Presumptuous Naturalism:
A Cautionary Tale

Professor Daniel D. Hutto
Professor of Philosophical Psychology

School of Humanities
University of Hertfordshire
de Havilland Campus
Hatfield,
Hertfordshire AL10 9AB

Telephone: +44 (0)1707 285655
Email: d.d.hutto@herts.ac.uk

Daniel D. Hutto

"The decisive movement in the conjuring

trick has been made,

and it was the very one we thought quite innocent"

1. Introduction

Naturalism is the word of the day. It's the 'ism' that most philosophers embrace (at least in English speaking climes). Wearers of the badge are a wildly diverse bunch. This is because there are quite different ways of being a naturalist and of conceiving of the naturalistic project.

　　Some naturalists take a special interest in our everyday or folk commitments. For them, the interesting philosophical project is to determine how much, if any, of what we ordinarily think about various subject matters (e.g. the mental, the moral, the aesthetic) is compatible with our best scientific understanding of what there is. To decide this, special methods have been created for (1) perspicuously representing our folk commitments and (2) examining if these outstrip, or go beyond, the commitments of a certain scientific understanding of what there is in nature. By these lights the philosophical task of the naturalist is to determine if the folk are committed to something over and above what is posited by a certain scientific world view. This naturalistic programme, known as the Canberra plan, relies on a special framework devised by its principal architect, David Lewis (1970, 1972) – a framework which has been extended by his followers, most prominently Frank Jackson (1998).[1]

Concentrating on their treatment of folk psychology, this paper seeks to establish that, in the form advocated by its leading proponents, the Canberra project is presumptuous in certain key respects. Crucially, it presumes (1) that our everyday practices entail the existence of implicit folk theories; (2) that naturalists ought to be interested primarily in what such theories say; and (3) that the core content of such theories is adequately characterized by establishing what everyone finds intuitively obvious about the topics in question. I argue these presumptions are a bad starting point for any naturalistic project and, more specifically, that in framing things in this way proponents of the Canberra plan have led us unnecessarily into philosophical quagmires.

The fundamental error is to suppose that our conceptual investigations ought to target (A) what the folk 'find obvious' about a given domain (which is putatively revelatory of a shared implicit theory) instead of (B) attending to what the folk do when competently deploying their concepts in dealing with that domain. Only the latter reveals the folk commitments. Focusing on what the folk find obvious, as Canberra planners claim to do, generates a host of methodological difficulties that are best avoided. Much worse than this, trying to identify what is 'intuitively known by all' typically results in contaminated pictures, of the genuine commitments of the folk, hogging our attention.

The cardinal sin of a presumptuous naturalism, as exemplified by the official versions of the Canberra plan, is that it makes it appear as if it is a simple matter to obtain an accurate understanding of folk commitments. Focusing on what anyone and everyone will find 'obvious' about some domain aids, abets and seemingly legitimizes certain popular but biased pictures of our folk commitments. This becomes dangerous when, by fuelling our intuitions, such pictures set important philosophical agendas and play a leading role in evaluating the adequacy of philosophical proposals.

In section three, using Jaegwon Kim as a stalking horse, I illustrate a clear instance of the kinds of difficulties that attend adopting this sort of starting point, focusing on the so-called 'problem' of mental causation. Examination of this case reveals that it is attachment to a certain popular understanding of our folk

psychological commitments – specifically, the idea that a productive notion of causation is required for making sense of human agency – that not only creates 'the problem' but also determines what any acceptable 'solution' *must* look like.

Against this, in section four, I show that a close examination of our everyday practice of deploying mental predicates to explain actions reveals no logically compelling grounds for supposing that the folk *must* be operating with a productive notion of causation. This is something inherited from elsewhere. Indeed, a quick review of how the philosophical community became convinced that reasons and reason explanations are causal in the first place (in *some* sense, in *some* cases) reveals that we would have no inclination to attribute a commitment to the notion of productive causation to the folk *if we did not already buy into* a certain picture of how mental states cause actions. Which picture? Well, the very one promoted by the standard Lewisian, commonsense functionalist analysis of the platitudes of folk psychology– an analysis that allegedly states *only* what the folk find obvious about the mental and nothing more.

If the diagnosis of this paper is correct then it serves as a cautionary tale. We ought to heed Wittgenstein's warnings about the dangers of being taking in by beguiling pictures – those that systematically obscure a clear vision of our actual use of concepts in various contexts.[2] That warning, it seems, is as pertinent today as ever. My aim is to show that any naturalism that takes a serious interest in understanding our folk commitments must adopt a different and more unassuming starting point than that proposed by the Canberra planners.

2. Naturalistic Ambitions

There is more than one way to conceive of the naturalistic project. Some forms of naturalism take no special interest in our everyday or folk conceptions. For example, this is true of Quine's scientistic naturalism. For him, the business of the naturalist is to provide adequate conceptions of what there is – conceptions that can economically account for the rich fabric of our sayings and doings (and hence adequate for all our practical scientific needs) in ways that are consistent with the

raw deliverances of sensory stimulation. Convinced by a series of arguments (most famously, one that allegedly dissolves the analytic/synthetic distinction), he saw the task of philosophy and science as the same: telling us not only what there is but also how it is possible for us to know what there is (in a way consistent with the first answer). Ultimately, the job is to make sense of the scientific worldview using nothing other than the best resources provided by the scientific worldview itself; "we seek no firmer basis for science than science itself; so we are free to use the very fruits of science in investigating its roots" (Quine 1995, 16).[3]

Quineians assume that we operate with theories in our everyday dealings with the world and that these theories are the inevitable starting points of inquiry.[4] It is conceded that these theories are never wholly abandoned (see Gibson 1996 for a useful discussion). Nevertheless, to satisfy Quineian ambitions, our lay conceptual schemes (our theories) require significant enrichment and extension to be of any interest to those engaged in the task of developing plausible stories about what there is. Folk conceptions are simply too ungainly, vague and unregimented to be serviceable for such work. For this reason Quineian naturalists take no special interest in what the folk say and think.

Other contemporary naturalists adopt a similar attitude to our folk conceptions despite having a more liberal, and less empiricist-based, project in mind. For example, Kim Sterelny (2003) offers a vision of the naturalistic project in which philosophy, although recognized as not one of the natural sciences, does indispensable service in enabling us to understand our place in nature. Working alongside the sciences, one of its major tasks is to integrate their findings into a consistent and coherent story – a story that would allow us to understand a wide range of interesting and important phenomena. On the list are such things as how "creatures manage to perform in a way that shows they are tracking the environment, planning ahead, responding to certain stimuli and ignoring others, grouping perceptual information together in particular ways, and so on" (Kingsbury and McKeown-Green 2009, 174). Sterelny-style naturalists are therefore interested in what the folk get up to as potential *explananda*. They might also be interested to discover if what the folk think about a specific subject matter happens to be true. But

this is not because they harbor any special, antecedent interest in folk conceptions in themselves. It is rather because such conceptions can provide initial inspirational fodder for the development of scientific theories in certain domains. Indeed, in some domains, such as cognitive science it is often the case that elements of folk thinking remain at the heart of certain advanced theories.

Matters are different with Canberra Planners. By way of direct contrast, their project depends on accurately characterizing our folk commitments. Their foundational first step begins with the effort to collect, collate and present our folk commitments in a logically perspicuous way. Ultimately, this is preparatory work for a second step; that of determining if – and to what extent – such commitments are satisfied by something in nature, whether perfectly, imperfectly or not at all.

In what follows, I concentrate exclusively on the presumptions associated with the first step of the Canberra plan. What does it involve? Focusing on Lewis' classic treatment of folk psychology (still the most influential instance of Canberra planning to date), we are famously told: "Collect all the platitudes … regarding the causal relations of mental states, sensory stimuli, and motor responses. … Add also all the platitudes to the effect that one mental state falls under another … Perhaps there are platitudes of other forms as well. Include only the platitudes which are common knowledge amongst us: everyone knows them, everyone knows that everyone else knows them, and so on" (Lewis 1972, 256).

