Abstract: What defence does the Narrative Practice Hypothesis (NPH) have against the charge that it is a covert form of externalist theory theory (TT)? I discuss and reject Dan Hutto’s own strategies and argue that the NPH remains vulnerable to a threat of collapse into externalist TT as long as narrative folk-psychological explanation is differentiated from simple belief-desire explanation merely by a degree of complexity, subtlety and/or context-sensitivity. It is entirely plausible, however, that there is a more principled distinction between these two types of explanation of human behaviour. I defend such a distinction and show how it eliminates the threat of collapse into TT entirely.

According to Dan Hutto’s narrative practice hypothesis (NPH) (Hutto, 2004; 2007; 2008a; 2008b) folk-psychology (FP) is fundamentally a narrative competence. This hypothesis is explicitly intended to replace the notion of FP as a folk-theory of mind (ToM). Given the variety of notions of ‘theory’ in this domain, though, it may not be entirely clear whether and to what extent the NPH is really incompatible with everyone’s understanding of FP as a theory. In this paper, I will argue that the NPH, as presented and defended by Hutto,

Correspondence: Prof. dr. M.V.P. Slors, Faculty of Philosophy, Radboud University Nijmegen, Erasmusplein 1, P.O. Box 9103, 6500 HD Nijmegen, The Netherlands. Email: marc.slors@phil.ru.nl

[1] Thanks to Leon de Bruin, Dan Hutto, Derek Strijbos and two anonymous referees for very useful comments.
faces a threat of collapse into some version of what is known as the ‘externalist’ variety of the so-called theory theory of folk-psychology (ETT). However, I will also argue that this threat can be avoided by accepting a principled distinction between simple belief-desire explanations of actions and full-blown narrative reason-explanations.

The paper is set-up as follows: in Section 1 I shall outline the NPH and its superficial resemblance to externalist TT. I shall also identify two arguments employed by Hutto against externalist TT, of which I will claim only one can serve as Hutto’s main defence against a possible collapse of the NPH into a form of TT: the rejection of mental holism. In section 2, I shall argue that this rejection of mental holism is incompatible with the ontological commitments of the NPH. Moreover, I shall argue that this rejection does not, as Hutto takes it, follow from the empirical facts. This renders the threat of collapse into TT all too real for the NPH. In Section 3, however, I will explain why the NPH does not collapse into TT. I shall argue that the narrative element in folk-psychological reason explanations is an addition to belief-desire psychology rather than an extension of it. While externalist TT may explain belief-desire psychology, this additional narrative element is out of its reach. Hence, the possible compatibility of externalist TT with the NPH does not amount to a collapse of the NPH into a form of TT. Moreover the claim of the NPH that TT does not fully capture real-life folk-psychological explanations remains intact.

1. The Narrative Practice Hypothesis and Externalist Theory Theory

Let me start by outlining the NPH, its possible compatibility with externalist TT and Hutto’s principled reason to reject this compatibility.

1.1 The NPH

The NPH is a theory about the nature and acquisition of folk-psychological (FP) competence that differs considerably from either of the only two options available up until recently: the theory theory (TT) and the simulation theory (ST). It has managed to carve out a theoretical niche for itself by subtly, and plausibly, changing the subject of the debate on FP. FP, according to Hutto, is not a spectator sport. It doesn’t serve the purpose of predicting or explaining behaviour from a disengaged third-personal standpoint. Rather, it serves the purpose of facilitating social interaction, which requires, most of the time, an engaged second-person perspective. Thus, what needs to be explained is primarily our ability to render actions intelligible in a socially acceptable...
fashion in order to facilitate our daily dealings with each other. This is done by giving reasons for actions. Reason-giving according to the NPH is storytelling, not theorizing.

The idea that reason-giving is theorizing is propounded by all versions of the TT. But also by the dominant version of ST (e.g. Goldman, 2006). For ‘if simulation plays a vital role in the process of understanding others, it does so by feeding the outcomes of operations involving recreative/enactment imagination or co-cognition into theorizing activity that brings ToM [Theory of Mind] principles into play’ (Hutto, 2008b, p. 177). When it comes to explaining our acquisition of FP competence, TT and dominant versions of ST commit us to innate or acquired ToM abilities; extreme cases of such a commitment are ToM modules (Leslie, 1992; 1994) or child-as-scientist hypothesis (Gopnik, 1996; Gopnik & Wellman, 1992).

