasticity. Nobody spontaneously mentioned the position effect as the cause of his preference – even when specifically asked whether their choice had been influenced by position (with the exception of a participant who was taking psychology courses). The authors concluded that participants seemed totally unaware of what was in fact the cause of their preference, and their claim about what caused it was merely a *confabulation*.

This is only the tip of the iceberg, and there have been many more studies on confabulation since. Gazzaniga (1992) and Bayes and Gazzaniga (2000), for example, have provided evidence for confabulation in split-brain patients, who had undergone surgical separation of their two hemispheres.[43] And Wegner (2002) has argued that confabulation, or what he calls 'intention invention', is also pervasive when it comes to our everyday self-ascription of consciously willed decisions. Whenever we explain our acts as the outcome of our conscious choice, we engage in intention invention, because our actions actually stem from countless causes of which we are completely unaware.[44] Wegner claims that 'When we apply mental explanations to our own behavior-causation mechanisms, we fall prey to the impression that our conscious will causes our actions. The fact is, we find it enormously seductive to think of ourselves as having minds, and so we are drawn into an intuitive appreciation of our own conscious will [...] The real causal sequence underlying human behavior involves a massively complicated set of

[43] These studies began as investigations of the abilities of people who have had their left and right brains surgically severed as a treatment for severe seizures. Such a treatment leaves mid and lower brain structures joining the two sides intact, but it creates a 'split brain' at the cortex.

[44] For example, consider the following study by Brasil-Neto et al. (1992). The experimenters exposed the participants to TMS (transcranial magnetic stimulation) of the motor area of the brain as the participants chose freely whether to move their right or left index finger. Surprisingly, although the participants showed a marked preference to move the finger contra-lateral to the site stimulated, they continued to perceive that they were voluntarily choosing which finger to move.

mechanisms' (2002, p.26f). This implies, according to Wegner, that an agent cannot be the real cause of his or her action. The agent self is only a virtual entity, an 'apparent mental causer' (2005, p.23).[45]

I certainly do not wish to defend Wegner's explanation of our everyday explanation of our own behavior, but I do think that the above experiments at the very least indicate that the commonsense use of introspection is far removed from the Cartesian ideal of clear and distinct perception. At the same time, however, the studies mentioned above do not seem to prove that we have no privileged access *whatsoever*. Although it might be true that we are not aware of the causes of our behavior, it could still be argued we do have a kind of privileged access to a great deal of information about ourselves, such as the content of our current thoughts and feelings, and the objects of our attention. Wilson (2002) has recently recanted part of his earlier confabulation story by admitting that 'the fact that people make errors about the causes of their responses does not mean that their inner worlds are a black box. I can bring to mind a great deal of information that is inaccessible to anyone but me. Unless you can read my mind, there is no way you could know that a specific memory just came to mind, namely an incident in high school in which I dropped my bag lunch out a third-floor window, narrowly missing a gym teacher who happened to walk around a corner at just the wrong time. Isn't this a case of having privileged "introspective access to higher order cognitive processes?" [...] Although we often have access to the results of these processes- such as my memory of the lunch-dropping accident- we do not have access to the mental processes that produced them. I don't really know, for example, why that particular memory came to mind, just as the participants in the panty-hose study did not know exactly why they preferred pair D over A' (p.150). And in a further passage, he claims that: 'To the extent that people's responses are caused by the adaptive unconscious, they do not have privileged access to the causes and must infer them, just as Nisbett and I argued. But to the extent that people's responses are caused by the conscious self, they have privileged access to the actual causes of these responses; in short, the Nisbett and Wilson argument was wrong about such cases' (p.106).

I think Wilson is correct in claiming that we do have some kind of privileged access to the contents of our own mind. But the last passage above is confusing, because it

[45] See also Wegner (2003), where he states that 'The theory of apparent mental causation turns the everyday notion of intention on its head [...] The theory says that people perceive that they are intending and that they understand behavior as intended or unintended - but they do not really intend' (p.10).

suggests that some responses are caused by the adaptive unconscious while others are caused by the conscious self. Later in this book, I will offer an alternative story about self-knowledge as an active and constructive process of *interpretation* (that also involves a certain amount of confabulation) instead of a passive introspection of one's own mental states (cf. chapter 5.2).