Once the collection of platitudes is in hand it is possible, through technical manipulation, to make our folk psychological commitments explicit. This is done by collating all of the individual platitudes into a single, conjunctive sentence. In the first instance this sentence will consist of a string of T-terms, in this case terms standing for mental states (e.g. beliefs, desires, etc.), and O-terms, in this case terms standing for other things that are not part of the theory (e.g. sensory promptings, behavioural responses, etc.). This sentence articulates the 'postulate' that is the core of our folk psychology, revealing the distinctive roles played by the various mental state concepts and how they stand in relation to other things.

Once each occurrence of the mental states terms is replaced by free variables and the whole is prefixed by an existential quantifier, we get the Ramsey sentence for

folk psychology. This is important to stage two of the Canberra project since it is possible in principle to replace the free variables (standing for mental state terms) with terms in some other vocabulary standing for non-mental phenomena. Thus the stage one work of collecting and analyzing folk platitudes is a prelude to determining what, if anything, in the natural world (e.g. as identified by neuroscience or physics) might play the sorts of network roles of the mental states identified by the folk theory. Of course, it is possible that nothing in reality plays exactly such roles and that the states in question will have miscellaneous occupiers. If so, the theoretical commitments of the folk will not be satisfied in the way those of our best final theory are hoped to be – allegedly only that theory has a real chance of being perfectly and uniquely realized by the natural properties upon which all else supervenes.

In this way, Canberra planners assume that folk psychology is a term-introducing theory and that its core commitments are captured by the relevant platitudes. The platitudes are meant to express what the folk find obvious about the mental. There are a number of problems with these assumptions.

One major worry is how to make sense of the claim that folk psychology is a theory. Lewis (1994) says that folk psychology is a body of tacit knowledge rather like our knowledge of syntax. Jackson tells us, "As I use the phrase 'folk psychology' it stands for a certain theory about what the world is like … very many people believe this theory, which is why I think it is right to call it a folk theory. In the same way I think that many people hold the theory that, as a rule, unsupported bodies fall" (Jackson, Mason and Stich 2009, 48). Either way folk psychology is thought to name a contentful and structurally well-defined theory implicitly held by the folk.

A tempting interpretation of such remarks suggests that the Canberra planners are presupposing the existence of bodies of knowledge that explain our classificatory abilities – i.e. bodies of knowledge composed of representations with truth evaluable contents that are stored in our minds or brains. It is easy to find support for such a reading. For example, Jackson holds that what becomes visible through the process of platitude analysis is the pattern that the folk implicitly *recognize* in making their everyday classifications and judgments. Additionally, he

tells us that "it *is the pattern that drives the classifications* – not some set of sentences. Of course, there will be two patterns – the one in the sentences and *the one in the brain*. This follows from the fact that the classifications go via the brain" (Jackson, Mason and Stich 2009, 61, emphasis added).

Elsewhere he writes of the project:

> It is all to do with the availability to the subject of sentences that capture what they[the folk] believe – of, that is, sentences that *represent as their minds do* when they believe that P, where P is the theory we are talking about (Jackson, Mason and Stich 2009, 59, emphasis added).

On the face of it, it looks as if taking folk psychology to be an implicit theory entails a rather substantive view about the nature of minds – i.e. that they contain truth evaluable representational contents that are capable of driving thought and actions. In particular, it appears that the having of certain beliefs (understood as contentful mental representations) *explains* why the folk endorse the various platitudes about the mental that they do. This would, of course, account for the proposed link between the platitudes and the content of folk's implicit theory of mind.

But this presumes too much, given the stated ambitions of Canberra planners. The representational theory of mind is a substantial (and much contested) theory in cognitive science. Indeed, many argue that the primary defect of such a theory is that it models mental representations too closely on linguistic representations, thus ruinously impeding a proper understanding of the actual basis of cognition.[5] The issue is hotly debated. Although we cannot know the outcome in advance, we do know that it depends on sustained, philosophical argument and scientific investigation. Thus it cannot be the case that the existence of a representational theory of mind is secured in advance (and can be known to be true) because the folk operate with implicit theories. The natural reading of the idea that folk psychology *just is* an implicit theory seems to entail, in *its very setup*, that eliminativism is ruled out (cf. Jackson, Mason and Stich 2009, 63). That begs important questions in a very

blatant way - questions that the Canberra planners fully intend to leave open. Hence, the tempting reading must be rejected.

A lighter reading is possible. Like Quine, many contemporary analytic philosophers take for granted that the making of *any* kind of assumption, judgment or classification constitutes operating with or having a theory. This fits with Jackson's understanding that "to have a theory is to have a certain view about how things are" (Jackson, Mason and Stich 2009, 87). Indeed it seems that this is what allegedly makes the truth of theory-theory 'near enough analytic' (Jackson 1999, 80). The thought here seems to be that simply by engaging in the practice of explaining actions by appeal to mental predicates the folk reveal their implicit commitment to the existence of such things as beliefs and desires. That is certainly true and uncontroversial; the folk surely believe in beliefs and desires. Since the folk make use of these concepts in making sense of others it is safe to say that they believe in beliefs and desires. For Jackson, this equates to their holding a theory about how things are. But the same is true of folk beliefs about golf clubs, forks, lamp shades, and so on. So the folk must hold theories (in Jackson's special sense of the word) about anything which they deal.

But if this is all that is meant by 'folk psychology's being a theory' then we must be careful to distinguish this claim from a stronger one with which it is likely to be confused. For it hardly follows that in believing in beliefs (having a view of how things are) the folk are thereby 'positing' the existence of mental states (or any other ordinary things) as *theoretical* entities. No compelling reason has been offered to suppose that *this* would be an appropriate characterization of their activities. Of course, one might claim that such a characterization provides the best description of what the folk are doing. But that is a substantive *empirical* proposal – one that, again, requires evidence and argument to back it up.[6] It doesn't come for free. So, again, on most standard renderings it is presumptuous to assume that folk 'hold theories'.[7]

There are other problems with this idea. If one assumes that the true quarry of stage one platitude analysis is to get at an implicit theory held by the folk then one faces a problem of selection. The trouble is that the folk tend to have lots of thoughts and say lots of things about various topics. Canberra planners are clearly not

interested in most of what the folk have to say or what they think on these topics. Hence, "P is *not* everything that the folk believe about the mind" (Jackson, Mason and Stich 2009, 88). Very well, but then one faces the question: Which platitudes genuinely reveal the folk theory? As is well known, deciding this – i.e. deciding which platitudes matter and the relative weights different platitudes are to be given – remains "a vexed question" (Nolan 2009, 278).

Worse still, there are serious methodological problems about how to discover the platitudes in the first place. Many suppose that deciding what folk psychology 'says' is, in the end, an empirical matter (Stoljar 2009, 126). But the truth is, as anyone who has tried to do the empirical work soon discovers, that commitments of folk psychology (and certainly as they are represented by Canberra planners as a cluster of neatly interlaced principles) cannot be "easily extracted from the kind of things people say" (Ratcliffe 2007, 49).[8]

Beyond all of this, there is a much more crucial reason for rejecting the idea that we should focus on the content of implicit folk theories, as revealed by what the folk find obvious or intuitive. For even if all of the aforementioned problems could be overcome, there is always the risk that the intuitions in question will be shaped by distorting pictures that have nothing to do with the way our concepts are used in our everyday practices. To probe what the folk find obvious when they think about the mental is hardly to tap a philosophically uncontaminated source of understanding. There is a clear and present danger that prominent folk intuitions about a given subject matter will be informed, not by what is integral to folk practices themselves, but by, say, certain popular pictures of the nature of mind – even if only indirectly. Indeed, there is an exceedingly high risk that this will be the case when the folk intuitions are really those of philosophers acting as spokespersons for the folk, i.e. calling on their own sense of what is obvious.

