CHAPTER 27

Reading 27.1 **Descartes, R. (1996). First meditation. *Meditations on First Philosophy*. Cambridge: Cambridge University Press, (Extract pp. 12–13).**

Reading 27.2 **Malcolm, N. (1958). Knowledge of other minds. *Journal of Philosophy*, 55. Reprinted in Rosenthal, D. (ed.) (1991). *The Nature of Mind*: Oxford: Oxford University Press, pp. 92–97. (Extract pp. 92–3).**

Reading 27.3 **Chihara, C.S. and Fodor, J.A. (1991). Operationalism and ordinary language: a critique of Wittgenstein. Reprinted in *The Nature of Mind* (ed. D. Rosenthal). Oxford: Oxford University Press, pp. 137-150 (Extract pp. 145–6).**

Reading 27.4 **Fodor, J.A. (1987). *Psychosemantics*. Cambridge, MA: MIT Press (Extract pp. 1-2).**

Reading 27.5 **Heal, J. (1995). Replication and functionalism. In *Folk Psychology* (ed. M. Davies and T. Stone). Oxford: Blackwell, pp. 45–59 (Extract pp. 45-7).**

Reading 27.6 **Stone, T. and Davies, M. (1996). The mental simulation debate: a progress report. In *Theories of Theories of Mind* (ed. P. Carruthers and P. Smith). Cambridge: Cambridge University Press, pp. 119-137 (Extract pp. 131–4).**

Reading 27.1 EXERCISE 1

From: Descartes, R. (1996). First meditation. *Meditations on First Philosophy*. Cambridge: Cambridge University Press, (Extract pp. 12–13).

Meditations on first philosophy

In which are demonstrated the existence of God and the distinction between the human soul and the body

First meditation

What can be called into doubt

Some years ago I was struck by the large number of falsehoods that I had accepted as true in my childhood, and by the highly doubtful nature of the whole edifice that I had subsequently based on them. I realized that it was necessary, once in the course of my life, to demolish everything completely and start again right from the foundations if I wanted to establish anything at all in the sciences that was stable and likely to last. But the task looked an enormous one, and I began to wait until I should reach a mature enough age to ensure that no subsequent time of life would be more suitable for tackling such inquiries. This led me to put the project off for so long that I would now be to blame if by pondering over it any further I wasted the time still left for carrying it out. So today I have expressly rid my mind of all worries and arranged for myself a clear stretch of free time. I am here quite alone, and at last I will devote myself sincerely and without reservation to the general demolition of my opinions.

But to accomplish this, it will not be necessary for me to show that all my opinions are false, which is something I could perhaps never manage. Reason now leads me to think that I should hold back my assent from opinions which are not completely certain and indubitable just as carefully as I do from those which are patently false. So, for the purpose of rejecting all my opinions, it will be enough if I find in each of them at least some reason for doubt. And to do this I will not need to run through them all individually, which would be an endless task. Once the foundations of a building are undermined, anything built on them collapses of its own accord; so I will go straight for the basic principles on which all my former beliefs rested.

Whatever I have up till now accepted as most true I have acquired either from the senses or through the senses. But from time to time I have found that the senses deceive, and it is prudent never to trust completely those who have deceived us even once.

Yet although the senses occasionally deceive us with respect to objects which are very small or in the distance, there are many other beliefs about which doubt is quite impossible, even though they are derived from the senses—for example, that I am here, sitting by the fire, wearing a winter dressing-gown, holding this piece of paper in my hands, and so on. Again, how could it be denied that these hands or this whole body are mine? Unless perhaps I were to liken myself to madmen, whose brains are so damaged by the persistent vapours of melancholia that they firmly maintain they are kings when they are paupers, or say they are dressed in purple when they are naked, or that their heads are made of earthenware, or that they are pumpkins, or made of glass. But such people are insane, and I would be thought equally mad if I took anything from them as a model for myself.