It is often claimed that the unreliability of introspective (phenomenological) properties poses a potential problem for those simulation theorists who rely on introspection to get their simulation routines off the ground. But a far more serious developmental constraint on the appeal to introspection is the fact that it presupposes *mental concept mastery*. If the introspection of our own mental states is the starting point for our intersubjective engagements, then this already presupposes that we are able to *identify* and *self-attribute* them. And if we are to distinguish between and clearly recognize the many varieties of mental states, thereafter to divine the connections they bear to our behavior, we must possess the concepts necessary for making such identifying judgments. We must grasp the meaning of the terms 'belief', 'desire', 'pain' and so forth. As we saw in the previous chapter, Goldman (2006) tries to avoid this requirement by putting forward *neural* states as suitable candidates for introspection. But it does not seem to make much sense to claim that we are able to introspect neural states in a conscious manner. Nor does it make sense to talk about the *unconscious* introspection of neural states, unless this process is construed as a kind of feedback or forward comparator (cf. chapter 4.3). In this case, however, it is not clear why the label 'introspection' should be used. In other words, if we insist on appealing to introspection, it seems we need a story about the acquisition of mental concepts in ontogeny.

Inference

Such a story about mental concept acquisition is also required if we wish to properly explain how human agents are able to make *inferences.* The latter ability is crucial for TT explanations of everyday social interactions, according to which we make sense of each other's actions by means of a folk psychological theory that specifies how beliefs and desires combine to give rise to intentions and actions. TT argues that this theoretical 'system of inferences' is the engine of everyday interpersonal understanding - even though

it must be supported by further auxiliary generalizations about what people typically do in a range of circumstances. At the core of the theory is the belief-desire principle: 'if A wants p and believes that doing q will bring about p, then ceteris paribus, A will q' (Borg 2007, p.6).[46] If we support this principle with other folk psychological generalizations such as 'persons who *want* to quench their thirst and *believe* that drinking water will satisfy their thirst, will tend to drink water', we can construct inferential arguments and use their conclusions for the purposes of behavior *prediction*:

1. Persons who want to quench their thirst, and believe that drinking water will satisfy their thirst, will tend to drink water (folk psychological law)
2. This person feels thirsty (first premise)
3. This person believes that drinking water will satisfy his thirst (second premise)
4. Normal conditions obtain (ceteris paribus)

---

5. This person is going to drink water (conclusion)

But we can also use these folk psychological generalizations to *explain* behavior. The question 'Why is he drinking water?' can be answered by referring to a belief-desire pair: 'Because he *wants* to satisfy his thirst, and he *believes* that drinking water will satisfy his thirst'. In both cases, we infer the conclusion from a folk psychological law, in combination with the starting premises (the initial conditions needed to connect this law to the specific explanation or prediction) and the ceteris paribus clause. It is often suggested that additional principles are needed in order to guarantee that we make these inferences in a *reliable* way. Botterill (1996), for example, gives us the following principle: '[Inference Principle] When an agent A acquires the belief that p and a rational thinker ought to infer q from the conjunction of p with other beliefs that A has, A comes to believe that q' (p.116).