Worryingly, appeal to revelation of this kind is precisely the Lewisian strategy. When asked why he thinks X must be part of the folk theory of Y, he answers: "Because so many philosophers find it so very obvious. I think it seems obvious because it is built into folk psychology. Others will think it is built into folk psychology because it is so obvious; but either way, the obviousness and the folk

psychological status go together" (Lewis 1999, 328). Of course, to make this story complete it would be handy if folk psychology existed as an implicit theory shared by all.[9] This would explain why it is that everyone finds certain platitudes about the mind – specifically those associated with the commonsense functionalist model of the mind – so very obvious. And, if the theory were true, these intuitions would be justified. But as argued above, Canberra planners are not entitled to assume the existence of such a theory as an explanatory construct.[10]

All of these problems dissipate if one abandons an interest in revealing what the folk 'find obvious' in favour of trying to discover what is integral to their competent use of concepts, when these concepts are put to good work in everyday practice. Taking the case in hand, the first step in making this shift is to stop thinking of folk psychology as an implicit theory. With far less baggage, we can assume that folk psychology denotes – at a bare minimum – the everyday business of making sense of intentional actions (i.e. our own and those of others) in terms of reasons. It requires being able to answer a particular sort of 'why'-question by competently deploying the idiom of mental predicates (beliefs, desires, hopes, fears, etc.). So conceived, folk psychology "is an account of how people actually understand each other's behaviour, rather than an account of how they think that they think in interpersonal scenarios. People may not have a clear idea about what is central to their thinking about others, and so [folk psychology] must be distinguished from what we might call 'folk folk psychology'" (Ratcliffe 2009, 381).

Folk psychology as a practice is a perfectly familiar, out-in-the-open, activity, not something hidden away in the minds of the folk. It is the sort of thing that Sterelny-style, integrative naturalists are interested in understanding. Indeed, any credible story they are likely to tell about it should harmoniously integrate the findings of a wide range of empirical sciences – e.g. anthropology, developmental psychology, comparative psychology, cognitive archaeology and neuroscience. If folk psychology is understood in this way then there is no point in trying to probe commonplace intuitions in order to reveal its core commitments. There is no need to try to discover the content of an implicit theory held by the folk, since we have no

reason for presuming that there is any such a thing (for detailed arguments on this score see Hutto 2008a).

At this juncture it is important to note that there are subtle discrepancies in standard accounts of what the Canberra planners are really interested in – discrepancies that reveal how easy it is to blur the distinction between what the folk do and what they think about a given subject matter. For example, in describing the stage one ambitions of the plan, Kingsbury and Mckeown-Green (2009) take it that the purpose of the first step is "to work out what, if anything, the ordinary people *are doing* when they ascribe and entertain beliefs and desires or *manage* to stay conscious" (174, emphases added). But note: working out what the folk are doing and working out what the folk think, as I have just been stressing, are really two quite different activities.

Arguably, if this is correct, any value in the analysis provided by commonsense functionalists derives from the fact that it manages to hit on *at least* some of what is integral to our folk practice.[11] It is possible to re-interpret some of its illuminating insights along the lines of what can be thought of as a 'practice' (as opposed to 'platitudes') analysis of folk psychology. Such an analysis would reveal the practical commitments of the folk – such as their commitments concerning the existence of beliefs, desires and other attitudes and of structural relations that hold between these attitudes, as revealed by the way that mentalistic concepts are used in prominent cases of making sense of actions in terms of reasons. Crucially, to engage in 'practice' analysis does not require assuming that the folk harbor implicit or tacit theories. It would be a mistake to imagine that it aims to detail the contents of some theory to which the folk quietly subscribe.

With this in mind, it may be possible to engage in something like the first stage work proposed by Canberra planners in a less presumptuous way. But there will be interesting consequences of doing so. Providing accurate descriptions of what the folk get up to when competently using their concepts requires painstaking investigation involving "philosophical work and attentiveness to the relevant phenomena" (Ratcliffe 2009, 385). Such work will not take the form of probing intuitions, for the reason given: Standard intuitions – even those of the allegedly

untutored folk – may be shaped by false pictures; pictures that can steal the limelight from and obscure the true character of folk commitments. Focusing on the standard assumption that our folk concepts are essentially causal, in what follows I illustrate the sorts of problems that arise from making false presumptions about the nature of folk commitments.

3. Mental Causation and the Competition Worry

Borrowing from Schopenhauer, Kim casts the problem of mental causation as an intractable *weltknoten* – a 'world knot' (Kim 2005, 7).[12] Like all good philosophical conundrums it gets its life from a clash between a certain set of intuitively plausible assumptions none of which we can surrender but which, when taken together, are apparently incompatible. Here the philosopher's job is to make the best fit possible, whatever the cost.

The way Kim sets up the problem can be usefully compared to the structure of a standard magic trick. First there is the set-up or 'the pledge'. At this point in the proceedings the magician reveals to the audience the subject matter to be dealt with and convinces them that his resources are perfectly ordinary and above board – i.e. he has nothing up his sleeve. Of course, this is also the point at which the audience is systematically duped by misdirection, and the success of the trick is guaranteed. In our case, attention is drawn to the fact that mental states are causally efficacious and hence mental concepts are causal concepts. Moreover, to underline this we are told that "a science that invokes mental phenomena in its explanations is presumptively committed to their causal efficacy; if a phenomenon is to have an explanatory role, its presence or absence must make a difference – a *causal* difference" (Kim 2005, 10).

Trouble and excitement immediately ensue in the form of the competition worry. For it looks now as if we have two different and sufficient causes, one identified by folk psychology and the other by some other science, both vying to do exactly the same explanatory work. On the assumption that over-determinacy is intolerable there is a problem. Assuming a functionalist framework, the root worry is that if mental properties are thought to be irreducible to but dependent on those of a more

basic sort then any causal contribution that any token mental properties might possibly make to the occurrence of any particular event will be systematically taken care of by its being realized by some or other subvenient base property. This is because of the relation of ontological dependence that is assumed to hold (vertically) between mental states and their realizers.

Thus exclusion is an unavoidable problem for anyone who assumes that mental states don't reduce to their realizers. The problem is utterly persistent. It arises as long as it is assumed that the realizers are heterogeneous (i.e. that mental properties are realized variously). It arises even if the supervenience base is assumed to be not just heterogeneous but also extended (e.g. perhaps mental properties depend not just on facts about X's brain, but facts about X's body, current environment, history, etc.). And it arises even if (as seems true) it is impossible that we will ever discover systematic laws couched either in concepts that apply to happenings at the subvenient level or in trans-ordinal psychophysical terms that would allow us to make interesting predictions or explanations concerning human reasons and actions, so described.

The competition worry arises just as long as it is assumed that there is an intimate metaphysical relationship of dependence – i.e. instantiation holding between mental properties and their realizers. Thus if we assume that mental properties have realizers then the latter apparently nullify, void or screen off any contribution a particular mental state property might make to the causal explanation of the occurrence of some other event.

At this point there is no way out. For this will be so no matter what practical purposes we might have for offering explanations that invoke mental predicates. Our purposes in giving such explanations matters not a jot when it comes to understanding the source of this metaphysical problem. This may not be obvious. Certainly the exclusion or competition worry *only* has teeth *if* we accept that reason explanations are causally explanatory and that reasons, or their associated mental states, are causes *in the same way as* explanations in the other sciences.[13] Consequently, it might be thought that the problem can be avoided if it turns out that reason explanations, in fact, provide explanations of a very different kind than

those of other, more basic sciences. Evidence for this is meant to come in the form that we offer such explanations for completely different purposes than we do when giving explanations in the sciences.

This is surely true. In many respects reason explanations are unlike causal explanations found in the sciences. Arguably, they function primarily as normalizing explanations (Hutto 2004, 2008a, Zawidzki 2008, Andrews 2009, Hacker 2009, Stoecker 2009). This is revealed by the fact that they are only called for when the other's actions are aberrant – i.e. when what is done violates our expectations and thus demands to be made intelligible. When this happens any of a wide range of explanatorily pertinent factors might require mention in order to make sense of X's reasons for acting (e.g. facets of X's character, X's mood, X's larger projects, the content of this or that propositional attitude of X, and so on).

Exactly what requires mention (and exactly how much elaboration is required) depends on the needs of the questioner. Crucially, like historical explanations, because reason explanations involve making sense of the other's idiosyncratic take on things they are not general and abstract but rather take the form of narratives that are personal and particular. In line with the maxim of conversational implicature, such narratives are selective; they cite details of a person's story – those that are worth mentioning in order to respond to specific, context-bound queries. Moreover, reason explanations are typically contrastive because in explaining why one Φed it is important to highlight certain factors over others that might have made a difference (e.g. this belief/desire pair as opposed to some other which could have equally explained why such an action was taken, etc).