Hutto forcefully argues against both (Hutto, 2008a, pp. 143–77). There is no need to buy into the strong cognitivism that infuses the traditional ToM views on FP acquisition. The basis from which we can explain the acquisition of FP competence, rather, is what Hutto labels ‘unprincipled bodily engagements’. These are our non-conceptual, embodied, primitive, but at times remarkably intelligent interactive capacities. Hutto explains how the use of language allows us to transform those engagements into the FP interactions of our daily lives (Hutto, 2008a, pp. 87–100; 129–42). In this explanation, typical folk-psychological concepts such as ‘belief’ and ‘desire’ are normally acquired in the course of our early development (see next section), but start to acquire their full use only when children develop the capacity to understand and create narratives. Hutto: ‘(…) children come into possession of all the pieces needed for playing the understanding-action-in-terms-of-reasons game before they can actually play it. What they are missing in their early years is not the necessary components, but knowledge of the basic rules’ (2008a, p. 27). But in fact, ‘rules’ still sounds too theoretical as the term designating the FP ‘principles’ that interrelate beliefs and desires. The point is that ‘[a]ccording to the NPH these “principles” are revealed to children not as a series of rules but by showing them in action, through narratives, in their normal context of operation’ (2008a, p. 29).

Thus — and that is one of its aims — the NPH explains the nature and acquisition of our ability to understand and explain actions in terms of beliefs, desires and reasons within an enactivist, embodied, non-representationalist view of ‘the mental’.
1.2 Externalist TT

In their (1994), Stich and Ravenscroft distinguish between what they call ‘internalist’ and ‘externalist’ conceptions of FP. Since I will be interested in the theory theory of FP only in this section, the relevant distinction to make is between internalist TT (ITT) and externalist TT (ETT). ITT is the prototypical sort of TT that the NPH argues against; it is the theory according to which our FP capacities are grounded in an innate or acquired theory of human psychology that is represented in our minds/brains. The NPH is incompatible with ITT. Those who suspect that the NPH may be compatible with TT must have a version of ETT in mind.

Stich and Ravenscroft introduce the externalist reading of FP and in effect ETT as follows:

We might equally well elect to use the term ‘folk psychology’ in a way that is more akin to Lewis’ usage — as a label for the collection of folk psychological ‘platitudes’ that people in our culture readily recognize and assent to. Or, since the collection of ‘platitudes’ is likely to be large and ungainly, we might reserve the label ‘folk psychology’ for a set of more abstract generalizations — a ‘theory’ if you will — that systematizes the platitudes in a perspicuous way and that (perhaps in conjunction with some other commonly known information) entails them. That systematization might well invoke terms and concepts that are quite unfamiliar to ordinary folk, in the same way that an attempt to systematize our linguistic intuitions probably would (Stich and Ravenscroft, 1994, p. 460).

According to ETT, FP-as-a-theory is present, not necessarily in our minds/brains, but in our FP practices. That is, our actual practice of understanding and explaining actions in terms of beliefs, desires and reasons, is supposed to be structured by a system of principles or rules. We need not be aware of this system, just like we are not aware of the rules of the grammar of the language we speak. But just like our actual speech and writing displays systematicity on closer inspection, so does our FP practice.

Is ETT ruled out by the NPH? It depends. Of course, the presence of FP-as-a-theory in our practices might be explained in terms of our possessing an internal theory of mind, that is, in terms of a theory represented in our brains (i.e. ETT and ITT are not necessarily mutually exclusive). Since the NPH is hostile to ITT, such a version of ETT will be incompatible with it. But to the extent that ETT does not invoke ITT, some version of it may well fit in with the enactivist antirepresentationalism of the NPH. Why would it not be the case that even though we acquire FP competences through exposure to
narratives, what we in fact learn through such exposure is how to use and apply the implicit rules of an (external) FP theory?

