THE DOMAIN OF FOLK PSYCHOLOGY
(Forthcoming in A. O'Hear (Ed.), *Mind and Persons*, Cambridge UP)

José Luis Bermúdez
University of Stirling

My topic in this paper is social understanding. By this I mean the cognitive skills underlying social behaviour and social coordination. Normal, encultured, non-autistic and non-brain-damaged human beings are capable of an impressive degree of social coordination. We navigate the social world with a level of skill and dexterity fully comparable to that which we manifest in navigating the physical world. In neither sphere, one might think, would it be a trivial matter to identify the various competences which underly this impressive level of performance. Nonetheless, at least as far as interpersonal interactions are concerned, philosophers show a rare degree of unanimity. What grounds our success in these interactions is supposed to be our common mastery of (more or less similar versions of) folk psychology.

Most philosophers would, I think, be inclined to agree on something like the following minimal characterisation of folk psychology.

> Human beings are social creatures. And they are reflective creatures. As such they continually engage in a host of cognitive practices that help them get along in their social world. In particular, they attempt to understand, explain and predict their own and others' psychological states and overt behaviour; and they do so by making use of an array of ordinary psychological notions concerning various internal mental states, both occurrent and dispositional. Let us then consider folk psychology to consist, *at a minimum,* of (a) a set of attributive, explanatory and predictive practices, and (b) a set of notions or concepts used in those practices. (Von Eckhardt 1994,)

No philosopher has, as far as I know, denied that this set of attributive, explanatory and predictive practices exists, nor that these practices implicate a network of psychological concepts. In the following I shall use the expression "folk psychology" in what I take to the standard way – namely, as picking out certain practices of ascribing propositional attitudes (and other mental states) to other agents and explaining/predicting their behaviour on the basis of those attributions.

The few challenges that have been levelled at folk psychology have been directed at its explanatory adequacy. Paul Churchland and other eliminative materialists have suggested that folk psychology will eventually be replaced by a theory capable of dealing with complexities that folk psychology cannot tackle – a theory that will be derived from neuroscience rather than from commonsense psychological concepts. They do not doubt, however, that folk psychology is currently our dominant tool for interpersonal cognition. The issue they contest is the security of its dominance. There is also considerable debate about the particular way in which folk psychology should be understood. The idea, popular among both philosophers and psychologists, that folk psychology is essentially theory-like has come under pressure from the suggestion that folk psychological understanding is essentially grounded in empathetic simulation rather than upon applying principles about how different types of mental state might interact and manifest themselves in behaviour. But participants in this debate still see it as a debate about how to understand folk psychology – that is to say about how we actually go about attributing attitudes and employing those attributions to explain and predict behaviour.[1]

Strikingly, the centrality of folk psychology in explaining what makes social interactions possible has rarely been challenged. It is taken for granted by almost everyone that we accommodate ourselves to the behaviour of our fellow human beings by attributing to them beliefs and desires (or other propositional attitudes), assuming that they will act in a way that is more or less rational given those beliefs and desires, and then working out the courses of action which that rationality assumption appears to dictate. Of course, everyone will agree that we don't *consciously* go about doing this, but is usual to think that, in the vast majority of cases, this practice of attribution, explanation and prediction proceeds tacitly. The analogy with linguistic understanding is suggestive at this point.[2] Gricean accounts of linguistic meaning and understanding propose that

[1] See, for example, the introduction to Davies and Stine 1995.
[2] In fact, it is not really an analogy, since neo-Gricean accounts of linguistic understanding reflect precisely the hegemony of folk psychological categories in social interaction and understanding that this paper is questioning.

the process of understanding an interlocutor involves making complex assessments of his conversational intentions. To the objection that we have no awareness of doing any such thing (except in non-standard cases of irony and so forth) proponents of such accounts think that our attributions reflect the tacit mastery and application of a theory of conversational implicature.

There can be no doubt, of course, but that we do engage in folk psychological explanation, sometimes working forwards from what we know about someone's beliefs and desires to what we think they will do, and sometimes working backwards from their behaviour and general knowledge of how their minds work to their particular motivations for acting in a certain way. The question is whether this provides a model for social understanding in general. Granted that we sometimes do make reflective and explicit use of the concepts of folk psychology in making sense of the behaviour of others, should we conclude that our *unreflective* social understanding involves an implicit application of the concepts of folk psychology in the interests of explanation and prediction? Should we conclude that all our social understanding involves deploying the concepts and explanatory/predictive practices of folk psychology?

With these questions in mind, let me distinguish two conceptions of the domain of folk psychology - or, more accurately, the two ends of a spectrum of conceptions of the domain of folk psychology. At one end lies the narrow construal of the domain of folk psychology. According to the narrow construal, the domain of folk psychology should not be presumed to extend further than those occasions on which we explicitly and consciously deploy the concepts of folk psychology in the services of explanation and/or prediction. At the other end of the spectrum lies the broad construal, which makes all social understanding a matter of the attribution of mental states and the deployment of those attributed states to explain and predict behaviour.

There are several reasons why the dominant conception in contemporary philosophy of mind of the domain of folk psychology is broad rather than narrow. The first is that the philosophy of mind tends to operate with a clearcut distinction between two ways of understanding behaviour.

We can either understand behaviour in intentional terms, as rationalized by propositional attitudes, or in non-intentional terms. The standard examples in philosophy of mind textbooks (and indeed in philosophy of action textbooks) of behaviours not to be understood in intentional terms are those which are either reflex or not properly attributable to the agent. It is standard to distinguish, for example, between an arm-raising that is intentional, comprehensible as issuing from a particular nexus of beliefs and desires, and one that is the result of a reflex response, or of someone else lifting my arm for me. It seems pretty clear that social understanding does not involve understanding the behaviour of others in either of these latter two ways. So, if the choice really is a stark one between behavior being unintentional in one of these senses, on the one hand, and its being intentional in the sense of being rationalized by propositional attitudes on the other, then it is easy to see why unreflective social understanding should be thought to involve the tacit application of folk psychology.