A brilliant piece of reasoning! As if I were not a man who sleeps at night, and regularly has all the same experiences[1] while asleep as madmen do when awake—indeed sometimes even more improbable ones. How often, asleep at night, am I convinced of just such familiar events—that I am here in my dressing-gown, sitting by the fire—when in fact I am lying undressed in bed! Yet at the moment my eyes are certainly wide awake when I look at this piece of paper; I shake my head and it is not asleep; as I stretch out and feel my hand I do so deliberately, and I know what I am doing. All this would not happen with such distinctness to someone asleep. Indeed! As if I did not remember other occasions when I have been tricked by exactly similar thoughts while asleep! As I think about this more carefully, I see plainly that there are never any sure signs by means of which being awake can be distinguished from being asleep. The result is that I begin to feel dazed, and this very feeling only reinforces the notion that I may be asleep.

[1] '. . . and in my dreams regularly represent to myself the same things' (French version).

Reading 27.2 EXERCISE 3

From: Malcolm, N. (1958). Knowledge of other minds. *Journal of Philosophy*, 55. Reprinted in Rosenthal, D, (ed.) (1991). *The Nature of Mind*. Oxford: Oxford University Press, pp. 92–97. (Extract pp. 92–3).

I believe that the argument from analogy for the existence of other minds still enjoys more credit than it deserves, and my first aim in this paper will be to show that it leads nowhere. J. S. Mill is one of many who have accepted the argument and I take his statement of it as representative . . .

Suppose this reasoning could yield a conclusion of the sort `it is probable that that human figure' (pointing at some person other than oneself) `has thoughts and feelings'. Then there is a question as to whether this conclusion can *mean* anything to the philosopher who draws it, because there is a question as to whether the sentence `That human figure has thoughts and feelings' can mean anything to him. Why should this be a question? Because the assumption from which Mill starts is that he has no *criterion* for determining whether another `walking and speaking figure' does or does not have thoughts and feelings. If he had a criterion he could apply it, establishing with certainty that this or that human figure does or does not have feelings (for the only plausible criterion would lie in behaviour and circumstances that are open to view), and there would be no call to resort to tenuous analogical reasoning that yields at best a probability. If Mill has no criterion for the existence of feelings other than his own then in that sense he does not understand the sentence `that human figure has feelings' and therefore does not understand the sentence `It is *probable* that the human figure has feelings'.

There is a familiar inclination to make the following reply: `Although I have no criterion of verification still I understand, for example, the sentence `He has pain'. For I understand the meaning of `I have pain' and `He has pain' means that he has the *same* thing I have when I have a pain.' But this is a fruitless manoeuvre. If I do not known how to establish that `someone has a pain' then I do not know how to establish that he has the *same* as I have when I have a pain. You cannot improve my understanding of `He has a pain' by this recourse to the notion of `the same', unless you give me a criterion for saying that someone *has* the same as I have. If you do this you will have no use for the argument from analogy; and if you cannot then you do not understand the supposed conclusion of that argument. (pp. 92-93)

Reading 27.3 EXERCISE 4

From: Chihara, C.S. and Fodor, J.A. (1991). Operationalism and ordinary language: a critique of Wittgenstein. Reprinted in *The Nature of Mind* (ed. D. Rosenthal). Oxford: Oxford University Press, pp. 137-150 (Extract pp. 145–6).

The Wittgensteinian argument of Section IV rest on the premiss that if we are justified in claiming that one can tell, recognize, see, or determine that '*Y*' applies on the basis of the presence of *X*, the either *X* is a criterion of *Y* or observations have shown that *X* is correlated with *Y*. Wittgenstein does not present any justification for this premiss in his published writings. Evidently, some philosophers find it self-evident and hence in need of no justification. We, on the other hand, far from finding this premiss self-evident, believe it to be false Consider: one standard instrument used in the detection of high-speed, charged particles is the Wilson cloud-chamber. According to present scientific theories, the formation of tiny, thin bands of for on the glass surface of the instrument indicates the passage of charged particles through the chamber. It is obvious that the formation of these streaks i not a Wittgensteinian criterion of the presence and motion of these particles in the apparatus. That one can detect these charged particles and determine their paths by means of such devices is surely not, by any stretch of the imagination, a *conceptual* truth. C.T.R. Wilson did not learn what "path of a charged particle" means by having the cloud chamber explained to him: he *discovered* the method, and the discovery was contingent upon recognizing the empirical fact that ions could ac as centers of condensation in a supersaturated vapor. Hence, applying Wittgenstein's own test for non-criterionhood (see above), the formation of a cloud-chamber track cannot be a criterion of the presence and motion of charged particles.