It is this possibility that I would like to investigate in this paper. Although I shall indicate in the third section why I think the NPH is not a covert version of ETT, I will first argue that Hutto’s own arguments against the compatibility of ETT with the NPH are not convincing. In order to do that, I will have to be more precise about which variety of ETT is most likely to be compatible with the NPH.

For even versions of ETT that shun an explanation of the alleged theoretical nature of our FP practices in terms of an internal FP-theory had by the practitioners might be incompatible with the NPH. ETT might construe FP such that ‘the folk’ conceive of beliefs and desires as theoretical postulates ‘similar to electrons, atoms, or gravity’ (Hutto, 2008a, p. 32). And such a view of the propositional attitudes is exactly what the NPH seeks to avoid and replace. To explain an action in terms of its reasons is not, according to the NPH, to hypothesize an explanans of that action in the form of private, internal beliefs and desires that are hidden from view. Rather, it is to situate that action in what Hutto calls a FP narrative. In FP explanations, according to Hutto, beliefs and desires operate as parts of narrative descriptions of actions; they are not theoretical postulates.

But there is a different class of theoretical postulates than electrons and atoms that Hutto seems to be overlooking. Consider Mendel’s postulate of genes at the time when he formulated his laws of genetics. Before the discovery that genes are in fact DNA (or are realized by DNA), genes are defined not as determinate entities, but rather as the causal roles played by some entity. As e.g. David Lewis (1966; 1972) noted, mental states such as beliefs and desires can be defined similarly, in terms of the causal role they play. ‘A mental state $M$ (say, an experience)’ according to Lewis ‘is definable as the occupant of a certain causal role $R$ — that is, as the state, of whatever sort, that is causally connected in specified ways to sensory stimuli, motor responses, and other mental states’ (Lewis, 1972, pp. 249–50). Accordingly, Lewis conceived of FP as a term-introducing theory that fixes the various causal roles mental state types are said to play relative to each other according to the users of FP. This is a form of ETT: Lewis says nothing about internal theories of mind (see also e.g. Jackson, 1998; Braddon-Mitchel & Jackson, 2007).

Some variant of this form of ETT may seem like a contender for being compatible with the NPH. But refinements are in order still. For, on the one hand it may be argued that electrons and atoms can also, on closer inspection, be defined in terms of their causal or theoretical
roles. On the other hand, genes turn out to be identical with DNA, a complex molecule, i.e. something with a status similar to atoms and electrons. In order to circumvent this problem and to keep the distinction between the types of theoretical postulate I wish to employ intact (in the present context this is the role-realiser distinction, but I will broaden the distinction later on), we may note that mental states as causal role states are multiply realisable: the same causal role state can be realised at different times or in different individuals by different neural structures. Mental states are thus not type-identical with neural states, and explanations of actions in terms of beliefs and desires cannot be reduced to explanations in terms of neural causes. Hence, mental states as causal role states have a status that is different from genes as a causal roles concept in that type-identification with some definite objective physical item is not feasible. Mental states as causal role states may be the postulates of a term-introducing theory, but that certainly doesn’t mean they are like electrons, atoms or gravity.

To be sure, the claim here is not that the NPH or Hutto’s position in general is compatible with Lewis’s views on FP or psychophysical identification in terms of a role-realiser relation. The claim is that a version of ETT that depicts FP as a term-introducing theory that defines mental states in terms of multiply realisable causal role states seems compatible with the NPH. Such a version of TT does not postulate an internal theory of mind and neither does it portray FP as a theory that postulates entities similar to electrons and atoms.