Secondly, a folk psychological construal of social understanding goes hand in hand with the dominant understanding of the "springs of action" - not surprisingly since in many cases the way in which we understand an action is intimately related to how we take that action to have come about.[3] Many philosophers assume that we act in virtue of our beliefs and desires and understand the "in virtue of" to be causal in nature. The way in which this view is developed usually leaves no room for "thinking behaviour" that is not causally generated by beliefs and desires. To the extent, then, that the activities of social coordination are thinking activities, the folk psychological construal seems to follow immediately.

At a more general level, and this is an idea which can be traced at least as far back as Kant's kingdom of ends, lies the background thought that it is a mark of the distinctiveness of persons that we treat them in a fundamentally different manner from the way in which we treat non-

[3] I mention this as a separate point because of the influential minority of philosophers who deny this reciprocal relation between the terms in which we understand action and how we take those actions to have been caused. The type of instrumentalism canvassed by Daniel Dennett falls into this category.

persons. The idea that we have this distinctive way of treating persons is bound up with the idea that there is a distinctive mode of explanation appropriate to the behaviour of persons and not appropriate to other parts of the animate world, namely, personal-level intentional explanation which treats agents as more or less rational beings striving to satisfy their desires and aspirations in the light of the information they possess about the world. So, one might think, it is an integral part of what it is to treat persons as persons that we should strive to explain their behaviour in terms of the broadly normative concepts of folk psychology.[4]

However, I can see little evidence to support the tacit assumption that belief-desire psychology is even universally applicable, let alone universally applied. There is a range of reasons for thinking that the application of folk psychology is a relatively circumscribed part of our social and cognitive life and the burden of this paper will be that our conception of the domain of folk psychology should be far closer to the narrow end of the spectrum than to the broad end. In section II I will discuss how the issue between the narrow and broad construal of the domain of folk psychology relates to the debate between theory-theory and simulationist conceptions of folk psychology. Both the simulationist and theory-theory camps can be understood in either broad or narrow terms – although, as it happens, they are almost invariably understood in broad terms. In section III I will present some general considerations counting against the broad construal. Section IV will explore concrete examples of how different types of social interaction might involve non-folk-psychological forms of social understanding.

II.

The issue between narrow and broad construals of the domain of folk psychology is orthogonal to the debate about whether folk psychology should be understood in terms of the application of a theory or in terms of empathetic simulation (see the papers collected in Davies and Stone 1995a

[4] For further discussion of the distinction between personal and subpersonal explanation see Bermúdez 2000 and the other essays in Bermúdez and Elton 2000.

and 1995b and in Carruthers and Smith 1997). According to proponents of the so-called theory-theory, folk psychological understanding involves the application of a tacitly known network of principles connecting mental states to each other and to behaviour. Simulationists, on the other hand, think that we explain and predict the behaviour of other agents by projecting ourselves into the situation of the person whose behaviour is to be explained/predicted and then using our own mind as a model of theirs. This involves running our own decision-making processes off-line taking as inputs the mental states that it seems appropriate for the other person to have in that situation. Both the simulation theory and the theory-theory think that social understanding proceeds essentially by the attribution of beliefs and desires. The issues separating the theory-theorist and the simulationist are (a) how we arrive at the attributions of beliefs and desires and (b) how we get from those attributions to explanations/predictions. The following passage from Gregory Currie makes clear both the differences between the two positions and the ground they share.

> Simulation theorists say that our access to the thoughts of others is not through the application of a primitive but effective theory, as advocates of the "theory-theory" of folk psychology suppose, but through a kind of internal, largely spontaneous, re-enactment that allows us to imagine ourselves in some rough approximation to the situation of another. In so imagining, we tend to acquire, in imagination, the beliefs and desires an agent would most likely have in that situation, and those imaginary beliefs and desires have consequences in the shape of further pretend beliefs and desires as well as pretend decisions that mimic the beliefs, desires and decisions that follow in the real case. (Currie 1995 p.158)

Both parties to the debate think that we arrive at predictions and explanations by moving from beliefs and desires, either through theoretical principles that link particular complexes of beliefs and desires to particular behaviours or through working out what one would oneself do in that situation with those beliefs and desires.

It is true that simulationists often write as if simulations proceed simply by "putting oneself into the perspective of another". Here is another, more ambiguous passage from Gregory Currie

> [The simulation view] says that I understand the minds of others by imaginatively projecting myself into their situations and using my own mind as a model of their's. Running my own mental states "off-line", I am able to *simulate* the mental processes

> of another, and thereby to learn, for example, what decision he will make. (Ibid. p.242)

The ambiguity concerns whether the "situation" into which I project myself includes some specification of what the other person believes, desires, hopes and so forth, or whether I simply adopt their perceptual perspective on the world and start simulating from there. It seems clear, however, that we cannot simulate someone, except in the most straightforward of cases, simply by adopting their perceptual perspective on the world. That would give us a fairly accurate way of determining how they are likely to be perceiving the world, but in order to work out how they might act on the basis of what they perceive we need to know how those perceptions are likely to feed into their propositional attitudes more widely understood. And nor, of course, can we work out what people are likely to believe simply by adopting their perceptual perspective - we need to attribute some attitudes to get started on the process of simulating the attitudes that someone might acquire in a given situation.

The debate between mainstream simulationists and theory-theorists is in an important sense orthogonal to the question of whether we should construe folk psychology in narrow or broad terms. The question of the domain of folk psychology is the question of whether we should make sense of our unreflective practices of social coordination in terms of the tacit application of the explanatory and predictive practices of explicit folk psychology. One's answer to this should not be affected by how one understands the explanatory and predictive practices of explicit folk psychology. The simulation theory is perfectly compatible with a broad construal of folk psychology. On such a view social coordination would be underpinned by constant off-line simulations of the mental states and processes of other participants, not to mention simulations of their simulations of oneself. But the simulation theory can equally be interpreted in narrow terms, as applying only to the relatively infrequent occasions when we make an explicit effort to make sense out of someone's behaviour. The same holds of the theory-theory. There is nothing about the theory-theory which demands that it be applied broadly rather than narrowly.