It is equally clear that the basis for taking these streaks as indicators of the paths of the particles is not observed *correlations* between streaks and some criterion of motion of charged particles. (What criterion for determining the path of an electron could Wilson have used to establish such correlations?) Rather, scientists were able to give compelling explanations of the formation of the streaks on the hypothesis that high-velocity, charged particles were passing through the chamber; on this hypothesis, further predictions were made, tested, and confirmed; no other equally plausible explanation is available; and so forth.

Such cases suggest that Wittgenstein failed to consider all the possible types of answers to the question, "What is the justification for the claim that one can tell, recognize, or determine that *Y* applies on the basis of the presence of *X*?" For, where *Y* is the predicate "is the path of a high-velocity particle," *X* need not have the form of either a criterion or a correlate.

Wittgensteinians may be tempted to argue that cloud-chamber tracks really are criteria, or symptoms observed to be correlated with criteria, of the paths of charged particles. To obviate this type of counter, we wish to stress that the example just given is by no means idiosyncratic. The reader who is not satisfied with it will easily construct others from the history of science. What is at issue is the possibility of a type of justification which consists in neither the appeal to criteria nor the appeal to observed correlations. If the Wittgensteinian argument we have been considering is to be compelling, some grounds must be given for the exhaustiveness of these types of justification. This, it would seem, Wittgenstein has failed to do.

It is worth noticing that a plausible solution to the problem raised in VI.5 can be given if we consider experiments with dreams and EEG to be analogous to the cloud-chamber case. That is, we can see how it could be the case that the correlation of EEG with dream reports was anticipated prior to observation. The dream report was taken by the experiments to be an indicator of a psychological event occurring prior to it. Given considerations about the relation of cortical to psychological events, and given also the theory of EEG, it was predicted that the EEG should provide an index of the occurrence of dreams. From the hypothesis that dream reports and EEG readings are both indices of the same psychological events, it could be deduced that they ought to be reliably correlated with one another, and this deduction in fact proved to be correct.

This situation is not at all unusual in the case of explanations based upon theoretical inferences to events underlying observable syndromes. As Meehl and Cronbach have pointed out, in such cases the validity of the "criterion" is often nearly as much at issue as the validity of the indices to be correlated with it. The successful prediction of the correlation on the basis of the postulation of a common etiology is taken both as evidence for the existence of the cause and as indicating the validity of each of the correlates as an index of its presence.

In this kind of case, the justification of existential statements is thus identical neither with an appeal to criteria nor with an appeal to symptoms. Such justifications depend rather on appeals to the simplicity, plausibility, and predictive adequacy of an explanatory system as a whole, so that it is incorrect to say that relations between statements which are mediated by such explanations are either logical in Wittgenstein's sense or contingent in the sense in which this term suggests simple correlation.

It cannot be stressed too often that there exist patterns of justificatory argument which are not happily identified either with appeals to symptoms or with appeals to criteria, and which do not in any obvious way rest upon such appeals. In these arguments, existential claims about states, events, and processes which are *not* directly observable are susceptible of justification despite the fact that no *logical* relation obtains between the predicates ascribing such states and predicates whose applicability *can* be directly observed. There is a temptation to hold that in such cases

there *must* be a criterion, that there must be some set of possible observations which would settle *for sure* whether the theoretical predicate applies. But we succumb to this temptation at the price of postulating stipulative definitions and conceptual alterations which fail to correspond to anything we can discover in the course of empirical arguments. The counter-intuitive features of philosophic analyses based on the assumption that there must be criteria are thus not the consequences of a profound methodological insight, but rather a projection of an inadequate philosophical theory of justification.

Reading 27.4 EXERCISE 5

From: Fodor, J.A. (1987). *Psychosemantics*. Cambridge, MA: MIT Press (Extract pp. 1-2).