As I already remarked in chapter 1, there are a number of problems with this theory-driven picture of mindreading. An important question is how we acquire the background knowledge needed to sensitively apply our folk psychological theory in the large variety of practical contexts, without having to claim that all this knowledge is simply innate. Another

[46] See also Botterill (1996), who claims that 'if belief-desire psychology has a central principle, it must link belief, desire and behavior. It could be formulated like this: [Action Principle] An agent will act in such a way as to satisfy, or at least to increase the likelihood of satisfaction, his/her current strongest desire in light of his/her beliefs' (p.115).

pressing question is how we acquire the theory itself. However, there are also problems with the idea that we understand others by means of an *inferential procedure*. Wittgenstein (1953), for example, argues that 'I know that a person who behaves in a particular way - who, for example, gets red in the face, shouts, gesticulates, speaks vehemently, and so forth - is angry precisely because I have learned the concept "anger" by reference to such behavioral criteria. There is no inference involved here. I do not reason "he behaves in this way, therefore he is angry" - rather "behaving in this way" is part of what it is to be angry and it does not occur to any sane person to question whether the individual who acts in this way is conscious or has a mental life' (§303). Wittgenstein's point is that our knowledge of the other mind is not primarily inferential in nature, but rather determined by public criteria that govern the application of psychological concepts. Inference seems only required under the Cartesian assumption that we have to work 'outwards' from the interiority of our own mind, to abstract from our own cases to the 'internal' world of others. This argument fits nicely with Gallagher's (2004) observation that there is no phenomenological evidence for the claim that we use inferential principles when we are interacting with other persons. (cf. chapter 1.3)

Proponents of ST might try to avoid these problems by proposing that mindreading is *process-driven*. We are capable of accurately simulating another person as long as (i) the process driving the simulation of the other is the same as the process that drives our own system, and (ii) our initial mental states are the same as those of the other person.[47] These requirements are representative for the analogical element that is characteristic for the argument of analogy - that the other is 'like me' in the relevant aspects. Since the simulator and its target are probably not exactly psychologically alike, we need to feed pretend inputs into the relevant psychological mechanisms in order to come up with decent predictions and explanations. This allows us to make 'adjustments for the relevant differences'.

The assumption of analogy allows us to understand others without theory. As Goldman (2006) puts it, 'to read the mind of others, they need not consult a special chapter on human psychology, containing a theory about the human decision-making mechanism. Because they have one of those mechanisms themselves, they can simply run their mechanism on the pretend input appropriate to the target's initial position. When

[47] However, as Fuller (1995) points out, this also implies that simulation routines still depend on 'a general premise stating that the model is relevantly similar to the [thing modeled]' (p.22).

the mechanism spits out a decisional output, they can use the output to predict the target's decision. In other words, mindreaders use their own minds to 'mirror' or 'mimic' the minds of others' (p.20). However, we noticed in the previous chapter that this makes simulation very vulnerable to a collapse into tacit theory. Mirroring processes still seem to require that 'some elements inside the attributor causally mediate between his explicit premises and conclusions, and that the causal structure of these elements mirrors the logical structure of psychological theory' (p.33). And this means that we cannot employ simulation routines without the help of some kind of inferential principle that enables us to reliably infer the logical conclusion from the general premise that the other is 'like me' and the other 'pretend' premises.

Another problem with the appeal to inference is that it comes with a severe *developmental constraint*. In order to infer the mental states of others, be it by means of a folk psychological theory or on the basis of an analogical premise, I already need to have some (mastery of) mental concepts. As Hutto (2004) points out, the inferential procedures employed by TT and ST make use of rather sophisticated abstract concepts such as: 'agent', 'rational thinker', 'belief' and 'desire'. It remains doubtful whether, let's say, four-year-olds, already have a handle on these concepts. One might try to sidestep this requirement by arguing that we should think of the relevant inferential processes as taking place at the *sub-personal* level, that is, in the brain. Goldman (2006), as we saw in the previous chapter, argues that mindreading is executed at the personal level by simulation, which is in turn implemented at the sub-personal level by a set of inferential principles. Simulation routines are executed by an 'algorithm' that is 'tacitly represented at some level in the brain' (p.33) The problem is, however, that such an algorithm still needs to operate upon *mental content* if we want to maintain that it functions like an *inferential* argument. And this in turn requires a sensible notion of content.