III

There are some very general reasons for being inclined towards a narrow rather than a broad conception of folk psychology. Some stem from considerations of cognitive architecture and the structure of the mind. Others stem from considerations of computational complexity. Both sets of reasons are ultimately suggestive rather than conclusive. The real argument, if it can be called that, will come from providing concrete examples of how areas of social understanding can be modelled in non-folk psychological terms. The general reasons will occupy us in this section and the concrete examples in section IV.

The computational argument is straightforward. It is motivated by the thought that the vast majority of our social interactions involve almost instantaneous adjustments to the behaviour of others, whereas folk psychological explanation is a complicated and protracted business, whether it is understood according to the simulation theory or the theory-theory. It is no easy matter to attribute beliefs and desires and then to work either backwards from those beliefs and desires to an explanation or forwards to a prediction. The point is perhaps easiest to see with respect to the theory-theory. To apply folk psychological explanation is to subsume observable behaviour and utterances under general principles linking observable behaviour to mental states, mental states to other mental states and mental states to behaviour. As many authors have stressed, the application of these principles requires identifying, among a range of possible principles which might apply, the ones that are the most salient in a given situation. It requires identifying whether the appropriate background conditions hold, or whether there are countervailing factors in play. It requires thinking through the implications of the principles one does choose to apply in order to extrapolate their explanatory/predictive consequences. It would be an overreaction to suggest that the need to do all these things makes folk psychological generalizations essentially useless. But it certainly makes them rather unwieldy. And it is no surprise that the paradigms of folk psychological explanations given by theory-theorists tend to be complicated inferences of the sort

either found in the final chapters of detective novels (e.g. Lewis 1972) or in dramatic and self-questioning soliloquies (e.g. Fodor 1987). These are indeed striking cognitive achievements, but it seems odd to take them as paradigms of interpersonal cognition. Do our everyday cognitive interactions with people really involve deducing hypotheses from general principles, drawing out the deductive consequences (more accurately: the *relevant* deductive consequences) of those general principles and then putting those hypotheses before the tribunal of experience? If that is what is required then it is a wonder that such a thing as social coordination exists.

The practical difficulties here are obscured by the narrow range of examples which tend to be considered. Folk psychological explanation is usually considered by philosophers to be a one-on-one activity. This is exactly what one would expect given that the paradigms are the detective drawing together the strands of the case, or the puzzled lover trying to decode the behaviour of her paramour. But social understanding is rarely as circumscribed as this. In many examples of social coordination there is a range of people involved and the behaviour of any one of them is inextricably linked with the behaviour of the others. Suppose that the social understanding involved in such examples of social coordination is modeled folk psychologically. This would require each participant to make predictions about the likely behaviour of other participants, based on an assessment of what those participants want to achieve and what they believe about their environment. For each participant, of course, the most relevant part of the environment will be the other participants. So, my prediction of what another participant will do depends upon my beliefs about what they believe the other participants will do. The other participant's beliefs about what the other participants will do are in turn dependent upon what they believe the other participants believe. And so on.

It is clear that there will be many layers in the ensuing regress, and that the process of coming to a stable set of beliefs which will allow one to participate effectively in the coordinated activity will be lengthy and computationally demanding. Of course, none of this shows that there are any objections in principle to modeling coordinative social understanding in folk psychological terms.

Any such claim would be absurd, not least because we have a well worked out mathematical theory that allows us to model social understanding in what are essentially folk psychological terms (or at least a regimentation of them). Game theory is a theory of social coordination and strategic interaction employing analogues of the folk psychological notions of belief and desires (in the guise of probability and utility assignments). What thinking about computational tractability should do, however, is at least to begin to cast doubt upon whether this could be a correct account of the form of social understanding in the vast majority of situations.

It is important to distinguish this point from another charge levelled at the theory-theory. Simulation theorists have sometimes suggested that issues of computational tractability work in favour of the simulation theory. Jane Heal, for example, has argued that theory-theorists run into difficulties analogous to the frame problem in computer science (Heal 1996). The frame problem is essentially the problem of determining which, among the myriad aspects and deductive consequences of a principle or of a belief, are relevant in a given situation (Dennett 1984). Any psychological theory incorporating a satisfactory response to the frame problem will of necessity incorporate a theory of relevance, specifying why certain psychological factors will be deemed relevant in particular situations but not in others, how changing the parameters of a situation can radically alter those aspects of it relevant to decision-making and how what is taken to be relevant can vary systematically with determinate aspects of the psychology of the individual. It is, according to Heal, a weighty consideration against the theory-theory that any such, presumably tacitly known, theory of relevance would be far more complex than any other postulated tacit theory to explain, for example, our grasp of grammar or of so-called naïve physics.

This worry is well grounded (although one might wonder whether a simulation theorist can avoid postulating at some level a tacitly known theory of relevance governing both our on-line decision-making processes and our off-line simulations). But it is orthogonal to the computational worry I am raising. That computational worry would still be there even if we granted the theory-theorist the legitimacy of postulating a tacitly known theory of relevance. The worry about

relevance is a worry about how it is even possible to tailor the generality of folk psychological principles to the particularity of specific situations. The computational worry, on the other hand, is about the combinatorial explosion that will occur when the situation in question involves several individuals who are potentially collaborating. Even if we can fix the parameters of relevance in a way that will permit folk psychological principles to come into play, the key problem comes from the fact that the application of folk psychological explanation to a multi-agent interaction will require a computationally intractable set of multiply embedded higher-order beliefs about beliefs.