The crucial move here is to distinguish between theoretical postulates that are obviously incompatible with the enactivist anti-representationalism of the NPH and those that are not. The distinction between roles and realisers helps, in this respect, but other distinctions may help too. Daniel Dennett, for instance, doesn’t conceive of beliefs and desires as causal role states. But he does conceive them as theoretical postulates that are different from electrons, atoms and gravity, i.e. different from what Reichenbach called *illata*. Beliefs and desires, according to Dennett, are *abstracta* — theoretical postulates like centres of gravity or vectors — that capture and predict our behavioural regularities, perhaps not accurately to the last detail but certainly fast, practical and effective. On his view, there is no need to postulate any neural realisation as concretely isomorphic to them as occupiers of causal roles:

---

[2] Two remarks. First: I diverge from Lewis’s 1972 proposal here. Lewis is interested in psycho-physical identification and hence in unique realisation. Second, the idea of multiple realisation is familiar from Fodor (1974). But it is worth emphasising that it does not imply or favour Fodor’s own ITT (e.g. Fodor, 1975).
Jacques shoots his uncle dead in Trafalgar square and is apprehended on the spot by Sherlock; Tom reads about it in the Guardian and Boris learns of it in Pravda. Now Jacques, Sherlock, Tom and Boris have had remarkably different experiences (…) but there is one thing they share: they all believe that a Frenchman has committed murder in Trafalgar square. (…) Ordinary folk-psychologists have no difficulty imputing such useful but elusive commonalities to people. If they insist that in doing so, they are postulating a similarly structured object in each head, this is a gratuitous bit of misplaced concreteness (Dennett, 1987, pp. 54–5).

Folk-psychologists may be depicted as theorists, in Dennett’s scheme, but only in a particular sense. FP is not a theory that refers to concrete entities in order to explain and predict behaviour, regardless of what either philosophers or FP users think (Dennett, 1995). But the abstracta invoked by FP are, according to Dennett, structured into what may be described as a theory about ‘the performance specifications of believers’ (Dennett, 1987, p. 59). Here too, the claim is not that Dennett’s position is entirely compatible with the NPH. The claim is merely that ETT might be conceived roughly along Dennettian lines, and that such a version of TT appears prima facie compatible with the NPH.

The ways in which Dennett and Lewis set apart causal role states and abstracta from the kind of theoretical postulate that is incompatible with the NPH requires reference to a form of mental holism. It is through defining beliefs and desires in terms of their holistic interrelation with each other and with contextualised behaviour that it becomes possible to conceive of them as something other than the concrete theoretical postulates — the illata — that the NPH denies to be at play in FP. Thus, Dennett explicitly claims that mental states are defined holistically within FP while Lewis conceives of FP as a huge theoretical sentence in which mentalistic terms such as ‘belief’ or ‘desire’ can be replaced by variables that play specific causal roles as defined by the Ramsey sentence of FP as a whole. Mental holism is a non-accidental aspect of the kind of ETT that may, as I argued, appear compatible with the NPH. This is important for what is to follow.

1.3 Hutto’s arguments against ETT

Hutto’s attacks on TT focus primarily on ITT and there is little attention for ETT. As far as I can see Hutto offers two arguments against the kind of ETT that I claimed to be prima facie compatible with the NPH in the previous subsection. One partly implicit strategy, to be found throughout Hutto’s (2008a), is his claim that the mental holism
that is crucial for the kind of ETT that may be compatible with the NPH is in fact refuted by facts from developmental psychology. Another strategy is Hutto’s later argument (2008b) that the Lewis-style version of ETT that might appear compatible with the NPH cannot be considered a proper theory. I take this argument to be applicable as well to a Dennett-style ETT. In this subsection I shall outline both arguments. I will start with the second one and argue that it fails, leaving the first argument as Hutto’s main buffer against the claim that the NPH may turn out to be another form of TT.

The basic idea behind the second argument is that when ETT conceives of FP as a theory that is implicitly present in our daily practices (rather than in our heads), then ‘(…) “theory” is surely the wrong word since the second T in TT would not denote a “theory” in any interesting or substantive sense. (…) [It] would not be an articulation of the “principles” used by folk psychological practitioners in making their everyday attributions, predictions or explanations. [They] would not be (…) used by the folk in their daily routines any more than Newtonian laws are used by planets in order to conduct their business’ (2008b, p. 180).