A Midsummer Night's Dream, act 3, scene 2.

Enter Demetrius and Hermia.

Dem. O, why rebuke you him that loves you so?
Lay breath so bitter on your bitter foe.

Herm. Now I but chide, but I should use thee worse;
For thou, I fear, hast given me cause to curse.
If thou hast slain Lysander in his sleep,
Being o'er shoes in blood, plunge in the deep,
And kill me too.
The sun was not so true unto the day
As he to me: would he have stol'n away
From sleeping Hermia? I'll believe as soon
This whole earth may be bor'd; and that the moon
May through the centre creep, and so displease
Her brother's noontide with the antipodes.
It cannot be but thou hast murder'd him;
So should a murderer look; so dead, so grim.

Very nice. And also very *plausible*; a convincing (though informal) piece of implicit, nondemonstrative, theoretical inference.

Here, leaving out a lot of lemmas, is how the inference must have gone: Hermia has reason to believe herself beloved of Lysander. (Lysander has told her that he loves her—repeatedly and in elegant iambics—and inferences from how people say they feel to how they do feel are reliable, ceteris paribus.) But if Lysander does indeed love Hermia, then, a fortiori, Lysander wishes Hermia well. But if Lysander wishes Hermia well, then Lysander does not voluntarily desert Hermia at night in a darkling wood. (There may be lions. "There is not a more fearful wildfowl than your lion living.") But Hermia was, in fact, so deserted by Lysander. Therefore not voluntarily. Therefore *in*voluntarily. Therefore it is plausible that Lysander has come to harm. At whose hands? Plausibly at Demetrius's hands. For Demetrius is Lysander's rival for the love of Hermia, and the presumption is that rivals in love do *not* wish one another well. Specifically, Hermia believes that Demetrius believes that a live Lysander is an impediment to the success of his (Demetrius's) wooing of her (Hermia). Moreover, Hermia believes (correctly) that if x wants that P, and x believes that not-P unless Q, and x believes that x can bring it about that Q, then (ceteris paribus) x tries to bring it about that Q. Moreover, Hermia believes (again correctly) that, by and large, people succeed in bringing about what they try to bring about. *So*: Knowing and believing all this, Hermia infers that perhaps Demetrius has killed Lysander. And we, the audience, who know what Hermia knows and believes and who share, more or less, her views about the psychology of lovers and rivals, understand how she has come to draw this inference. We sympathize.

In fact, Hermia has it all wrong. Demetrius is innocent and Lysander lives. The intricate theory that connects beliefs, desires, and actions—the implicit theory that Hermia relies on to make sense of what Lysander did and what Demetrius may have done; and that *we* rely on to make sense of Hermia's inferring what she does; and that Shakespeare relies on to predict and manipulate our sympathies (*'deconstruction' my foot*, by the way)—this theory makes no provision for nocturnal interventions by mischievous fairies. Unbeknownst to Hermia, a peripatetic sprite has sprung the ceteris paribus clause and made her plausible inference go awry. "Reason and love keep little company together now-a-days: the more the pity that some honest neighbours will not make them friends."

Granting, however, that the theory fails from time to time—and not just when fairies intervene—I nevertheless want to emphasize *(1) how often it goes right, (2) how deep it is, and (3) how much we do depend upon it*. Commonsense belief/desire psychology has recently come under a lot of philosophical pressure, and it's possible to doubt whether it can be saved in face of the sorts of problems that its critics have raised. There is, however, a prior question: whether it's worth the effort of trying to save it. That's the issue I propose to start with.

Reading 27.5 EXERCISE 6

From: Heal, J. (1995). Replication and functionalism. In *Folk Psychology* (ed. M. Davies and T. Stone). Oxford: Blackwell, pp. 45–59 (Extract pp. 45-7).

1 The Functional Strategy versus the Replicative Strategy

In this paper I want to examine two contrasted models of what we do when we try to get insight into other people's thoughts and behaviour by citing their beliefs, desires, fears, hopes, etc. On one model we are using what I shall call the *functional strategy* and on the other we use what I label the *replicative strategy.* I shall argue that the view that we use the replicative strategy is much more plausible than the view that we use the functionalist strategy. But the two strategies issue in different styles of explanation and call upon different ranges of concepts. So at the end of the paper I shall make some brief remarks about these contrasts.