The worry about combinatorial explosion is not confined to the theory-theory. Let us suppose that the simulation theory can get by without having to assume a tacitly known theory of relevance, so that a simulation simply involves using one's own mind as a model of the minds of the other participants in the interaction. One would still need to plug into the decision-making processes an appropriate set of inputs for all the other participants and then run simultaneous simulations for all of them. This is multiply problematic. There is, first of all, a straightforward question about how many simulations it is actually possible to run simultaneously. Since the practical details of how the process of simulation might work have not really been explored, there is little concrete to say about this. *Prima facie*, however, one might think that there will be some difficulties with the idea of multiple simultaneous simulations, given that a simulation is supposed to work by running one's own decision-making processes off-line and those processes are presumably designed to give an output for a single set of inputs. But there is a more serious problem. The simultaneous simulations will not be independent of each other. Suppose that the interaction contains three participant, A, B and C, in addition to me. In order to simulate B properly I will need to have views about what A and C will do - without that information I will not have any sense of what initial beliefs it would be reasonable to attribute to B. But, by parity of reasoning, this information about what A and C will do will depend upon each of them having information about what the other participants will do. It is very difficult to see how the notion of simulation can be stretched to accommodate, not just simultaneous simulations, but simultaneous

simulations which are interdependent. So, the simulation theory, no less than the theory-theory, is bound to confront problems of computational tractability if it adopts a broad construal of the domain of folk psychology.

Let us turn now to the second general reason for scepticism about the broad intepretation of the domain of folk psychology. Here I will be painting with very broad strokes of the brush indeed. Folk psychological reasoning is a paradigm of metarepresentational thinking, where metarepresentational thinking involves thinking about thoughts – taking thoughts as the objects of thought, attributing them to other subjects, evaluating their inferential connections with other thoughts, and so on. It has been suggested that metarepresentational thinking is in some sense language-dependent (Dennett 1996, Bermúdez 2003). One might argue, for example, that thoughts must be vehicled in a way that is consciously and reflectively accessible if they are to feature in metarepresentational thinking, and that the only possible vehicles are linguistic. I have developed this line of thought in more detail elsewhere (Bermúdez 2003 Ch. 7). If the thesis of language-dependence is correct then it seems highly likely, on the basis of our best current theories of cognitive archeology, that many of the cognitive skills involved in social coordination emerged long before the capacity for metarepresentational thinking, and hence long before folk psychological explanation was even possible.[5] Early hominids appear to have been capable of an impressive range of types of collective behaviour, involving the social transmission of knowledge (e.g. knowledge of the natural world); the tracking of social relations within social groups; complex forms of social coordination (in hunting and migratory behaviour) and technical training in tool manufacture (Mithen 1996). All these forms of social coordination require high degrees of social understanding. *Ex hypothesis* this social understanding could not have involved the concepts and explanatory/predictive strategies of folk psychology.

[5] In fact, this suggestion is independently plausible even without the thesis of language-dependence.

Of course, this does not allow us to draw any immediate inferences about the current state of our social cognition – perhaps the metarepresentational abilities that emerged with language acquisition (or at any rate relatively late in cognitive evolution) simply wrote over their primitive precursors, in the way that some developmental psychologists think that the earliest conceptions of the physical world acquired in infancy are completely superceded by the "naïve physics" emerging later in development (Gopnik and Meltzoff 1997). That would doubtless be the position of those who adopt what I have termed the broad construal of the domain of folk psychology. But much of what we know about the evolution of cognition suggests that this is unlikely to have happened. Evolution works by tinkering, grafting new structures onto already existing ones, changing the function of structures that are already there. All the evidence is that our cognitive architecture is a palimpsest of superimposed structures of varying phylogenetic pedigree.

Moreover, the points about computational tractability made earlier in this section suggest good reasons for thinking that these primitive structures have not only persisted but in fact continue to play an important role in our social lives. It is not, I have suggested, feasible to think that all or even most of our social interactions can be modelled in folk psychological terms. It is natural to think, therefore, that much of our current social cognition may well reflect a residue of skills and abilities that long preceded the emergence of metarepresentation and folk psychology. There is little to be gained, however, from pursuing this line of thought without providing concrete examples of the form that these non-folk-psychological skills and abilities might take. We will turn to that task in the next section.

IV

In thinking about how social understanding might be modelled in non-folk psychological terms a good place to start is with emotion perception. The form and level of one's participation in social interactions is frequently a function of one's assessment of the emotional states of other

participants. This is most clear when the interaction is a competitive one - a zero-sum game, for example (taking a game in the technical sense as a strategic interaction among players). It may be to my advantage, for example, to press ahead to take advantage of another participant's dilatoriness - or to retreat and retrench when I notice the aggressiveness of one of the competitors. But something similar holds for cooperative interactions. My own commitment to a shared project is likely to be at least partly determined by my sense of the extent to which my partners value the shared goal. And the exact form of my participation in the shared activity will be tailored to how I read my partners' varying and changing levels of enthusiasm. I need to be sensitive to whether and when my partners are bullish, bearish, frustrated or enthusiastic. Without this we will not be able to work together effectively.

There are three points to make about the various types of emotion perception implicated in social interactions. The first is the most obvious. The type of emotional sensitivity at play in social interactions is highly diachronic. Social interactions are extended processes in which the relevant affective valences are constantly changing. Successfully negotiating such interactions is not in any sense a matter of identifying relatively long-term dispositional states or character traits. The affective indices are in constant flux. Think, for example, of the emotional dynamics of a team game or a committee meeting. This makes folk psychological attributions, whether derived through a process of simulation or through the application of a theory, particularly inapposite. The processes by which folk psychological attributions are reached are too unwieldy to permit of rapid real-time monitoring and revision. The second point is that in many social interactions the actual content of the relevant affective and emotional states will be apparent from the context. Participants need to be sensitive not to what is represented but rather to the fine details of the attitudes taken to what can be presumed to be shared representational contents. The third point is the least obvious, but perhaps the most important. What matters in social interactions and coordinated activities is that the participants succeed in acting with due sensitivity to the affective and emotional states of other participants. There is no need for those affective and emotional states

to be explicitly identified and attributed. These forms of social understanding do not require forming judgements about the emotional states of the other participants.