I am not convinced that this is a good argument. Consider a parallel example: Suppose we reject a Chomskian view of natural language according to which all language users possess some internally (i.e. neurally) stored generative grammar; language is grammatical, to be sure, but the grammar is present in our linguistic practices, so to speak, not in our heads. We gain knowledge of our grammar, on such a view, by carefully observing and analysing the principles at play in our linguistic practices. But does this view imply that if we gain such knowledge, ‘it would not be an articulation of the “principles” used by [linguistic] practitioners’? There surely is a strong tendency to speak of principles used by people here. So if there is not a similar tendency to describe the planets as using Newtonian laws we must explain the difference. Hutto’s argument assumes that we can’t explain the difference. I think we can. I will stick to FP while explaining it.

Crucially, planets are objects and folk-psychologists are agents. That is, unlike planets, folk-psychologists initiate their own actions, in this case the acts of interpretation of each other’s behaviour. In that sense, planets are governed by Newtonian laws (at least for the sake of this argument) while folk-psychologists are not in a similar way governed by the principles of externalist FP. Folk-psychologists may, in Dennett’s terminology, choose to apply the intentional strategy (i.e. apply FP), but they may also opt for a design strategy or a physical strategy. Or they may apply a faulty or idiosyncratic version of FP.
Precisely because there are options and precisely because the folk-psychologist herself decides on one of these options while initiating the act of FP interpretation, it is natural to express this by saying that she *uses* the FP strategy (*rather than* some other strategy) to interpret someone’s actions.

The fact that the folk-psychologist is likely not to be able to state explicitly the principles of the strategy she is using does not block this conclusion. Many language users are unable to formulate the grammatical principles at play in their speech. To demand that folk-psychologists be able to state the principles of FP in order to count as using FP-as-a-theory would be to discount almost all forms of TT. For the huge majority of TT-ists claim that FP-as-a-theory is used implicitly.

Hutto’s later argument against ETT, then, fails to convince. What remains is the strategy applied against ETT in his 2008 book: the rejection of mental holism. Here Hutto reasons as follows: The ability of children to understand moderately complex narratives (of the kind Hutto claims are involved in the acquisition of FP competence) develops after children acquire the concepts of desires and beliefs. For this reason, Hutto emphasizes at different occasions that these narratives are *not* responsible for introducing an understanding of mental concepts, such as desire and belief for the first time, rather (...) they put on show how these attitudes can integrate with one another (...). The NPH assumes that kids already have a practical grasp on what it is to have a desire or belief before learning how to integrate their discrete understanding of these concepts in making sense of actions in terms of reasons. FP narratives enable this by showing how these core attitudes and other mental states behave *in situ* (Hutto, 2008b, p. 178).

Conveniently, this separation of (the grasp of) FP concepts (e.g. beliefs and desires) from (the grasp of) the ‘principles’ interrelating them in FP explanations of actions allows Hutto to distance himself from the kind of mental holism involved in the kind of ETT that I argued to be possibly compatible with the NPH. Hutto:

(...) the claim that the meaning of mental predicates depends wholly on their lawful relations is apparently undermined by the fact that children develop a practical understanding of the different propositional attitudes at distinct stages in their careers. Thus they have a grasp of the concept of desire, quite independently of and prior to having an understanding of the roles they play in making sense of a person’s reason for action. And an understanding of both of these attitudes appears to precede an understanding of the roles they play in making sense of a person’s action (Hutto, 2008a, p. 31).
This, then, is the principled ground on which Hutto is able to claim that the NPH is not a covert version of ETT.

2. Belief-Desire Interaction and the Threat of Collapse

So, in (my reconstruction of) Hutto’s view, in order to safeguard the NPH from collapsing into a form of TT, it is crucial to claim that ‘children come into the possession of all the pieces needed for playing the understanding-actions-in-terms-of-reasons game before they can actually play it’ (Hutto, 2008a, p. 27). In this section I will argue that this principled distinction between ‘the pieces’ and ‘the game’ or ‘the FP components’ (propositional attitudes and other psychological states) and the ‘principles’ interrelating them, cannot be upheld. I will argue that the distinction is hard to square with Hutto’s anti-internalistic, interpretationist view of the mind, and that the developmental psychological data Hutto invokes to argue for this distinction do not in fact show what he takes them to show. Finally, I will trace the NPH’s threat of collapse into TT back to what I take to be its root: the assumption that narrative reason explanation is continuous with simple belief-desire psychology on a scale of increasing complexity. Rejecting that assumption, I will argue, should be the strategy for avoiding a collapse.