The core of the functionalist strategy is the assumption that explanation of action or mental state through mention of beliefs, desires, emotions, etc. is causal. The approach is resolutely third personal. The Cartesian introspectionist error—the idea that from some direct confrontation with psychological items in our own case we learn their nature—is repudiated. We are said to view other people as we view stars, clouds or geological formations. People are just complex objects in our environment whose behaviour we wish to anticipate but whose causal innards we cannot perceive. We therefore proceed by observing the intricacies of their external behaviour and formulating some hypotheses about how the insides are structured. The hypotheses are typically of this form: 'The innards are like this. There is some thing or state which is usually caused by so and so in the environment (let us call this state "*X*") and another caused by such and such else (let us call this "*Y*"); together these cause another, "*Z*", which, if so and so is present, probably leads on to . . . ' And so on. It is in some such way as this that terms like 'belief' and 'desire' are introduced. Our views about the causes, interactions and outcomes of inner states are sometimes said to be summed up in 'folk psychology' (Stich, 1982a, p. 153ff). Scientific psychology is in the business of pursuing the same sort of programme as folk psychology but in more detail and with more statistical accuracy. On this view a psychological statement is an existential claim—that something with so-and-so causes and effects is occurring in a person (Lewis, 1972). The philosophical advantages, in contrast with dualism and earlier materialisms such as behaviourism and type-type identity theory, are familiar. It is via these contrasts and in virtue of these merits that the theory emerged. See Putnam (1967) for a classic statement.

This is a broad outline. But how is psychological explanation supposed to work in particular instances? What actual concepts are employed and how, in particular, are we to accommodate our pre-theoretical idea that people have immense numbers of different beliefs and desires, whose contents interrelate?

Functionalists would generally agree that there is no hope of defining the idea of a particular psychological state, like believing that it is raining, in isolation from other psychological notions. Such notions come as a package, full understanding of any member of which requires a grip on its role in the system as a whole (Harman, 1973). This is true of any interesting functional concepts, even, for example, in explaining functionally something as comparatively simple as a car. If we try to build up some picture of the insides of a car, knowing nothing of mechanics and observing only the effects of pushing various pedals and levers and inserting various liquids, we might well come up with ideas like 'engine', 'fuel store', 'transmission', etc. But explanation of any one of these would clearly require mention of the others. Similarly we cannot say what a desire is except by mentioning that it is the sort of thing which conjoins with beliefs (and other states) to lead to behaviour.

But something more important than this is that the number of different psychological states (and hence their possibilities of interaction) are vastly greater than for the car. There is no clear upper bound on the number of different beliefs or desires that a person may have. And, worse, we cannot lay down in advance that for a given state these and only these others could be relevant to what its originating conditions or outcome are. This 'holism of the mental' (Quine, 1960, Davidson, 1970) which is here only roughly sketched, will turn out to be of crucial significance and we shall return to it. But for the moment let us ask how the functionalist can accommodate the fact that, finite creatures as we are, we have this immensely flexible and seemingly open-ended competence with psychological understanding and explanation. A model lies to hand here in the notions of axioms and theorems. We have understanding of hitherto unencountered situations because we (in some sense) know some basic principles concerning the ingredients and modes of interaction of the elements from which the new situations are composed.

What can the elements be? Not individual beliefs and desires because, as we have seen, there are too many of them. Hence the view that having an individual belief or desire must be, functionally conceived, a composite state. This is one powerful reason why the idea of the having of beliefs and desires as relations to inner sentences seems attractive (Field, 1978, pp. 24–36). The functional psychologist hopes that, with a limited number of elements (inner words), together with principles of construction and principles of interaction (modelled on the syntactic transformations of formalized logic), the complexity of intra-subjective psychological interactions can be encapsulated in a theory of manageable proportions.

But, however elegantly the theory is axiomatized the fact remains that it is going to be enormously complex. Moreover we certainly cannot now formulate it explicitly. There should therefore be some reluctance to credit ourselves with knowing it (even if only implicitly) unless there is no alternative account of how psychological explanation could work. But there is an alternative. It is the replicating strategy to which I now turn.