This is primarily a concern for simulation theorists who employ a *broad* notion of simulation as being essentially a mindreading process. Currie and Ravenscroft (2003) note that 'simulation, as it is currently used, is ambiguous; it has a narrower and a broader meaning. Suppose I try to predict your behavior by imagining myself in your situation. There are three things that must go on if I am to get the answer by simulation. The first is to acquire knowledge, or at least some beliefs, about your situation. The second thing is for me to place myself, in imagination, in that situation and to see, what, in imagination, I decide. The third is to draw a conclusion from this about what you will do. Sometimes

"simulation" refers to the whole three-tier process, sometimes just to the bit in the middle' (p.54). Simulation theorists who adopt a broad understanding of simulation construe it as involving an inferential procedure that follows the steps of a logical argument, and are therefore not really different from theory theorist.[48]

In this respect, it is far less demanding to articulate a narrower notion of simulation. Gordon (1995), for example, argues that simulation is not a process of *transportation* but rather one of *transformation*. This is a 'hot' methodology because it involves the exploitation of one's own motivational and emotional resources.[49] Crucially, such an imaginative transformation does not require, as he puts it, any 'inference from me to you'. In proposing his radical kind of simulation, Gordon rejects the assumption that our social encounters mainly take place against the backdrop of strong first/third-person divide. He points out that the mirror neuron processes constitutive for what Gallese calls the 'shared manifold' or 'we-space' implicitly express the *similarity* of self and other rather than their *distinctness*. They show us how the other's observed behavior and the self's matching response become intelligible *together*, that is, in the same process. When we engage in social interaction, it is not necessary for to us make any assumptions about our similarity to them, implicit or otherwise. Gordon (2005) suggests that we do not infer from the first to the third-person, but rather 'multiply the first person'.

The question is to which extent mirror neuron processes can still be interpreted as instances of simulation (cf. chapter 2.3). This is a problem for all simulation theorists insofar they hold that simulation operates 'primarily at the sub-verbal level' (Gordon 1986, p.170) and claim that, for a large part, simulation is 'non-conscious or minimally conscious'

[48] Of course, this does not mean that both positions are vulnerable to the same objections. For example, an objection against the formal validity of the argument from analogy is that it only enables me to understand myself in the situation - I don't understand the other. Wittgenstein (1953), for example, observes that 'If one has to imagine someone else's pain on the model of one's own, this is none too easy a thing to do: for I have to imagine pain which I do not feel on the model of the pain which I do feel. That is, what I have to do is not simply to make a transition in imagination from one place of pain to another. As, from pain in the hand to pain in the arm. For I am not to imagine that I feel pain in some region of his body (Which would also be possible)' (§302). Although this argument has some force against ST, it cannot be used against TT. This is because TT is committed to inference but not to analogy.

[49] See Gordon (1992), where he writes that 'In seeking an explanation of your friend's action, you were looking for features of the environment (features you believed it to possess) that were menacing, frightening, attractive, and the like. This is not a matter of looking dispassionately for features believed to produce certain characteristic actions or emotions. Rather, it is a search that essentially engages your own practical and emotional responses' (p.15).

(Goldman 2006, p.151). But those who employ a narrow notion of simulation at least have the benefit of *parsimony*, in the sense that they do not have to postulate tacit inferential principles and (non-)conceptual mental contents in order to explain our basic intersubjective engagements. Hutto (2004) argues that the tendency of simulation theorists (i.e., those who employ a broad notion of simulation) to do this in fact reveals a *theoretical bias* in their view of intersubjectivity, and he warns against the assumption that '[...] the processes involved in basic acts of recognition, even intersubjective ones, tacitly mimic those of mature reasoners who would tackle the same problem using a set of abstract concepts and general principles so as to make explicit inferences. We are systematically misled on this score because in the very act of classifying such behavior we must employ our own conceptual scheme of reference. But it is nothing more than an intellectual bias to suppose that, for example, young children or animals must be tacitly employing it' (p.557). According to Hutto, this intellectual bias is particularly hard to resist as long as it is assumed that we always start from a detached point of view in our dealings with others. The question is precisely whether such a viewpoint does justice to our everyday social engagements with others. Most of the time, we already know what to expect from others and they know what to expect from us. We do not need any mediating knowledge or inferential principles because 'much of the work of understanding one another in day-to-day interactions is not really done by us at all, explicitly or implicitly. The work is done and carried by the world, embedded in the norms and routines that structure such interactions' (McGeer 2001, p.119).