2.1 Mental holism in view of Hutto’s interpretationism

One of the features (and according to many, including myself, attractions) of the NPH is that it offers a complete account of human action and interaction without drawing on a traditional notion of ‘mind’ as involving internal representations (such as in the mental realist positions that gave rise to ITT), or even a notion of mental states that calls for a theoretical account of tractable neural implementation (such as offered e.g. by Lewis). On the one hand, the unprincipled engagements from which our narratively guided FP practices grow involve intentional but not propositional attitudes. To apply the notion of mental content at this level would be to miss the point (Hutto, 2008a, ch. 3). On the other hand, the high-level reason talk of FP gets its explanation at the socio-cultural rather than the neuro-psychological level. At this level there are mental contents, but these cannot be traced back or reduced to the neural structures of the people to whom the relevant mental states are ascribed.

What, ontologically speaking, is it to ascribe a belief or a desire to someone on this picture? I think the answer here should be that the position Hutto advocates is some form of interpretationism, probably...
of a more Davidsonian than Dennettian kind. Just like Hutto argues that to have a concept of ‘belief’ is to have certain social abilities (2008a, p. 129–31), I take it that on his account to ascribe a belief is to interpret someone’s behaviour so as to allow or express the possibility of specific interactions. FP in Hutto’s conception is, as Dennett (1995) puts it, ‘a craft’, not a description or theory (it should be said, though, that Hutto’s interpretationism is not as permissive as Dennett’s; see 2008a, pp. 42–3).

The point I wish to make in this section is that interpretationism is usually taken to involve a degree of mental holism that appears to contradict the assumption that it is possible to possess the components of FP, i.e. be able to apply the concepts of ‘belief’ and ‘desire’, in abstraction from knowing how these components interact (cf. Malpas, 1992, ch. 3). This is best illustrated by cases in which the FP interpretation is underdetermined by the action at issue. An example may help to convey the idea. Suppose I am at a birthday party and I see my 11 year old nephew reach for a glass of beer that is on the table. Since I do not believe he desires beer, I will attribute to him the belief that there is, say, apple juice in the glass, because I know that’s what he likes. Then I realize that it is common these days for young children to boast about drinking alcohol. Maybe he is impressed by stories of classmates and doesn’t want to stay behind. In that case I should attribute to him the desire to drink beer and the belief that there is beer in the glass. What a simple example such as this shows is that the ascription of beliefs and desires are linked. I cannot, when altering my interpretation of my nephew, merely alter, say, the desire I ascribe to him and leave the belief intact; there’s no point in reaching for a glass of apple juice when you want a beer. Typically, it is the assumption of rationality that interpretationists refer to when explaining why we ascribe beliefs and desires in pairs, and not atomistically.

But are beliefs and desires always invoked in tandem? Take another example:

The boy’s sign says ‘LEMONADE — 12 cents a glass.’ I hand him a quarter, he gives me a glass of lemonade and then a dime and a penny change. He’s made a mistake. Now what can we expect from him when we point out his error to him? That he will exhibit surprise, blush, smite his forehead, apologize, and give me two cents. Why do we expect him to exhibit surprise? Because we attribute to him the belief that he’s given me the right change — he’ll be surprised to learn that he hasn’t (…). Why do we expect him to blush? Because we attribute to him the desire not to cheat (or be seen to cheat) his customers (Dennett, 1987, p. 84).
Here it seems that the surprise is explained by the boy’s belief that he has given the right change. The blushing on the other hand is explained by his desire not to cheat. At first glance there is no belief-desire interaction. And since this example is from an uncontaminated interpretationist source, it may seem to show that interpretationism does allow for the possibility to ascribe beliefs and desires without being aware of how these interrelate. If that were the case, Hutto can have his cake and eat it; he can stick to an interpretationist ontology of the propositional attitudes and accept (at least in some cases) the kind of mental atomism that is required by his separation of FP components from FP ‘principles’.