On the replicating view psychological understanding works like this. I can think about the world. I do so in the interests of taking my own decisions and forming my own opinions. The future is complex and unclear. In order to deal with it I need to and can envisage possible but perhaps non-actual states of affairs. I can imagine how my tastes, aims and opinions might change, and work out what would be sensible to do or believe in the circumstances. My ability to do these things makes possible a certain sort of understanding of other people. I can harness all my complex theoretical knowledge about the world and my ability to imagine to yield an insight into other people *without any further elaborate theorizing about them.* Only one simple assumption is needed: that they are like me in being thinkers, that they possess the same fundamental cognitive capacities and propensities that I do.

The method works like this. Suppose I am interested in predicting someone's action. (I take this case only as an example, not intending thereby to endorse any close link between understanding and prediction in the psychological case. Similar methods would apply with other aspects of understanding, for example, working out what someone was thinking, feeling or intending in the past.) What I endeavour to do is to replicate or recreate his thinking. I place myself in what I take to be his initial state by imagining the world as it would appear from his point of view and I then deliberate, reason and reflect to see what decision emerges.

Reading 27.6 EXERCISE 7

From: Stone, T. and Davies, M. (1996). The mental simulation debate: a progress report. In *Theories of Theories of Mind* (ed. P. Carruthers and P. Smith). Cambridge: Cambridge University Press, pp. 119-137 (Extract pp. 131–4).

4 The shape of the debate

If we leave the concept mastery question to one side, and just focus on the question about the resources that are drawn upon, then the debate seems to come down to this. The theory-theory says that our ability to negotiate the social world depends upon our possessing a body of empirical knowledge about how people's situations, mental states, and behaviour are related. The simulation alternative needs to find a distinctive place for the ability to engage in imaginative identification while also denying that our folk psychological practice relies upon possession of a body of psychological knowledge.

There are then a number of ways in which the distinction between the two sides of this debate might be blurred. One particular kind of threat of collapse arises if the theory-theory makes use of the idea of tacit knowledge of a theory. For the notion of possessing tacit knowledge has to be defined, and once a definition is offered it is a substantive question whether someone who engages in mental simulation—as that is described by Goldman or Gordon, say counts as having tacit knowledge of a psychological theory. If the notion of tacit knowledge is defined so thinly that a simulator also counts as a tacit knower of a psychological theory then, in a quite strict sense, the distinction between the two sides collapses (see Davies, 1994; Heal, 1994; Perner, 1994, this volume).

But, even supposing that the basic distinction between simulating or reenacting and drawing upon a body of psychological knowledge remains intact, still there are ways in which the theory-theory approach and the simulation alternative might be argued to overlap.

4.1 Opposed or complementary approaches?

One line of argument that has been present from the outset of the debate is that a simulator needs to draw upon a body of psychological knowledge in order to carry through a simulation. Thus, Dennett (1981/1987, pp. 100 1):

> How can it [sc. simulation] work without being a kind of theorising in the end? For the state I put myself in is not belief but make-believe belief. If I make believe I am a suspension bridge and wonder what I will do when the wind blows, what 'comes to me' in my make-believe state depends on how sophisticated my knowledge is of the physics and engineering of suspension bridges. Why should my making believe I have your beliefs be any different? In both cases, knowledge of the imitated object is needed to drive the make-believe 'simulation' and the knowledge must be organised into something rather like a theory.

Consider again Gordon's example in which my friend and I are walking along the trail. Surely, the theory-theorist will say, in order to explain or predict my friend's action of running away, I need to know something about the typical causal relations between recognising a bear, being afraid, and taking evasive action. In addition, I need to know something about my friend's psychological make-up, and that knowledge, too, will be dependent upon pieces of theory (about the attitude towards bears that tends to be produced by a certain kind of education, for example).