But these observations are irrelevant when it comes to thinking about the problem of mental causation. All of this can be true (and indeed it all appears to be true) and yet reason explanations might be causal in the relevant sense after all. This will be so just in case, despite having all of the features just mentioned, what one cites when citing one's reason *also* picks out something that makes a difference – a *causal* difference. This is why Kim gives short shrift to the general strategy of trying to avoid the problem of mental causation by freeing "ourselves from our

15

metaphysical preoccupations *and attend[ing] to our explanatory practices involving mental phenomena*" (Kim 2000, 60).

With all of this in place we are brought to see that there is simply no escaping "the metaphysical problem … [which] is the problem of showing how mental causation is possible, not whether it is possible, although of course what happens with the how-question may in the end induce us to reconsider our stance on the whether-question" (Kim 2000, 61).

Now that we are well and truly trapped the magician's next step is to show us a way out. Once known in magical circles as 'the turn', this is where the trick seems to be performed. The resolution, according to Kim, is to accept reductionism. Really, it seems we have no other choice. Luckily, this is possible (he claims) – at least for intentional mental states – since they are amenable to functionalization. To functionalize is to show that a given mental state can be wholly defined in terms of the causal patterns of its characteristic inputs, network relations and outputs between relata such as perceptions, mental states and actions. Thus Kim's solution assumes that analytic functionalism (as described in section two) provides an appropriate starting point for analysis for at least some mental properties. If so, it follows that if his solution to the problem of mental causation is accepted then perhaps the best known brainchild of the Canberra plan will have been put to good use, yielding important results. Thus, "the problem of mental causation is solvable for a given class of mental properties if and only if these properties are functionally reducible with physical/biological properties as their realizers" (Kim 2005, 29 – see also 24).

And so, at this stage the trick is concluded. This is 'the Prestige', where all that is left for the magician to do is to accept the applause. Kim's reasoning is impeccable and his solution elegant. The set-up, the problem and solution are all very tidy - perhaps, too tidy. Canny readers will have noticed that invoking functionalization as a solution depends on antecedently accepting a familiar story about the nature of mental states. For "Functionalism views mental properties as *defined in terms of their causal roles* in behavioural and physical contexts, and it is evidently committed to the thesis that systems alike in intrinsic physical properties must be alike in their mental

or psychological character. The reason is simple: we expect identically constituted physical systems to be *causally indistinguishable* in all physical and behavioural contexts" (Kim 2005, 14, emphases added).[14]

Equally, as Kim stresses, thinking that there is a problem about mental causation doesn't depend on accepting any sophisticated philosophical premises (e.g. causal closure of the physical, etc). Rather it gets its life from acceptance of "a *perfectly intuitive* and ordinary understanding of the causal relation" (Kim 2000, 61, emphasis added); one which Kim more recently designates as our "strong intuitive conception of causation" (Kim 2007, 235). The troublesome, allegedly homely intuitive conception of causation that lies at the heart of this problem is that of productive or generative causation. Kim is admirably clear about this. He consistently reminds us that:

Causation as generation or effective production and determination, is in many ways stronger than mere counterfactual dependence, and it is *this sense* that is *fundamentally involved* in the problem of mental causation (Kim 2005, 18, emphases added).[15]

Fundamentally, these worries arise, I believe, from the question whether mentality has the power to bring about its effects in continuous processes of generation and production (Kim 2007, 236, emphasis added).

All will be well for Kim if our everyday activity of making sense of actions by appeal to mental predicates incorporates the assumption that mental states are causal in a productive sense. But what if this is a presumptuous mischaracterization of what is implied by the use of folk concepts in practice? What if this is an imposition – a kind of Lewisian overlay – on what is integral to folk practice? What if reason explanations need not be interpreted as causal in this sense? Well, then there would be no competition between reason explanations and those in other domains, no problem of mental causation, and no prestige for solving that problem after all.

4. Lessons from History and Alternative Possibilities

There was a time, not very long ago, when it wasn't obvious to philosophers that our everyday thought and talk about the mental was wedded to the idea that psychological phenomena cause and are caused by physical events (e.g. events in our bodies and environment). Indeed, it was widely supposed by those who gave detailed attention to the way our concepts operate in their everyday contexts that reasons were not causes and that reason explanations were not causal.

Famously, Davidson changed all of that. He paved the way for today's causalist orthodoxy, convincing us first that there are good reasons to think that reason explanations are causal in important respects and making respectable the idea that reasons are causes.[16] Crucial to his argument is the idea that citing *a* reason – understood as a particular belief/desire pair (which he calls a primary reason) – is not the same as citing *the* reason for an action. Usually a person has more than one primary reason for acting at any given time, yet in many circumstances only one of these will be *the* reason for the action. Relatedly, there are many alternative belief/desire pairs which might make a given action intelligible but it's the selective citing of one such pair as opposed to some other that explains, on any given occasion, why the person acted (Davidson 1987, 42).

Consider this scenario: I promise my wife that I will wash the dishes before five o'clock and being a man of my word I firmly intend to do so. According to Davidson this counts as a primary reason, a rationale, for my doing the washing up before the appointed time. However, it may be the case that, despite my intention to keep this promise, the reason I actually do the washing up before five o'clock is because, unexpectedly, I get a phone call from a friend who has decided to come round for tea at four. Since the tea cups are at the bottom of the pile, and I don't want my friend to see the kitchen in a state, I choose to do the washing up before he arrives. The latter reason (my desire for some clean cups and to keep a tidy house for my visitor) apparently pre-empts the former reason (my desire to be true to my word) in

bringing about, and hence causally explaining, this action. In other words, citing the latter reason explains (causally) why in this instance, I actually did the washing up.

In light of such cases, Davidson argues that we must read the 'because' in such common remarks as "I did the washing up *because* I wanted to impress my visitor" as denoting a causal relation. What distinguishes giving *a* possible reason for acting from giving *the* actual reason for acting is that in the latter case we are attempting to pick out the event which, as far as we can tell, in fact 'produced', 'brought about', 'was responsible for' or 'caused' the action in question. This is how the 'champion of the causal analysis' intends to show that "when we offer the fact of the desire and belief in explanation, we imply not only that the agent had the desire and belief, but that they were efficacious in producing the action" (Davidson 1980, 232). We are not interested in what *might have been* responsible for my Φing but what in fact was. We are interested in identifying the historical, actual cause of my Φing.

But many philosophers find Davidson's proposal unsatisfactory – the consensus is that it doesn't go far enough. This is because it is wedded to a story about event causation when what is apparently wanted is a story about property causation, as Honderich (1982) first showed. Imagine one event (call it c) – putting green, French pears on a scale – and another (call it e) – the hand on the scale moving to the two pound mark. No doubt c caused e. But we also know that not all of c's properties played a part in bringing about e. Event e was caused in virtue of event c's having certain instantiated properties and not others – i.e. it was the placing something *that weighed x amount, in these circumstances* that was responsible for the positioning of the scale's hands and not, say, the nationality or colour of the pears that made the difference.

By comparison and extension by analogy, many naturally suppose that taking Davidson's initial causal analysis seriously requires the conclusion that when citing mental predicates in explaining actions the folk must be (at least trying) to identify causally relevant properties. If not, it seems we would have to conclude that the folk are open to the possibility that mental properties, as a class, may be causally inert; that the mental *qua* mental is causally impotent. It is with this in mind that it is widely concluded that "Davidson's account *goes against our intuitions* about the

causal efficacy of the mental. It seems that only by recognizing that events cause other events by virtue of instantiating certain properties, can our intuitions about causal relevance be at all expressed" (Corbí and Prades 2000, 13).[17]

But surely this is too quick. Our folk practice tolerates other interpretations. Jackson and Pettit (1990, 1988) introduced the idea of programme explanations. They observe that: "We can and often do explain by citing a feature which causally programmes without causing. Features which causally explain need not cause. This is typically what happens when we explain in terms of highly relational properties" (Jackson and Pettit 1988, 392). The basic idea is some property F 'programs for' or 'ensures' the presence of some other property, P, which is causally efficacious with respect to G, although F lacks causal efficacy with respect to G.[18] So, to use their own example, 'fragility' is a dispositional property that is instantiated by a range of possible base properties, any one of which would cause (and one which does), e.g., glass breaking, in certain specifiable situations – by involving certain impacts. Of course, explanations that cite mental predicates are more informative than those that only cite dispositional properties; but plausibly explanations of both sorts might share the feature that they convey information that is not conveyed by explanations citing base properties that do the direct causal labour.