Mental concept mastery

Both the ability to introspect my own mental states and the ability to make appropriate inferences over them presuppose a certain level of mental concept mastery. To introspect a specific sensation or to infer that someone is in pain, I need to know what the mental concepts 'pain' and 'sensation' mean. But how do we acquire this knowledge? Some proponents of ST suggest that these terms get their meaning by 'inner ostension' - by being directly associated with a specific quality of internal and privately experienced mental states. This is the view of Meltzoff (2002), for example, who is often interpreted as a theory theorist, but in this respect defends a simulation approach. Meltzoff proposes that our

understanding of mental states develops as follows: ‘As infants perform particular bodily acts they have certain mental experiences. Behaviors are regularly related to mental states. For example, when infants produce certain emotional expressions and bodily activities, such as smiling or struggling to obtain a toy, they also experience their own mental states. Infants register this systematic relation between their own behavior and underlying mental states’ (p.35). In a further step, infants use these ‘behavior-mental states mappings’ to make inferences about the mental states of others on the basis of an analogy.

The most fundamental flaw in proposals like these is precisely the assumption of inner ostension, i.e. that one learns from one’s own case what thinking, feeling, sensation are. Wittgenstein (1953) already showed how this leads first to solipsism, and then to nonsense. He illustrated the difficulty of inner ostension by scrutinizing the following quote from Augustine: ‘When they (my elders) named some object, and accordingly moved towards something, I saw this and I grasped that the thing was called by the sound they uttered when they meant to point it out. Their intention was shown by their bodily movements, as it were the natural language of all peoples: the expression of the face, the play of the eyes, the movement of other parts of the body, and the tone of the voice which expresses our state of mind in seeking, having, rejecting, or avoiding something. Thus, as I heard words repeatedly used in their proper places in various sentences, I gradually learnt to understand what objects they signified; and after I trained my mouth to form these signs, I used them to express my own desires’ (Confessions I 8).

The above passage indicates that Augustine assumes that language learning occurs through ostensive definition, i.e., that the meaning of a term is learned by pointing out examples. But Wittgenstein argues that this assumption is very problematic. One problem of learning by ostensive definition is that this by itself can never fix the meaning of a word. ‘No one can ostensively define a proper name, the name of a color, the name of a material, a numeral, the name of a point of the compass and so on. The definition of the number two, "That is called 'two'" - pointing to two nuts - is perfectly exact. But how can two be defined like that? The person one gives the definition to doesn't know what one wants to call "two"; he will suppose that "two" is the name given to *this* group of nuts! He *may* suppose this; but perhaps he does not. He might make the opposite mistake; when I want to assign a name to this group of nuts, he might understand it as a numeral. And he might equally well take the name of a person, of which I give an ostensive definition, as

that of a color, of a race, or even of a point of the compass. That is to say: an ostensive definition can be variously interpreted in *every* case' (1953, §28).

In order to learn from ostensive definition, the learner already needs to have some grasp of what the teacher intends when pointing to something. An everyday ostensive definition is embedded in a public language, and in a social community in which that language is used. '[...] the ostensive definition explains the use - the meaning - of the word when the overall role of the word in language is clear. Thus if I know that someone means to explain a color-word to me the ostensive definition "That is called sepia" will help me to understand the word [...] One has already to know (or be able to do) something in order to be capable of asking a thing's name' (§30).

For Augustine, by contrast, language 'expresses our state of mind', and he seems to assume that language learning is essentially of matter of *understanding*. According to Wittgenstein, Augustine's account of how we learn our first language actually resembles how we learn a *second* (foreign) language (cf. §32). Learning by ostensive definition seems to imply the translation of an inner private language into an outer conventional language. The terms of this inner private language get their meaning through inner ostension, by being directly associated with a specific quality of privately experienced mental states - independently of a public language.