It has been known for some time that emotion perception is highly dependent upon cues operating far below the threshold of conscious awareness. Emotional states can be transmitted directly from person to person, something which plays an important role in many types of social interaction, particularly those involving collective behaviour. We have a reasonably worked out understanding of how this transmission of emotional states can take place. The role of facial expression in the communication and detection of emotion has been systematically studied since Charles Darwin's pioneering study (Darwin 1872). Recent neuroscientific research, based on the study of brain-damaged patients and on lesion studies in animals has postulated the existence of neural circuits dedicated to the production and understanding of expressive behaviour, the so-called limbic system.

It is important, however, to separate out two possible claims that might be made. The simple claim that emotion perception is frequently subliminal and a matter of pattern recognition should be uncontroversial, and in itself does not count against what I have been terming the broad conception of folk psychology. Directly perceived emotional states can easily serve as inputs to the processes of simulation, or as the raw material to which the generalizations of theoretical folk psychology are applied. The more interesting, and potentially controversial, suggestion is that we frequently act upon the perception of emotional and affective states without explicitly identifying them. The idea here is that we regulate our own behaviour as a function of our sensitivity to the emotional and affective states of those with whom we are interacting without at any point making explicit the identifications on which our behaviour rests. The understanding of emotional expression feeds directly into behaviour. Sensitivity to emotional states feeds directly into action without any attribution of emotional states.

This second suggestion certainly is incompatible with the broad construal of the domain of folk psychology, for the essence of the broad construal is that social understanding requires categorising the behaviour of others in the concepts of folk psychology, in order to bring to bear either the mechanisms of simulation or the appropriate tacitly known theory. It seems to me that this second suggestion is a better characterisation of what is going on in very many cases of social interaction and coordination. The case for the broad construal of folk psychology is correspondingly weakened.

It might be wondered, however, just how significant this conclusion actually is. The types of social understanding that we have been considering are all highly circumscribed in at least the following sense. The issue is often not what other participants will do but how they will do it, since we may well know that other participants are constrained to act within narrowly prescribed limits. These are often not situations in which issues of explanation and prediction arise in the sort of ways for which one might think that folk psychological forms of social understanding would be required. Moreover, the fact that many social interactions involve an element of "affect attunement" which is achieved without recourse to folk psychology hardly shows that no element of those interactions is controlled folk psychologically. Even somone sympathetic to the general line that many basic forms of social interaction fall outside the domain of folk psychology might pose the question of whether the type of deflationary account I have been offering really counts against the broad construal of the domain of folk psychology.

There are two different issues to be distinguished here. The first has to do with the nature of those interpersonal situations that are *not* circumscribed by shared goals or a relatively small number of clearly defined possible outcomes. We might ask whether these are always situations in which we can only act effectively by actively explaining and/or predicting the behaviour of other participants in terms of what they believe and desire. I shall suggest that in fact there are many situations in which one can act effectively without having any beliefs about the propositional attitudes of others. Nonetheless, it seems clear that there will always be situations in which we do

need to predict and explain the behaviour of other people. The second question to be addressed, therefore, is how we generally proceed in those situations. Do we exploit the mechanisms of folk psychology, whether construed in simulationist or in theoretical terms? Or can we explain and predict what is going to happen without exploiting those mechanisms?

Let us start with the question of what happens in interpersonal situations that are *not* circumscribed by shared goals or a relatively small number of clearly defined possible outcomes. Those favouring the broad construal of the domain of folk psychology will suggest that, as soon as we move beyond highly circumscribed collaborative enterprises such as games or mending an aeroplane, we enter a realm of interpersonal interaction that can only be successfully negotiated by fitting the behaviour of other participants into the conceptual framework of folk psychology. In fact, however, it is far from clear that this is the case. The well-studied game-theoretical problem of how to behave in an indefinitely iterated prisoner's dilemma is a case in point, and one that has plausibly been argued to have wide application.

A prisoner's dilemma is any strategic interaction in which the dominant strategy for each player leads inevitably to an outcome in which each player is worse off than he could otherwise have been. A dominant strategy is one that is more advantageous than the other possible strategies, irrespective of what the other players do. In the standard example from which the problem derives its name, the two players are prisoners being separately interrogated by a police chief who is convinced of their guilt, but as yet lacks conclusive evidence. He proposes to each of them that they implicate the other, and explains the possible consequences. If each prisoner implicates the other then they will both end up with a sentence of 5 years in prison. If neither implicates the other, then they will each be convicted of a lesser offence and both end up with a sentence of 2 years in prison. If either prisoner implicates the other without being implicated himself, however, then he will go free while the other receives ten years in prison. The dominant strategy for each player is to implicate the other. Since we are dealing with rational players who know each other to

be rational, it follows that each will implicate the other, resulting in both spending 5 years in prison - even though had they both kept quiet they would have ended up with just 2 years apiece.

Although some authors have tried to argue otherwise (e.g. Gauthier 1986), it is hard to see how it can be anything but rational to follow the dominant strategy in a *one-off* strategic interaction obeying the logic of the prisoner's dilemma. But what about social interactions that have the same logic but are repeated? Repeated interactions create the possibility of one player rewarding another for not having implicated him (or whatever the relevant non-cooperative activity might be). Surely this will change what it is rational to do. In fact, however, it only does so in a limited range of situations. The well-known backwards induction argument suggests that the rational course of action where each player is rational, knows the other player to be rational and is certain in advance how many strategic interactions there will be will be to defect on the first play. It is only when it is not known how many plays there will be and/or the rationality of the other participant is not known that scope opens up for cooperative play. And this is where we rejoin the question of the domain of folk psychology. Suppose that we find ourselves, as we frequently do, in social situations that have the structure of an indefinitely repeated prisoner's dilemma. The issue may simply be how hard one pulls one's weight in the department.[6] It will be to my advantage to cut the examination meeting, provided that my colleagues do my work for me. But how will that affect their behaviour when we next need to wine and dine a visiting speaker? Will I find myself dining tete-a-tete and footing the bill on my own? Before I decide whether or not to cut the examination meeting I had better think about that possibility, and all the other possibilities when some or all of us applying dominance reasoning will lead to a sub-optimal outcome. But how do I do this?