Emphatically, I am not seeking to defend or even place bets on the ultimate success of the programme explanation proposal. That is well beyond the scope of this paper. It will suffice for my purposes to use it as a foil to assess the standard reasons for rejecting the very possibility that it might tell an adequate story about our folk practice. Kim tells, for example, that: "explanatory relevance or efficacy … is too weak to be satisfying. To my mind any vindication of psychological explanation worth having must do justice to the 'because' in "She winced because she felt a sudden pain in the elbow" and to do this we need a more robust sense of because than is provided by programme explanation" (Kim 2000, 77). Apparently the trouble with that account is that: "the only relevance that is present here seems to be informational relevance" (Kim 2000, 75).

Allegedly, this is not good enough. It fails to supply what is really required. Thus, "If you are willing to give up on mental causation and a robust notion of

mental causal explanation and live with a looser and weaker model of explanatory relevance, you can perhaps make use of David Lewis' idea that 'to explain an event is to provide some information about its causal history'" (Kim 2000, 75). This won't do. Why not?

Kim assumes that the only kind of mental causation worth wanting is the productive sort, where mentality has the "power to bring about its effects in a continuous process of generation and production" (Kim 2007, 236). He insists on the need for a 'thick' conception of causation that involves 'real connectedness'. Thus attempts such as Jackson's and Pettit's (and others inspired by Lewis) are dismissed as too weak and too relaxed; they threaten to cheapen or dilute the idea that reason explanations are genuine causal explanations – i.e. that they are causal explanations at all.

> Explanation is a pretty loose and elastic notion … no one should legislate what counts and doesn't count as explanation, excepting only this namely when we speak of 'causal explanation', we should insist … that what is invoked as a cause *really* be a cause (Kim 2000, 76).

Real causation is apparently "productive causation, which respects the locality/contiguity condition" (Kim 2007, 236). We need this because what we are after is *real connectedness.* What is the source of these convictions about genuine 'robust mental causation' (Kim 2005, 22)? Why should we accept Kim's demands in this case?

As the possibility of an alternative explanation (by means of programme explanation) makes clear the issue cannot be settled by attending to our folk practices *alone*. Without additional support to secure it Kim's demand is problematic. For the fact is that there are stronger and weaker ways to make sense of the claim that when we give and receive reason explanations "it is only worth mentioning those properties of the event that are *relevant* to the causal processes at stake" (Corbí and Prades 2000, 13, emphasis mine).

It is an unwarranted imposition to insist that in highlighting 'what's worth mentioning' the folk are *necessarily* offering explanations that aim to directly pick out causal properties – properties that are productive of actions in a broadly mechanical sense. Such a conclusion is simply not justified by looking at the folk practice in isolation. It is entirely possible that what the folk find worth mentioning may be only, strictly speaking, informationally relevant and not causally productive. Nothing explanatorily pertinent would be lost if this were so. Put starkly, that the factors cited by the folk are worth mentioning does not entail that they pick out (or attempt to pick out) causally relevant properties *per se*. Hence, it does not follow that in talking of beliefs and desires the folk are referring to mental properties that denote causally efficacious mental types – properties that when appropriately instantiated productively and mechanically bring about actions in the way supposed by traditional analytic functionalists. To think otherwise is to simply assume a certain narrowly based, input/output version of commonsense functionalism that defines mental states as internal causes is a *bona fide* characterization of our folk commitments. But if this assumption is the true basis of Kim's demand then it is not independently justified, and his rejection of viable alternatives (such as programme explanation) is blatantly question-begging.

Indeed, the mere existence of a plausible rival interpretation shows that there is no logical link between what the folk find worth mentioning and the idea that mental states are causal 'in the productive sense'. Nevertheless, Kim is obdurate that: "metaphysics still won't go away. For the only way in which I believe that we can understand the idea of causal explanation presupposes the idea that the event invoked in a causal explanation is in reality a cause of the phenomenon to be explained" (Kim 2000, 61). He tells us that without this stronger, robust sense of mental causation in play our attempts to understand human agency and knowledge will be brought to ruin (Kim 2005, 9-10). Clearly, we should be unmoved by such fears (at least at this stage), since the claim that productive mental causation is required for an adequate understanding of human agency is the very point in dispute.

It might be thought that there is some general argument waiting in the wings that would establish that the *only* viable notion of causation is the productive sort. It might be supposed that such an argument would supply the requisite additional support. Of course, it wouldn't, unless it was already conceded that our folk practice requires a robust, productive notion of causation and not merely informational relevance. And, again, to assume that would be clearly question begging.

In any case, this is not the strategy Kim adopts. For other good reasons, he elects to take the more direct route. He is prepared to treat mentality as a special case, stating that "we need not concern ourselves with this general issue about causation" (Kim 2007, 236). Although he does (naughtily) assume that we need to invoke some notion of causation or other if we are to do justice to our folk practice, he attempts to establish positively that *only* the productive notion of causality will do. He states:

> It seems to me that counterfactual dependence is not enough to sustain the causal relation involved in our idea of *acting upon the natural course of events* and *bringing about changes so as to actualize what we desire and intend*. An agent is someone who, on account of her beliefs, desires, emotions, intentions and the like has the capacity to perform actions in the physical world: that is to *cause her limbs* and other bodily parts (e.g., the vocal cords) to move in appropriate ways (Kim 2007, 236, emphases added).

One it might be thought that Kim is claiming in this passage is that because we *experience* our agency as productively causal, this gives us compelling grounds for thinking that it is. To establish this would have the advantage of providing an independent basis for rejection other non-productive proposals about the nature of mental causation. But there are reasons to doubt the truth of that claim. First and foremost, *experiencing* our agency whenever we act (let alone experiencing it as productively causal) is not a regular feature of our phenomenology. Those who have extensively explored both the character of our sense of agency, and the ways that it makes itself manifest, argue that it takes the form of pre-reflective, implicit or tacit experience. It is a regular feature of acting only in the sense that if it went missing

one would suddenly notice it – one would find oneself doing things that would feel alien (see Gallagher 2007, and also Horgan and Tienson 2002).

If so, there is no secure route from facts about our standard phenomenology of agency to the idea that mental states productively cause actions. Worse still, even if experiences of productive agency do sometimes occur, noticeably so, when we 'act on the natural course of events' or 'bring about changes' in the world this is not universally the case. A sense of our *productively* bringing about changes is not always and everywhere a feature of our acting based on what we think, desire or intend. Of course, we are aware that changes occur but surely, we do not always *feel* that *we* bring about such changes when we act based on our beliefs and desires – the parade cases under examination. To recall my earlier example, although it may be true that my belief that 'a friend will visit at four o'clock' (and not my earlier promise to my wife) was decisive in my acting appropriately with respect to the dirty dishes it hardly follows that for this to be so I must have experienced myself as doing it for this reason and not the other.

If our commitment to mental causation as productive is not built into our experience of agency, then perhaps there is some other source for the idea behind Kim's conviction that "agency requires productive causation" (Kim 2007, 236). After all, in the cited passage Kim speaks, not of experiences but of 'our *idea* of acting on the natural course of events'. But if the thought is that this idea is simply built into 'our' or 'the' concept of agency then we are dangerously close, yet again, to making appeal to what everyone and anyone finds obvious. We are thrown back on intuition mongering and the idea that we all share an implicit folk theory. There is little doubt that many folk subscribe to the idea that mental phenomena are causal in a productive, mechanical way. It is a popular thought and may indeed be said to form part of the commonsense *picture* of the mind. Many riders of the Clapham omnibus would probably endorse it, if asked. But that surely does not matter. What matters is *why* they think this.