However, Wittgenstein gives us a powerful argument for the *impossibility* of such a private language. This is how it goes: suppose that at a certain point in time, you decide to endow the term W with meaning, solely by associating it with a certain sensation you feel at that time. At a later time, upon feeling the same sensation, you say: 'Hey, there is another W.' But how can you determine whether you have used the term correctly on this occasion? Perhaps you misremembered the fist sensation. Or perhaps you saw a close similarity where in fact there was none. In order to distinguish between the correct use of the term W and the incorrect use of W, one must have a *criterion for identification*. This entails that one must be able to follow a rule privately in isolation from others. But this is impossible, according to Wittgenstein, because *seeming* to follow a rule can never be tantamount to actually following that rule. Whatever *seems* right will *be* right, which only means that here we can't talk about right (cf. §258).

The private language argument is obviously problematic for ST insofar as the latter assumes that we first learn to identify and self-attribute mental states *in private,* and then use this as a starting point for our knowledge of the other mind. However, it is often

objected that the argument draws a stronger conclusion than its premises justify. For if a public check on a correct application is all what is required for meaningfulness, then all one's understanding of 'W' need include is some connections between the occurrence of this sensation and the occurrence of other phenomena. But these other phenomena need not be publicly observable phenomena per se; they can be other mental states and still serve as checks on the correct application of 'W'. This idea is at the core of TT, according to which the meaning of folk psychological concepts such as beliefs and desire depends on their role in a larger theoretical framework. Meaning is not just given, but created as a function of prediction and explanation. And this is not necessarily a public process.

Consider Fodor's (1979) 'Language of Thought', for example. According to this MTT proposal, humans are born with a content-processing system built into their nervous system, which resembles the machine language that is hard-wired into a computer. But instead of computing binary code, this system processes mental representations by means of specific rules. These representations have contents by virtue of their ability to correspond with (things in) the world. At the same time, they are sensitive to computational processing due to their 'lingual' nature. Fodor (1979) explicitly defends the Augustinian idea that learning a first language is a translation process from an inner language to an outer language. He argues that the language of thought is what eventually enables us to learn our 'natural' language: 'You cannot learn a language whose terms express semantic properties not expressed by the terms of some language you are already able to use' (p.61). One possible objection against the language of thought is that it leads to a regress: if we cannot learn a language unless we already have one, we also need another language in order to learn the first one, and so on ad infinitum. But Fodor dismisses this objection by saying that in order to use a language, you don't need to *learn* a language – you need to *know* it. And the language of thought is known but not learned, since each of us is simply *born* with it.[50]

What about the private language argument? Fodor's language of thought is private in the sense that it is not being governed by public conventions. However, Fodor argues, this is not necessary a problem as long as it is employed in a tacit way. According to the private language argument, we are not able to follow a rule privately in isolation from others. But what if all human beings are born with the same set of folk psychological rules?

[50] Notice that such an appeal to innateness is typical for the rationalist approach to the problem of knowledge (cf. chapter 1.2).

What if we are natural born rule-followers? Again, the appeal to tacit theory is tempting. However, we already saw in chapter 1 that the assumption of a tacit set of (innately acquired) theoretical principles still does not solve the problem of how we are able to sensitively apply our mindreading skills in a large range of practical contexts. It seems that we need more than simple belief-desire syllogisms in order to select the specific contents of the mental states over which our folk psychological theory quantifies in particular situations. How do we acquire the background knowledge needed to pull this off? Fodor has only one answer to this question: innateness. But if everything is already in place before we acquire our 'natural' linguistic skills, another question pops up. How can we make sense of the mental content that is needed to fuel our tacit theory of mind? It is notoriously difficult to spell out what is precisely meant by a tacit belief or desire. This is precisely why some have suggested that, when it comes to specifying the content of the belief and desires of nonverbals, folk psychology increasingly comes 'under stress' (cf. Godfrey-Smith 2003). The root problem, according to Hutto (2007a), is that the very idea of content as something 'given' in perceptual encounters, 'acquired' by mental states, and 'manipulated' in sub-personal cognitive processing is deeply problematic if not incoherent. Remark that the notion of mental content is not only problematic for modular TT (MTT), but also for scientific TT (STT). Although STT rejects the claim that mental content is innate, it still needs to explain how our folk psychological principles facilitate the acquisition of non-conceptual mental content during development.[51]