It is natural to think that I will need to make a complex set of predictions about what my colleagues will do, based on my assessment of their preference orderings and their beliefs about the probability of each of us defecting as opposed to cooperating, and then factor in my own

[6] This is not, strictly speaking, a prisoner's dilemma, since it involves more than two players. The multi-person equivalent of the prisoner's dilemma is usually known as the tragedy of the commons.

beliefs about how what will happen in future depends upon whether or not I come to the examination meeting - and so on. This, of course, would be an application of the general explanatory framework of folk psychology (again on the simplification that utilities and probability assignments are regimentations of desires and beliefs). The broad construal of the domain of folk psychology is committed to saying that this is the way decision-making proceeds in strategic situations of these kinds - simply because these strategic situations depend upon social understanding and, according to the broad construal, all social understanding more complex than simple sensitivity to the emotional and effective states of others has to be a matter of folk psychological explanation and prediction.

Even if we can make sense of the idea that strategic interaction involves these kinds of complicated multi-layered predictions involving expectations about the expectations that other people are expected to have, one might wonder whether there is a simpler way of determining how to behave in that sort of situation. And in fact game theorists have directed considerable attention at the idea that social interactions taking the form of indefinitely repeated prisoner's dilemmas might best be modelled through simple heuristic strategies in which, to put it crudely, one bases one's plays not on how one expects others to behave but rather on how they have behaved in the past. The best known of these heuristic strategies is TIT-FOR-TAT, which is composed of the following two rules:

A. Always cooperate in the first round

B. In any subsequent round do what your opponent did in the previous round

The TIT-FOR-TAT strategy is very simple to apply, and does not involve any complicated folk psychological attributions or explanations/predictions. All that is required is an understanding of the two basic options available to each player, and an ability to recognise which one of those strategies has been applied by other players in a given case. The very simplicity of the strategy explains why theorists have found it such a potentially powerful explanatory tool in explaining

such phenomena as the evolutionary emergence of altruistic behaviour (see Axelrod 1984 for an accessible introduction and Maynard Smith 1982 and Skryms 1996 for more detailed discussion).

For our purposes, the lesson to be learnt from this example is not that we should model extended social interactions in terms of TIT-FOR-TAT. TIT-FOR-TAT has only a limited applicability to practical decision-making, simply because, in a situation in which two players are each playing TIT-FOR-TAT, a single defection will rule out the possibility of any further cooperation. This is clearly undesirable, particularly given the possibility in any moderately complicated social interaction that what appears to be a defection is not really a defection (suppose, for example, that my colleague misses the examination meeting because her car broke down). So any plausible version of the TIT-FOR-TAT strategy will have to build in some mechanisms for following apparent defections with cooperation, in order both to identify where external factors have influenced the situation and to allow players the possibility of building bridges back towards cooperation even after genuine defection.

The important point is that strategies such as TIT-FOR-TAT do not involve any exploitation of the categories of folk psychology. They can be followed without the attribution of folk psychological states to those with whom one is interacting. In fact, a stronger conclusion is warranted. Such strategies do not involve any processes of explanation or prediction at all. It is clear that no prediction is required, given that what I do in any particular situation is determined by how I interpret what the other player did in the previous encounter. It may seem that this introduction of the notion of interpretation allows folk psychological notions of explanation can get a grip, but this would be a mistake. In order to apply TIT-FOR-TAT, or some descendant thereof, all I need to do is to work out whether the behaviour of another player should best be characterised as a cooperation or a defection - and indeed to work out which previous behaviours are relevant to the ongoing situation. This will often be achievable without going into the details of why that player behaved as they did. Of course, sometimes it will be necessary to explore issues of motivation before an action can be characterised as a defection or a cooperation - and

sometimes it will be very important to do this, given that identifying an action as a defection is no light matter. But much of the time one might well get by perfectly well without going deeply at all into why another agent behaved as they did.

From a game-theoretical point of view, therefore, there is nothing mysterious about the idea that one can act effectively in complicated social interactions without bringing to bear the explanatory and predictive apparatus of folk psychology. Within game theory, construed as a normative theory of rational behaviour, it can make perfectly good sense to adopt strategies which are, in an important sense, folk psychologically "blind". The real question is with the extent to which the normative theory applies descriptively. How frequently do we employ heuristically simple strategies in social interactions, taking our cue from very simple understandings of what other people have done - rather than from complicated attributions of folk psychological states? The general considerations canvassed in the previous section seem to suggest that it's likely that we do. At the very least this brief excursion into game theory gives us a way of interpreting in non-folk-psychological terms a large class of social interactions that are *not* circumscribed by shared goals or a relatively small number of clearly defined possible outcomes.

Let us take stock. A case has been made for the following claims

(I) The form and level of one's participation in many social interactions is often a function of one's assessment of the emotional states of other participants in a way that feeds directly into action without any attribution of emotional states. This frequently occurs in social interactions circumscribed by shared goals or a relatively small number of clearly defined possible outcomes. Many such activities are controlled without anything that looks like a folk psychological attribution at all.

(II) We can participate effectively in social interactions which are not so circumscribed without making use of the predictive and explanatory apparatus of folk psychology.

Let us turn now to the second of the two questions identified earlier. Let us consider social interactions which do not fall under either (I) or (II). *Ex hypothesi* these social interactions require explaining and predicting the behaviour of others. Have we now arrived within the domain of folk psychology? As matters are generally understood by philosophers we must have arrived there,

simply because it is pleonastic that explanation and prediction proceed in folk psychological terms.