If the preceding observations are correct it is not because the folk have closely attended to what they do when using their folk concepts in practice. Rather it is much more likely that this commonplace view is, ultimately and quite indirectly,

sponsored by "a certain picture of the world; the mechanical/causal world picture. This picture sees the whole of nature as obeying certain general causal laws – the laws of physics, chemistry, biology, etc. – and it holds that psychology too has its laws, and that the mind fits into the causal order" (Crane 1995, 62). If acceptance of this sort of framework is the ultimate source of the commonsense idea that mental *must* be causal in a productive way then there is a clear problem of circularity in the standard presentation, ala Lewis and Kim, of this as a philosophically neutral characterization of what is built-into the very idea of mental agency.

Furthermore, if this analysis is correct many philosophers will need to abandon the pretence that they are in fact engaged in one of their favourite games – i.e. reviewing the prospects for the future of folk psychology. In recent years debates on this topic have been of focal interest. But it looks as if we've been sold a bill of goods in accepting as genuine certain familiar assumptions about the character of our folk commitments. It is generally assumed that the folk are committed to the idea that mental states cause actions in a mechanical way – i.e. that the folk are committed to the mechanical model of the mind. On this basis bets are placed on the likelihood of showing that other properties in the folk conception can be made compatible with the mechanical mind story. Usually this takes the form of trying to show that an account of mental representational content is compatible with a computational account of mental processing.

Fodor has done more than most to marry these two. Nevertheless, he realizes that to make a secure bond adjustments are required to what he takes to be the folk picture. Thus he tells us, "It is obscure how externalist intentional laws could be computationally implemented. Very well, then, let there be another kind of intentionality – let there be, as one says, 'narrow content' as well as 'broad' content – such that narrow content is ipso facto not externalist" (Fodor 1994, 17). The proposed revision is motivated by the assumption that the laws of folk psychology must be 'narrow': i.e. that they should only invoke intrinsic, local properties that are internal to agents. This is required if they are to be causal in the right sort of – productive, mechanical - way.

Others are far less sanguine about folk psychology, so understood, finding a future secured in a mature scientific psychology. They hold that the explanatory posits of a mature scientific psychology are unlikely to smoothly accommodate folk psychology. On this basis, its wholesale rejection or elimination is anticipated.

Taking it for granted that folk psychology incorporates a commitment to the idea that mental states are productively causal they doubt that its *other* commitments are compatible with this idea. Thus Ramsey remarks, "Folk psychology is committed to the existence of mental representations. Therefore, for folk psychology to be vindicated, the correct scientific theory needs to invoke, at the very least, inner cognitive representations *as well*" (2007, 114, emphasis added). He holds that this is unlikely, arguing persuasively that the best theories in cognitive science do not posit anything resembling the kinds of truth-evaluable representations that folk psychology trades in. With more punch, Churchland expresses the same concern in the following way: "neuroscience is unlikely to find 'sentences in the head'... on the strength of this assumption, I am willing to infer that folk psychology is false" (1991, 65). The bottom line is that for these thinkers folk psychology will be in good standing, if and only if, its commitments can be shown to be consistent with a mechanical model of the mind.

All parties to this dispute about the future of folk psychology subscribe to a fairly restrictive understanding of the requirements of a properly scientific psychology. But more than that, they are really only placing bets on the likelihood of incorporating a particular picture of folk psychology into (what they take to be) a proper scientific psychology. The picture of folk psychology in question is, as argued above, presumptuous. It presumes that folk practice is committed to the view that mental states are productively causal. Hence, these thinkers regard the folk as wedded to a particular understanding of the way that contentful mental states cause actions – i.e. *they assume* that the folk assume that there are such things as 'inner', representational mental states that 'mechanically produce' behaviour. I have been at pains to show that this is an *interpretation* of the commitments of practicing folk psychologists, but despite its seductive attractions it is not supported by careful attention to what the folk do (see Hutto 2004, 2008a). Of course, it is also

an interpretation that is unhealthily, and circularly, bound up with the standard supposition that folk psychology just *is* a kind of *low-level theory* – one that posits the existence of 'inner' mental states, such as beliefs and desires, as a means by which to causally explain outward behaviour and action.

Close scrutiny of what is truly implied by the commitments of our folk practices – i.e. as revealed by the way the folk use their concepts – shows that these familiar interpretations are without support. The bottom line is that determining whether it is coherent to suppose that mental properties, such as intentionality, are causally efficacious in a way that is compatible with a vision of the mind as mechanical is, despite appearances, simply not the same project as determining whether our folk understanding of mind, as revealed in our everyday practices, can be naturalized. To suggest otherwise is to risk a charge of false advertising.

6. Forget the Folk

An attractive option at this juncture, at least for philosophers with a particular agenda, will be to shout 'Forget the folk'. Perhaps, there is a problem with Canberra-style naturalism. Perhaps it is presumptuous about the nature of folk commitments in the very framing of its project. Perhaps it makes unwarranted assumptions about the nature and content of folk psychology. Surely, the cure is to abandon all interest in folk conceptions, full stop. After all, why should serious metaphysical musings about the mental be beholden to our ordinary use of concepts?

It seems that one might coherently reject such an interest and still be concerned with whether minds can be mechanical simply because one accepts, as an axiomatic general truth, that all of nature is mechanical. The universe might be imagined, along Newtonian lines, as an enormous clockwork device. Accordingly, if minds are part of the great natural machine then they *must* be mechanical – on the assumption that only properties that are mechanical, or reduce to mechanical, are natural. The job of the naturalist who endorses these assumptions is to determine if minds, given their special properties, really are natural, in the light of the above restriction. Even Quineians recognize that it is incumbent on naturalists to explain the sorts of belief-

forming and reason-giving activities that ground scientific investigations. Nothing guarantees in advance that the 'mechanical mind' story is best suited for, or even capable, of that task; indeed all the evidence of its incompatibility with truth evaluable accounts of representational content suggests just the opposite.

At first glance, pursuing this path seems to be a wholly coherent naturalistic project, at least if we measure coherence by popularity. It or something very like it looks, *de facto*, to have launched thousands upon thousands of philosophical speculations. Presented in this bold form, without qualification, it has the advantage of being revealingly unpretentious – revealingly, because any naturalist intent on motivating this sort of project will ultimately need to justify their starting assumption about the nature of nature. This is a tall task – and not one that is likely to succeed if based solely on scientific considerations. Ultimately, it would require establishing that the seventeenth century clockwork analogy is still viable in the light of twenty-first century thinking about the nature of reality – and, *prima facie*, that hardly looks credible. The risks are high for naturalists of this kind since they may be operating with mistaken assumptions about the best scientific conception of nature.

Still, there is a deeper problem. A major issue which divides naturalists concerns the referent of the term 'natural'. This is evident in current debates concerning the truth of physicalism. Popular formulations of physicalism state that 'all natural phenomena are physical phenomena' or that 'physical features and facts exhaust all the facts' (such that once the physical details of the world are in place and configured everything else follows automatically by strict entailment). Nevertheless, there are naturalists who coherently reject physicalism, arguing that 'natural' and 'physical' are not co-extensive terms. They hold that not all natural properties can be equated to or identified with physical properties. For them some natural – non-physical (e.g. phenomenal) – properties are primitive, existing in their own right alongside or in addition to physical properties.

It is clear that one of the crucial hinges in the debate about what falls under the concept 'natural' is the way disputants understand the characteristics of mental properties. It is the legitimacy of their characterizations of such properties that

determines if they are right to suppose that mental properties are irreducibly special. Yet, if deciding the fate of mental properties is unavoidable (which arguably, it is) and any understanding of mental properties requires making appeal to and interrogating our everyday or folk conceptions (which arguably, it does) then we are back to square one. The trouble is that it is unclear what else could possibly ground such conclusions if not an appeal to the way that the folk use their concepts. Without such an appeal what else could possibly decide how we ought to think about the nature of mental properties? What could possibly ground such debates?

In this light, it seems there is no option but for naturalists to take an interest in what the folk talk about when putting their concepts to good use. We ought to be interested in the possibility of making sense of things and how they stand to our best theories of what there is in nature. But if we are to investigate these questions properly we will need to adopt a more unassuming style of naturalism than has been popular of late.[19]

References

Andrews K. 2009. Understanding Norms without a Theory of Mind. *Inquiry* 52: 433-448.