A question of analogy

So far I have not dealt with something that is crucial to the argument of analogy. This is the idea of analogy *itself*. My evaluation of this requirement entirely depends on how it is

[51] Many proponents of TT think that the basic perceptual acts of nonverbal creatures are content-involving. Evans (1982) gives us an adequate description of what this means: 'In general, we may regard a perceptual experience as an informational state of the subject: it has a certain *content* -- the world is represented a certain way - and hence it permits of a non-derivative classification as *true* or *false.* For an internal state to be so regarded, it must have appropriate connections with behavior - it must have a certain motive force upon the actions of the subject [...] The informational states which a subject acquires through perception are *non-conceptual,* or *nonconceptualised.* Judgements *based upon* such states necessarily involve conceptualisation' (pp.226-7). The idea is that perception involves a translation (or conceptualization) of the contents of one language into another.

interpreted. As long as we conceive of analogy in terms of a (tacit) premise, which states that the interpreter is similar to the agent under consideration and serves as a starting point for a (conscious) inferential procedure, I think the requirement of analogy is very problematic. Such an interpretation is usually endorsed by simulation theorists that employ a broad notion of simulation. It not only introduces a number of developmental constraints (most importantly, that of mental concept mastery), but is also vulnerable to a number of standard objections against the *argument* from analogy. (cf. chapter 2.2) But what if it is possible to have the argument from analogy without the actual argument? What if there is an analogy between ourselves and others that is non-conceptual and non-inferential in nature?

Gallagher (2003a) points out that such an analogy, a kind of 'common code', may be found at the level of sensory-motor mechanisms. He claims that developmental studies suggest that this common code is already operative from the very beginning of life: 'What I see is automatically registered in a code that is common to other sense modalities, including proprioception; and in the case of seeing biological movement, perception includes motoric, kinaesthetic activation. So when I see the other's body moving in a certain way, I have a kinesthetic-proprioceptive sense of what that is like.' Analogy, thus understood, seems to be far less demanding. And it can even be used against those who are critical of the argument from analogy. For example, Zahavi (2001) argues that for the argument from analogy to work, there has to be a similarity between the way in which my own body is given to me, and the way in which the body of the other is given to me. But Zahavi points out that my own body, as it is *felt* proprioceptively for me, does not at all resemble the other's body as it is *perceived* visually by me. However, if we can find a non-conceptual and non-inferential analogy between ourselves and others at the sub-personal level, this argument appears to be off-base in an important way. It is tempting to argue that such an analogy would in fact prove that we are born with some kind of inner language of thought. This would be a mistake, however, since it would be a very strange language – a language of which we are not conscious, and which does not involve inference, concepts or content. In other words, this sub-personal language would lack all the important properties we normally attribute to language.

3.4 Beyond the problem of the other mind

Given that our introspective and inferential abilities presuppose a rather sophisticated knowledge of mental concepts, the bottom line question is how we acquire these concepts and come to learn what they mean. Many proponents of TT and ST think that this is the achievement of the *individual agent.* Meaning is primarily a private affair – it is 'given' through introspection (Goldman's ST), explained in terms of innateness (modular TT) or picked up from the environment through perception (scientific TT).[52] With such a narrow Cartesian focus on *subjectivity*, it seems almost natural to assume that intersubjectivity is *derivative* – a matching process between individualized mental states that share the same meaning. It also seems inevitable that, on such a view, intersubjectivity turns out to be very problematic. As long as we are inspired by a Cartesian phenomenology of uncertainty, the pressing issue remains how we can access other minds, i.e. what sort of intervention process (inference, introspection, or analogy) is needed between our initial perception of others and our active response towards them.