But on closer inspection it is not true that there is no belief-desire interaction in examples such as these. Why would the boy blush if he didn’t believe he might be seen as having cheated? Or look at the bracketed phrase in the last sentence of the quote. It matters whether the boy doesn’t want to cheat or whether he does want to cheat but doesn’t want to be seen cheating. In the former case the expression of surprise is explained by his belief to have given the right change. In the latter case, the expression is explained by his attempt to hide his real intentions and his wish to make it look like he believed he had given the right change whereas he didn’t in fact believe that. The point of this second example is that if one were to claim of a given situation that it can best be viewed as a situation in which someone understands the behaviour of someone else merely through the attribution of a single belief or desire, one needs to take great care to show that no implicit background beliefs or desires are overlooked.

The relation between holism and interpretationism is a strong one. If the point of attributing beliefs and desires is not to refer to an independent mental reality but instead to make sense of the behaviour of others, as Hutto has it, and if, as the examples above show, this requires ascription of specific sets of interrelated beliefs and desires then it is hard to imagine how one can be an interpretationist without accepting mental holism. Having said that, though, the purpose of this brief subsection is not to present a knock-down argument to the effect that Hutto’s ontological commitments force him to abandon the components-principles distinction. There are various forms of interpretationism and it does not seem a priori impossible that on some loose version of it some form of mental atomism is feasible. But if

---

atomism and interpretationism are held to be compatible, that compatibility certainly needs to be argued for.

2.2 Do empirical facts indicate a distinction between FP components and FP ‘principles’?

Hutto does argue, in some sense at least, for the fact that young children acquire the concepts of ‘belief’ and ‘desire’ as full blown propositional attitudes prior to knowing how to interrelate them by rehearsing and interpreting some empirical evidence from developmental psychology. In Chapter 7 of his (2008a) he sketches the emergence of these concepts in children from what I take to be a largely interpretationist point of view. In some detail he shows how linguistic competence extends the abilities involved in having a primitive grasp of desires and beliefs as ‘mere’ intentional attitudes into abilities that can be understood as having a grasp of beliefs and desires as propositional attitudes. In his account, Hutto takes care to emphasise the extent to which the acquisition of full-blown concepts of ‘desires’ and ‘beliefs’ are independent: ‘Desires come first in our linguistically scaffolded mentalistic understanding of things: children are capable of attributing these long before they can make competent belief ascriptions’ (Hutto, 2008a, pp. 135–6). In this subsection I want to argue that it is not at all clear that this does indeed show that the concepts of ‘belief’ and ‘desire’ are indeed acquired before children grasp the principles that interrelate them. Instead, I will argue, there is much to be said for the opposite.

First, a preliminary observation. Though it is well documented that the ‘linguistically scaffolded’ concept of ‘desire’ precedes the ‘linguistically scaffolded’ concept of ‘belief’ by some six months, this does not mean that the development of these concepts can be separated entirely. For both concepts build on earlier acquired non-linguistic abilities that are often (in my view overconfidently and somewhat misleadingly) referred to as the ‘nonverbal’ grasp of ‘beliefs’ and ‘desires’. The first occurrences of these abilities are not easily or uncontroversially separated temporarily. At any rate, both abilities (I deliberately avoid ‘concepts’ which I shall reserve for linguistically scaffolded abilities) occur long before their linguistic counterparts do. Thus, though controversial, it is claimed that 18-month olds already have some understanding of the desire of others:

(…) [C]hildren were shown two bowls, one of goldfish crackers and one containing broccoli. The experimenter would look cheerfully at the broccoli and say ‘Yum!’ and make a disgusted face toward the crackers
and say ‘Yuck!’ Then the experimenter would hold out her hand and ask, ‘Can you give me some?’ (…) 18-month-olds succeed in giving the experimenter the one that she showed a preference for, broccoli, even though it is not their own choice (Bloom, 2004, pp. 18–9).