The simulation theorist has a number of responses to make at this point. Particularly, the simulation theorist argues that the theory-theorist is making an unwarranted and unparsimonious assumption. This assumption is that, over and above any actual thinking (say, thinking about tracks, bears and escape) that takes place whether *in propria persona* or within the scope of a simulation—there are also general psychological principles, knowledge of which explains the movements of thought that occur during a period of thinking, or within some episode of simulation. The simulation alternative sees no need for these two layers of thought: a layer that is the actual episode of thinking about the world, and a layer of meta-thinking that brings about the movement of thought in the first layer. All that is needed, according to the simulation theorist, is that some thinking take place, in accordance with the canons of rational cognition. The dynamics of thought require no meta-cognitive engine.

This line of argument is used to reject the suggestion that mental simulation will inevitably be 'theory driven' rather than 'process driven' (Goldman, 1989). But it can also be used to undermine the theory-theory itself. For what the theory-theory appears to be committed to is not just knowledge of some general psychological principles, but also knowledge of indefinitely many indefinitely detailed principles about thought concerning specific subject matters (Heal, 1995, this volume). Botterill (this volume), Carruthers (this volume), and Perner (this volume) all concede, on behalf of the theory-theory, that the intrusion of something like simulation will be needed—what Carruthers calls 'simulation within a theory' and Perner calls 'content simulation'.

On the other side, advocates of the simulation theory typically acknowledge that inductively based generalisations play a role in real life use of mental simulation (Goldman, 1989, Harris, 1992), and Perner (this volume) presents empirical data that points in the same direction. So, to that extent, we seem to be moving towards a measure of agreement over the need for hybrid theories. Future research will need to address in detail the ways in which simulation and deployment of knowledge interact. As Perner says (this volume, p. 103): [S]ince any theory use involves an element of simulation and since simulation on its own cannot account for the data, the future must lie in a mixture of simulation and

theory use. However, what this mixture is and how it operates must first be specified in some detail before any testable empirical predictions can be derived.'

4.2 **Cognitive penetrability: The crucial test?**

The point about the relative lack of economy involved in adoption of the theory-theory approach is made vivid by an example introduced by Paul Harris (1992). Suppose we are asked to predict the grammaticality judgements that a speaker of the same language would make when confronted with a range of sentences in the language. Harris reasonably claims that predictive success would be high, and offers this explanation (1992, p. 124):

> The most plausible answer is that you read each sentence, asked yourself whether it sounded grammatical or not, and assumed that other English speakers would make the same judgements for the same reasons. The proposal that you have two distinct tacit representations of English grammar, a first-order representation that you deploy when making your own judgements, and a metarepresentation (i.e. a representation of other people's representations) that you deploy in predicting the judgements made by others, so designed as to yield equivalent judgements, strains both credulity and parsimony.

Furthermore, we might suppose, what goes for the prediction of grammaticality judgements goes also for the prediction of belief formation on the basis of inference and for the prediction of intentions and behaviour.

However, Stich and Nichols (1995) respond to Harris's example by distinguishing cases. They are inclined to make concessions to the simulation theory over the explanation and prediction of a person's judgements and beliefs in a particular perceptual situation, and perhaps also in the case of belief-formation on the basis of inference. But they argue strongly for the theory-theory as providing a better account of the prediction of behaviour. Given the role of theoretical inference in any decision-taking process, we are not convinced that different stances towards prediction of belief formation on the basis of inference and prediction of behaviour can be justified. But that is not our main concern here.

Stich and Nichols' argument hinges on the phenomenon of cognitive penetrability, introduced in an earlier paper (1992). The issue is this. If predictions are based upon deployment of a theory (a body of information or misinformation), then those predictions are liable to be false if the theory is incorrect in any way. Theory-based predictions are subject to error introduced by misinformation. But a flawed theory will obviously have no impact upon predictions that do not draw upon it. Even if I have deeply flawed psychological views, if I use mental simulation to generate predictions about what people will do given their beliefs and desires, then my flawed theory need introduce no error into my predictions.

So the crucial empirical question is, apparently, whether we are liable to make false predictions about other people's decisions, intentions, and actions. If we are then, according to Stich and Nichols (1992, 1995), this favours the theory-theory. And, as they point out, there are indeed examples where folk psychological prediction lets us down.