Braddon-Mitchell D, Nola R. 2009. Introducing the Canberra Plan. In *Conceptual Analysis and Philosophical Naturalism*, ed. D Braddon-Mitchell, R Nola, 1-20. Cambridge, MA: MIT Press.

Churchland, P. M. 1991. Folk Psychology and the Explanation of Human Behaviour. *The Future of Folk Psychology*. J. D. Greenwood. Cambridge, Cambridge University Press.

Churchland, P. 2007. The Evolving Fortunes of Eliminativist Materialism. *Contemporary Debates in Philosophy of Mind*. B. McLaughlin and J. Cohen. Oxford, Blackwell Publishing: 160-182.

Corbí JE, Prades JL. 2000. *Minds: Causes, and Mechanisms*. Oxford: Blackwell

Crane, T. 1995. *The Mechanical Mind*. London: Penguin Books.

Davidson D. 1980. *Essays on Actions and Events*. Oxford: Clarendon Press.

Davidson D. 1987. Problems in the Explanation of Action. In *Metaphysics and Morality*. ed. P Pettit, R Sylvan, J Norman. Oxford: Blackwell.

Dilman I. 1996. Existence and Theory: Quine's Conception of Reality. In *Wittgenstein and Quine*, ed. RL Arrington, H-J Glock. London: Routledge.

Dretske F. 2009. What Must Actions Be for Reasons to Explain Them? In *New Essays on the Explanation of Action*, ed. C Sandis, 13-21. Basingstoke: Palgrave Macmillan.

Fodor JA. 1994. The Dogma that Didn't Bark (A Fragment of a Naturalized Epistemology). In *Naturalizing Epistemology 2nd Edition*, ed. H Kornblith, 192-216. Cambridge, MA: MIT Press.

Fodor JA. 1994. *The Elm and the Expert: Mentalese and its Semantics*. Cambridge, M.A: MIT Press.

Gallagher S. 2007. Sense of Agency and Higher-order Cognition: Levels of Explanation for Schizophrenia. *Cognitive Semiotics* 0: 32-48.

Gibson RF. 1996. Quine, Wittgenstein and Holism. In *Wittgenstein and Quine*, ed. RL Arrington, H-J Glock. London: Routledge.

Greenwood JD. 1991. Introduction: Folk Psychology and Scientific Psychology. In *The Future of Folk Psychology*, ed. JD Greenwood. Cambridge: Cambridge University Press.

Hacker PMS. 2009. Agential Reasons and the Explanation of Human Behaviour. In *New Essays on the Explanation of Action*, ed. C Sandis, 75-93. Basingstoke: Palgrave Macmillan.

Honderich T. 1982. The Argument for Anomalous Monism. *Analysis* 42: 59-64

Horgan T, Tienson J. 2002. The Intentionality of Phenomenology and the Phenomenology of Intentionality. In *Philosophy of Mind: Classical and Contemporary Readings*, ed. D Chalmers, 520-33. . New York: Oxford University Press

Hutto DD. 1999. A Cause for Concern: Reasons, Causes and Explanations. *Philosophy and Phenomenological Research*: 59(2) 381-401.

Hutto DD. 2004. The Limits of Spectatorial Folk Psychology. *Mind and Language* 19: 548-73.

Hutto DD. 2007. The Narrative Practice Hypothesis: Origins and Applications of Folk Psychology. In *Narrative and Understanding Persons*, ed. DD Hutto, 43-68. *Royal Institute of Philosophy Supplement*. 60. Cambridge: Cambridge University Press.

Hutto DD. 2008a. *Folk Psychological Narratives: The Sociocultural Basis of Understanding Reasons*. Cambridge, MA: MIT Press.

Hutto DD. 2008b. Limited Engagements and Narrative Extensions. *International Journal of Philosophical Studies*. 16: 419-44.

Hutto DD. 2008c. The Narrative Practice Hypothesis: Clarifications and Implications. *Philosophical Explorations* 11: 175-92.

Hutto DD. 2009a. ToM Rules, but it is not OK. In *Against Theory of Mind*, ed. A Costall, I Leudar, 221-38. Basingstoke: Palgrave.

Hutto DD. 2009b. Folk Psychology as Narrative Practice. *Journal of Consciousness Studies*. 16: 9-39.

Hutto DD. 2009c. Philosophical Clarification, Its Possibility and Point. *Philosophia*. 37: 629-652.

Jackson F. 1998. *From Metaphysics to Ethics*. Oxford: Oxford University Press.

Jackson F. 1999. All That Can Be at Issue in the Theory-Theory Simulation Debate. *Philosophical Papers* 28: 77-96.

Jackson F, Mason K, Stich S. 2009. *Folk Psychology and Tacit Theories: A Correspondence between Frank Jackson and Steve Stich and Kelby Mason*. In Conceptual Analysis and Philosophical Naturalism, ed. D Braddon-Mitchell, R Nola, 45-97. Cambridge, MA: MIT Press.

Jackson F, Pettit P. 1988. Functionalism and Broad Content. *Mind* 97: 381-400

Jackson F, Pettit P. 1990. Program Explanation: A General Perspective. *Analysis* 50: 107-17.

Kim J. 2000. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: MIT Press.

Kim J. 2005. *Physicalism, Or Something near Enough*. Princeton: Princeton University Press.

Kim J. 2007. Causation and Mental Causation. In *Contemporary Debates in Philosophy of Mind*, ed. B McLaughlin, J Cohen, 227-42. Oxford: Blackwell Publishing.

Kingsbury J, McKeown-Green J. 2009. Jackson's Armchair: The Only Chair in Town? In *Conceptual Analysis and Philosophical Naturalism*, ed. D Braddon-Mitchell, R Nola, 159-82. Cambridge, MA: MIT Press.

Lewis D. K. 1970. How to Define Theoretical Terms. *Journal of Philosophy* 67: 427-46.

Lewis D. K. 1972. Psychophysical and Theoretical Identifications. *Australasian Journal of Philosophy* 50: 249-58.

Lewis D. K. 1995. Reduction of Mind. In *A Companion to the Philosophy of Mind*, ed. S Guttenplan, 412-31. Oxford: Blackwell.

Lewis D. K. 1999. *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

McLaughlin B. P. 1993. On Davidson's Response to the Charge of Epiphenomenalism. In *Mental Causation*, ed. J Heil, AR Mele, 27-40. Oxford: Oxford University Press.

Menzies P, Price H. 2009. Is Semantics in the Plan? In *Conceptual Analysis and Philosophical Naturalism*, ed. D Braddon-Mitchell, R Nola, 159-82. Cambridge, MA: MIT Press.

Nolan D. 2009. Platitudes and Metaphysics. In *Conceptual Analysis and Philosophical Naturalism*, ed. D Braddon-Mitchell, R Nola, 267-300. Cambridge, MA: MIT Press.

O'Leary-Hawthorne J, Price H. 1996. How to Stand Up for Non-Cognitivists. *Australasian Journal of Philosophy* 74: 275-92.

Pigden C. 2009. A Niggle at Nagel: Causally Active Desires and the Explanation of Action. In *New Essays on the Explanation of Action*, ed. C. Sandis, 220-40. Basingstoke: Palgrave Macmillan.

Quine WV. 1995. *From Stimulus to Science*. Cambridge, MA: Harvard University
Press.

Ramsey WM. 2007. *Representation Reconsidered*. Cambridge: Cambridge University
Press.

Ramsey, W., S. Stich and J. Garon. 1991. Connectionism, Eliminativism and the
Future of Folk Psychology. *The Future of Folk Psychology*. J. Greenwood.
Cambridge: Cambridge University Press.

Ratcliffe M. 2007. *Rethinking Commonsense Psychology: A Critique of Folk Psychology,
Theory of Mind and Simulation*. Basingstoke: Palgrave Macmillan.

Ratcliffe M. 2008. Farwell to Folk Psychology: A Response to Hutto. *International
Journal of Philosophical Studies* 16: 445–51.

Ratcliffe M. 2009. There are No Folk Psychological Narratives. *Journal of
Consciousness Studies* 16: 379-406.