TT and ST assume that in order to solve this problem, we have to start with the primacy of a theoretical, third-person stance towards others. But a much more basic question is whether the Cartesian context is the primary context in which intersubjectivity takes place. Gallagher (2001) is right to stress that it is questionable whether our ordinary attempts to understand other people are best characterized as explanations and predictions. Most of our intersubjective encounters are firmly rooted in *second-person interactions*, in which we directly engage with others and already know to some extent what we can expect from them. Of course, there are situations in which we can be perplexed by their actions, and try to predict their next move or explain what exactly motivated them to behave in a certain way. As Hutto (2007a) points out, 'driven by suspicion we may be left with nothing but speculation and supposition about their motives. That is, we may be forced to make third-party predictions and explanations of actions precisely in the sorts of cases in which we do not know what to expect from others or when we cannot engage with them directly. But, for this very reason, these sorts of approaches

[52] Externalist TT (cf. chapter 1.2) is clearly the exception here, and I fully agree with their objection to internalist versions of TT that folk psychological principles 'ain't in the head'. At the same time, however, externalist TT still takes the prediction/explanation of behavior by means of a folk psychological theory to be central to intersubjectivity, and in this respect remains firmly rooted in the Cartesian tradition discussed in the previous sections.

are bound to be, on the whole, much less reliable than our second-person modes of interaction' (p.13).

The moral is that, in practice, there is no *general* problem of the other mind. Why should we assume that intersubjectivity is intrinsically problematic and best characterized in terms of a phenomenology of uncertainty? If we pay attention to practice, we find that most of the time we already have some basic understanding of what to expect from others, and we also know what they expect from us. We do not need to engage in inferential or introspective procedures to make sense of what they mean or what they are doing. Gallagher (2001) claims that 'before we are in a position to theorize, simulate, explain or predict mental states in others, we are already in a position to interact with and to understand others in terms of their gestures, intentions and emotions, and in terms of what they see, what they do or pretend' (p.91).

Importantly, these interactions provide us with a basic understanding of other minds that is not subject to *reasonable* doubt. This is well expressed by Thomas Reid (1983), who was much more practically minded than John Stuart Mill in this respect. Reid dismissed the problem of the other mind by arguing that 'No Man thinks of asking himself what reason he has to believe that his neighbor is a living creature. He would be not a little surprised if another person should ask him so absurd a question: and perhaps could not give any reason which would not equally prove a watch or a puppet to be a living creature. But, though you should satisfy him of the weakness of the reasons he gives for his belief, you cannot make him in the least doubtful. This belief stands upon another foundation than that of reasoning and therefore, whether a man can give good reasons for it or not, it is not in his power to shake it off' (pp.278-9). Reid is right that the problem of the other mind does not show up in our common-sense encounters with others. But of course, we might wonder what to make of this foundation that secures our understanding of the other mind. How do we interact with others 'in terms of their gestures, intentions and emotions'?

In the next chapter, I attempt to answer the above questions by making a case for the importance of *second-person practices*. These intersubjective engagements embody our baseline understanding of others, and enable us to relate to them in a direct way - without mindreading or other cognitive/conceptual interventions. Such a pragmatic approach is able to avoid the severe developmental constraints that need to be met by TT and (most) ST accounts, and gives us more insight in the context-sensitivity of our intersubjective capacities. Moreover, it also does more justice to the empirical evidence on their *actual*

development. From a pragmatic point of view, this is one of the first requirements that a plausible account of intersubjectivity has to satisfy. And last but not least, my pragmatic approach allows for a richer *phenomenology* of intersubjectivity - one that does not need to be characterized solely by means of prediction and/or explanation. This is because many of the anticipatory and predictive processes that enable our meetings with other minds take place at the neurobiological level, and can be described in *sub-personal* terms.