There is an important class of social interactions, however, in which it is true both that they involve predicting and/or explaining the actions of other participants and that the relevant predictions and explanations do not proceed via the attribution of folk psychological states. These are situations involving stereotypical routines and behaviour patterns. Let us start with two very simple examples. Whenever one goes into a shop or a restaurant, for example, it is obvious that the situation can only be effectively negotiated because one has certain beliefs about why people are doing what they are doing and about how they will continue to behave. I can't effectively order dinner without interpreting the behaviour of the person who approaches me with a pad in his hand, or buy some meat for dinner without interpreting the person standing behind the counter. But do I need to attribute folk psychological states to these people in order to interpret them? Must these beliefs about what people are doing involve second-order beliefs about their psychological states? Surely not. Ordering meals in restaurants and buying meat in butcher's shops are such routine situations that all one needs to do is to identify the person approaching the table as a waiter, or the person standing behind the counter as a butcher. That is all the interpretation required. These are both cases in which simply identifying social roles provides enough leverage on the situation to allow one to predict the behaviour of other participants and to understand why they are behaving as they are. There is no need to make any folk psychological attributions. There is no need to think about what the waiter might desire or the butcher believe – any more than they need to think about what I believe or desire. The point is not that the routine is cognitively transparent - that it is easy to work out what the other participants are thinking. Rather, it is that we don’t need to have any thoughts about what is going on in their minds at all. The social interaction takes care of itself once the social roles have been identified (and I've decided what I want to eat).

The basic lesson to be drawn from highly stereotypical social interactions such as these is that explanation and prediction *need not* require the attribution of folk psychological states. It would

be too strong even to say that identifying someone as a waiter is identifying him as someone with a typical set of desires and beliefs about how best to achieve those desires. Identifying someone as a waiter is not a matter of understanding them in folk psychological terms at all. It is understanding him as a person who typically behaves in certain ways within a network of social practices that typically unfold in certain ways. The point is that this is a case in which our understanding of individuals and their behaviour is parasitic on our understanding of the social practices in which their behaviour takes place. Nor, of course, is this understanding of social practices a matter of mastery of a primitive theory. We learn through experience that certain social cues are correlated with certain behaviour patterns on the part of others and certain expectations from those same individuals as to how we ourselves should behave. Sometimes we have these correlations pointed out to us explicitly -- more often we pick them up by monitoring the reactions of others when we fail to conform properly to the "script" for the situation.

One of the interesting characteristics of this type of social understanding is that it involves a type of reasoning clearly different from the way in which folk psychological reasoning is understood acccording to either the theory-theory or the simulation theory. For proponents of the theory-theory, social understanding involves what is essentially subsumptive reasoning. Folk psychology is a matter of subsuming patterns of behaviour under generalizations and deducing the relevant consequences. For proponents of the simulation theory, in contrast, folk psychological reasoning is a matter of running one's own decision-making processes off-line and feeding into them appropriate propositional attitude inputs for the person one is interpreting. For those types of social understanding which involve exploiting one's knowledge of social routines and stereotypes, however, the principal modes of reasoning are similarity-based and analogy-based. Social understanding becomes a matter of matching perceived social situations to prototypical social situations and working by analogy from partial similarities. We do not store general principles about how social situations work, but rather have a general template for particular types of

situation with parameters than can be adjusted to allow for differences in detail across the members of a particular social category.

Some researchers in computer science defeated by the practical difficulties of trying to provide rule- and logic-based models of common-sense reasoning – difficulties associated with the "frame problem" discussed earlier – have moved towards what are known as *frame-based* forms of knowledge representation. Here is Minsky's original articulation of the notion of a frame:

> Here is the essence of the theory: when one encounters a new situation (or makes a substantial change in one's view of the present problem) one selects from memory a structure called a *frame*. This is a remembered framework to be adapted to fit reality by changing details as necessary.
>
> A *frame* is a data structure for representing a stereotyped situation, like being in a certain kind of living room, or going to a child's birthday party. Attached to each frame are several kinds of information. Some of this information is about how to use the frame. Some is about what one can expect to happen next. Some is about what to do if those expectations are not confirmed.
>
> We can think of a frame as a network of nodes and relations. The top levels of a frame are fixed, and represent things that are always true about the supposed situation. The lower levels have many *terminals* – slots that must be filled by specific instances or data. Each terminal can specify conditions its assignments must meet. (The assignments themselves are usually smaller sub-frames.) Simple conditions are specified by *markers* that might require a terminal assignment to be a person, an object of sufficient value, or a pointer to a sub-frame of a certain type. More complex conditions can specify relations among the things assigned to several terminals. (Minsky 1974, pp. 111-112)

The frame-based approach is not, of course, confined to the representation of social situations and interpersonal configurations. Frames can have patterns of behaviour built into them. They provide a concrete example of the form that a routine-based approach to social understanding and social coordination might take.

Once again, we should separate out different possible claims here. Conceding that much of our social understanding may be frame-based rather than rule-based is not automatically to provide a further narrowing of the domain of folk psychology. It may be that the parameters in the frame that need to be set (what Minsky calls the terminals or slots) include specifications of the mental states of the other parties in the interaction. However, the claim I am putting forward, albeit

tentatively, is that this will not be the case (or at least will not be the case for many of our frame-based social interactions). The parameters associated with the other participants are set by specifications of roles and behaviour, rather than by specifications of beliefs and desires.

The frame-based approach has obvious applicability to scenarios such as that in the restaurant. But it is natural to ask how much of our everyday social interaction can be modelled in this way? How much of our social understanding is a function of our mastery of social roles, frames and routines? My inclination here would be to say: rather more than we think. It would be odd, given the element of repetition in all our social lives, if we had to start *ab initio* each time we participate in a repeated social interaction – if we operated with general principles which need to be tailored to meet the demands of specific situations, with all the difficulties of relevance that such tailoring involves. Again, it is hard to see what could count as a knock-down argument in this area, but the considerations of computational tractability canvassed in the previous section, together with the observation that many of our social interactions involve a considerable amount of high-speed, real-time adjustment to the behaviour of others suggests that the routine-based approach to social understanding should be taken very seriously.