Sehon S. 2005. *Teleological Realism: Mind, Agency and Explanation*. Cambridge, MA:
MIT Press.

Slors M. 2009. The Narrative Practice Hypothesis and Externalist Theory Theory: For
Compatibility, Against Collapse. *Journal of Consciousness Studies* 16: 335-59

Sterelny K. 2003. *Thought in a Hostile World*. Oxford: Blackwell.

Stoecker R. 2009. Acting for Reasons - A Grass Root Approach. In *New Essays on the
Explanation of Action*, ed. C. Sandis, 276-92. Basingstoke: Palgrave Macmillan.

Stoljar D. 2009. The Argument from Revelation. In *Conceptual Analysis and
Philosophical Naturalism*, ed. D Braddon-Mitchell, R Nola, 113-37. Cambridge, MA:
MIT Press.

Wittgenstein L. 1953. *Philosophical Investigations*. Oxford: Basil Blackwell.

Zawidzki T. 2008. The Function of Folk Psychology: Mind Reading or Mind
Shaping? *Philosophical Explorations* 11: 193-210.

[1] It was so named by its critics O'Leary-Hawthorne and Price (1996) for the following reason: "Since Canberra is a planned city founded originally as the seat of the Federal Government of Australia, its detractors complain that it lacks the features that arise naturally in cities that grow organically and have diverse inhabitants that are not government bureaucrats. Originally an ironic suggestion, the metaphor turns on the allegation that the Canberra plan adopts a view of language that misses its fundamental diversity" (Braddon-Mitchell and Nola, 2009, 1).

[2] We were warned, some time ago, that: "when we have got a picture of our ordinary way of speaking we are tempted to say that our way of speaking does not describe the facts as they really are" (Wittgenstein PI § 402).

[3] Many nowadays are quite critical of this Quineian vision and the sort of instrumentalism about theories that it promotes. As Fodor quips: "Could the goal of scientific activity really just be to make up stories that accommodate one's sensory promptings? Would that be a way for a grown man to spend his time?" (Fodor 1994, 193).

[4] Everyday thought and talk is regarded as theoretical because it involves commitment to some or other conceptual scheme. For Quine, this is true of any intellectual activity that goes beyond mere responding to the deliverances of sensory stimulation or the making of rudimentary (i.e. non-referential, non-reified) observations. Hence, even the most basic classifications made in ordinary life about objects and their properties are deemed inferential, hypothetical and theoretical.

[5] Thus "In ascribing propositional attitudes to ourselves as our basic cognitive states, and explaining our behaviour in their terms, we are evidently trying to characterize a truly amazing cognitive machine – the brain – in terms of exactly one parochial game that only one species of animal has recently learned to play: language. From this perspective, why should we suspect, even for a moment, that human language would reflect the basic elements and structure of the brain's cognition?" (Churchland 2007, p. 180, see also Ramsey 2007).

[6] I believe there are compelling philosophical as well as empirical reasons for thinking that all explanatory versions of theory theory are false (see Hutto 2008a, ch. 8, 9 and 11, 2009a).

[7] It is a mistake to assume that believing in beliefs (as the folk do) necessarily constitutes having an implicit theory. Nor should we automatically suppose that learning how to make sense of actions in terms of reasons necessarily involves the acquisition of one. Simply put: "Learning to identify things is not coming to have theories" (Dilman 1996, 185). For further discussion of this point see Hutto (2009b, p. 33-34).

[8] Philosophers have been moved to try to take experimental soundings of folk thinking. For example, to ascertain whether or not desires are commonly conceived as causes Pigden (2009) surveyed two groups of university students (one from England, the other from New Zealand). The aim was to probe their conceptual intuitions. Questionnaires focusing on three vignettes were used. The participants were asked to provide 'commonsense obvious' answers, i.e. those they were inclined to give "without too much thinking" (p. 237). Based on what they said in response to these probes it was concluded that "the majority of students are inclined to think of desires as causes" (p. 239). Not only is the set-up of this sort of experiment questionable the results stand in need of serious interpretation. There is clearly more than one possible understanding of cause - which one are these folk operating with? And, why should we presume that the folk are best placed to answer that question?

---

[9] Jackson seems to propose exactly this. For him, the sort of conceptual analysis involved in stage one of the Canberra plan allegedly depends on making "appeal to what seems most obvious and central [about the concept in question], as revealed by our intuitions about possible cases" (Jackson 1998, 31). "My intuitions about possible cases reveal my theory … Likewise, your intuitions reveal your theory. To the extent our intuitions coincide with those of the folk, they reveal the folk theory" (Jackson 1998, 32).

[10] See Hutto (2008c) for a discussion of problems that arise for a plausibly weaker, more formal and wholly external rendering of the idea that folk psychology 'contains an implicit theory', one that could be made explicit.

[11] Consider the 'central action principle'. It states how the beliefs and desires must be related in the production of actions. To the extent that it reveals a core commitment of folk psychology it reveals a structure that is integral to the way the folk actually use their concepts when making sense of actions in terms of reasons. Although this structure is not easy to discern in isolated acts of reason ascription or understanding, it can be found allegedly by giving careful attention and analysis to the way the folk deploy mental predicates when doing so competently (or at least in producing or comprehending reason explanations of a certain familiar sort). See Slors (2009) for a discussion in favour of the idea that folk psychology competence embeds a belief-desire structure (p. 344-345). See Ratcliffe (2007, 2008, 2009) for arguments against this idea. Ratcliffe holds that although the folk make use of the concepts of belief and desire in their sense making practices, nothing like the belief-desire model, which assigns these concepts distinctive roles, plays a part in those practices. He and I disagree about this (see Hutto 2008b). We do agree, more fundamentally, that the question is a live one that cannot be settled by purely empirical means – i.e. if by that one thinks it would be settled by probing folk intuitions. Settling it requires substantial philosophical investigation into the way our concepts are used when they are being put to good work.

[12] To be precise he sometimes regards it as a cluster of inter-related problems stemming from various physicalist commitments.

[13] The trouble is that: "If both common-sense and physical science are offering causal explanations of human behaviour, it is hard to avoid seeing them as competitors" (Sehon 2005, 13).

[14] The attempt to naturalize folk psychology by functionalizing it looks promising precisely because the roles played by the propositional attitudes are assumed to be causal. Indeed, understood as an argument for physicalism (as Lewis first presented it), the emphasis on "causation plays an essential role … it is vital that the theoretical roles identified at the first stage of the Lewis program be causal roles" (Menzies and Price 2009, 189). The concern of this paper is not to assess the success or otherwise of Lewis' argument for physicalism. Rather it is to ask on what grounds was it decided that the roles played by mental states concepts are, in fact, causal?; And if causal, causal in what sense?

[15] Kim notes that in some ways the productive notion of causation is also weaker than counterfactualist renderings, hence it has an easier time in dealing with pre-emption and over-determination (Kim 2005, 18).

[16] This sea change has entered into philosophical lore: "Although philosophers once denied that thoughts could be causes (Anscombe 1957, Melden 1961), today there is agreement that in some sense, our ordinary understanding of thoughts attributes to them various causal roles" (Ramsey 2007, 18). Or again: "Davidson's (1963, 1967) seminal papers on the explanation of human action have led to a new consensus that folk-psychological explanations referencing 'propositional attitudes' are a species of causal explanations that may play an entirely legitimate role in a scientific psychology (whatever Davidson's reservations about its prospects)" (Greenwood 1991, 4). For a fuller discussion see Hutto (1999).

[17] Similarly, McLaughlin (1993) observes that since Davidson only offers us an event-based story about causation at best his version of identity theory only guards against token-epiphenomenalism and not type-epiphenomenalism. Hence he regards Davidson's anomalous monism solution as 'cold comfort'.

[18] This is reminiscent of Dretske's proposal to understand mental events as structuring causes. For the most up to date statement of this approach see Dretske 2009.

[19] I am grateful to Daniele Moyal-Sharrock, Joanne Gilles, Britt Harrison, Axel Seemann, Derek Strijbos, Marc Slors and the audience of the British Wittgenstein Society conference on 'Wittgenstein and Naturalism' for helpful comments on earlier versions of this paper and the ideas in it.