Ex hypothesi, however, the understanding of social roles and routines associated with frame-based reasoning is restricted to familiar social situations. What happens when we find ourselves in unfamiliar social situations? What happens when none of our frames can be brought to bear; when we have no obvious contextual cues which will allow us to get a handle on the likely behaviour patterns of the other people with whom we are engaging; when the interaction is open-ended and the potential pay-offs and trade-offs too unclear for it to count as an instance of a prisoner's dilemma-type strategic interaction? It is natural to think that here we have arrived at the proper domain of folk psychology. One might think that social understanding is a complex tool for negotiating the social world. The social world is often transparent, easily comprehensible in terms of frames, social roles and social routines. Other agents can be predicted in terms of their participation in those routines and roles, while their emotional and affective states can simply be

read off from their facial expression and the "tenor" of their behaviour. When the social world is in this way "ready-to-hand", to borrow from Heidegger's characterisation of the practical understanding of tools, we have no use for the reflective apparatus of folk psychology. We do not need to bring to bear the machinery of folk psychological attribution to navigate through the social world, to accommodate ourselves to the needs and requirements of other people and to succeed in coordinated activities. But sometimes the social world becomes opaque. We find ourselves in social interactions where it is not obvious what is going on; that cannot easily be assimilated to prototypical social situations; where we cannot work out what to do simply on the basis of previous interactions with the other participants. And it is at this point that we find ourselves in need of the type of metarepresentational thinking characteristic of folk psychology – not as a mainstay of our social understanding, but rather as the last resort to which we turn when all the standard mechanisms of social understanding and interpersonal acommodation break down.

Early formulations of the notion of folk psychology stressed the idea that folk psychology is an explanatory theory. This is very much to the fore, for example, in Sellars's influential mythical account of how folk psychology might have emerged (Sellars 1956). For Sellars, the attribution of propositional attitudes is a process akin to the postulation of unobservables in science. We attribute folk psychological states as a way of trying to make sense of what is going on in the "black box" of other people's minds. Beliefs and desires are explanatory posits put forward to make sense of observable behaviour in a manner analogous to the way in which, say, subatomic particles are postulated to make sense of observable effects. Of course, on one way of taking it, this way of looking at folk psychology is closely aligned with the theory-theory. The theory-theory is a natural consequence of this way of understanding folk psychology if we stress the idea that the explanatory posits of folk psychology are given their content by the principles and laws in which they feature. But there is a broader issue here which applies equally to the simulationist understanding of folk psychology. It is a natural corollary of Sellars's understanding of folk psychology that we need only embark upon the process of folk psychological attribution when

other people are inscrutable - when it is not possible to read off what they will do and why they are doing it from the social context and from observable behaviour. But it is a philosophical myth that all our social interactions fall into this category.

References

Axelrod, R. 1984, *The Evolution of Cooperation* (Harmondsworth: Penguin)

Bermúdez, J. L. 2000, 'A Difference Without A Distinction', *Philosophical Explorations* 2, 63-82

Bermúdez, J. L. 2003, *Thinking Without Words* (New York: Oxford University Press)

Bermúdez, J. L. and Elton, M. 2000, *Personal and Subpersonal: Essays on Psychological Explanation*. Special issue of *Philosophical Explorations* (January 2000).

Carruthers, P. M. and Smith, P. K. 1997. *Theories of Theories of Mind* (Cambridge: Cambridge University Press)

Churchland, P. M. 1981, 'Eliminative Materialism and the Propositional Attitudes', *Journal of Philosophy* 78, 67-90.

Currie, G. 1995, 'Imagination and Simulation' in Davies and Stone (eds.) 1995b.

Damasio, A. 1994, *Descartes's Error: Emotion, Reason and the Human Brain* (New York: G. P. Putnam)

Darwin, C. 1872, *The Expression of the Emotions in Man and Animals* (London: Murray)

Davies, M. and Stone, T. 1995a, *Folk Psychology* (Oxford: Basil Blackwell)

Davies, M. and Stone , T. 1995b, *Mental Simulation* (Oxford: Basil Blackwell)

Dennett, D. 1969, *Content and Consciousness* (London: Routledge Kegan Paul)

Dennett, D. 1987, 'Cognitive Wheels: The Frame Problem of AI', in *Minds, Machines, and Evolution*, C. Hookway (ed.) (Cambridge: Cambridge University Press)

Dennett, D. 1996, *Kinds of Minds* (New York: Basic Books)

von Eckhardt, B. 1994, 'Folk Psychology and Scientific Psychology', in S. Guttenplan (ed.), *A Companion to the Philosophy of Mind* (Oxford: Blackwell)

Fodor, J. A., 1987, *Psychosemantics* (Cambridge MA: MIT Press)

Gauthier, D. 1986, *Morals by Agreement* (Oxford: Oxford University Press)

Goldman, A. 1995, 'In Defense of the Simulation Theory' in Davies and Stone 1995a.

Gopnik, A. and Meltzoff, A. 1997, *Thoughts, Theories and Things* (Cambridge MA: MIT Press)

Heal, J. 1996, 'Simulation, Theory and Content', in Carruthers and Smith 1996.

Lewis, D. 1972, 'Psychophysical and Theoretical Identifications', *Australasian Journal of Philosophy* 50, 249-258.

Maynard Smith 1982, . *Evolution and the Theory of Games* (Cambridge: Cambridge University Press)

Minsky, M. 1974/1997, 'A Framework for Representing Knowledge' in *Mind Design II,* J. Haugeland (ed.) (Cambridge MA: MIT Press)

Mithen, 1996, *The Prehistory of the Mind* (London: Thames and Hudson)

Rolls, E. T. 1999, *The Brain and Emotion* (Oxford: Oxford University Press)

Sellars, R. W. 1956/1997 *Empiricism and the Philosophy of Mind* (Cambridge MA: MIT Press)

Skryms, B. 1996, *The Evolution of the Social Contract* (Cambridge: Cambridge University Press)