Similarly, (and also controversially) Onishi and Baillargeon (2005) have claimed that 15 month-olds have some grasp of the false beliefs of others, using a violation-of-expectation version of the false belief task. Southgate et al. (2007) criticise this study for being ambiguous: the results can also be explained by the assumption that children attribute ignorance instead of false beliefs. But they themselves present a disambiguated version of the non-verbal false belief task, showing that 25-month olds appear to pass it. It is far from clear how we should interpret these data (see e.g. Herschbach, 2008). But it does seem clear that long before children acquire a linguistic concept of ‘desire’, which is six months prior to acquiring the concept of ‘belief’, they have been able to make non-linguistic ‘attributions’ that are at least in some sense predecessors of both the concepts of belief and desire.

So, when children develop the linguistic concept of ‘desire’ (around 3.5 years of age) it is not entirely true that this concept exists in isolation. Still, it is a well established fact that 3.5-year olds are able to attribute desires linguistically while not being able to attribute beliefs in that way. The first point I’d like to make is that, also in view of the previous remark about non-linguistic predecessors, there does not seem to be a clear demarcation of when a child can be said to have acquired a full-blown concept of a specific propositional attitude — it all depends on what one calls ‘a concept’. And given this, one begs the question against the mental holist when it is said that the concept of ‘desire’ is acquired before knowing how to interrelate it with e.g. the concept of ‘belief’. It makes perfectly good sense, from the point of view of the holist, to say that a child hasn’t acquired the full blown concept of ‘desire’ until she knows how it interacts with beliefs. If it is the case that a 3.5-year old does not know how a desire and a belief interrelate (the possibility of a linguistic concept of ‘desire’ interacting with a non-linguistic context in which there is a primitive grasp of beliefs, remains open), then she has acquired a proto-version of the concept of ‘a desire’ but not the real thing yet. The case for isolated desire concepts is inconclusive to say the least.

As for beliefs, Hutto’s case is even less strong, for the simple reason that by the time children are said to acquire that concept (by the age of 4), the linguistic ability to attribute desires is already in place. There may certainly seem to be instances in which children explain actions...
in terms of isolated belief attribution. But here we need to ask whether the lesson of the second example of the previous section is sufficiently heeded: are we sure there are no implicit desire attributions that play a role in the background? Consider what may seem to be a paradigmatic example of action explanation/prediction based on the attribution of a single belief: Baron-Cohen, Leslie and Frith’s (1985) version of the false belief task, the so-called Sally-Ann test. Here’s the standard picture that is used to illustrate the test:

![Sally-Ann Test Diagram]

The semi-final sentence makes the background-desire, in co-operation with which the false belief is supposed to explain Sally’s looking in the basket, entirely explicit. And even if we were to erase that sentence, the transition from ‘now Sally comes back’ to ‘where will Sally look for the ball’ obviously presupposes that Sally wants her ball. It would be absurd to claim that children do not understand the belief-desire interaction going on here. It is overwhelmingly likely that they do. And the possibility that they co-determine each other’s role and meaning is entirely open.

Figure 1
The standard picture illustrating the Sally-Ann test
Hutto does not deny this. He gives a wonderful three-page description of how the concept of ‘belief’ develops out of the cognitive friction that is certain to emerge in the conversational interplay that children in between 3 and 4 years are capable of. It is almost unthinkable (though Hutto is silent on this) that this interplay at all times avoids what we would characterise as belief-desire interactions e.g. of the kind illustrated by the first example in the previous section. Indeed Hutto writes that ‘children are more or less simultaneously learning what beliefs are while also becoming familiar with the specific role they play in folk-psychology. For it is during the period of 3 to 4 years of age that they start having the relevant kinds of conversational interplay with caregivers as well as being introduced to people narratives’ (Hutto, 2008a, p. 138). This statement is, at least at first glance, at odds with the separation of FP components and FP principles. The consistency of Hutto’s position might appear to be saved by emphasizing that narratives are introduced as an item that is logically distinct from the acquisition of beliefs. But the same cannot be claimed of conversational interchange: Hutto does describe the emergence of the concept of belief as requiring interchange and hence in all likelihood as being a product of having witnessed many belief-desire interactions.

Thus, even to some extent by Hutto’s own standards, it is very plausible that children know how beliefs and desires interact from the moment they have the full concept of ‘a belief’; the separation of FP components and FP principles cannot be illustrated by means of referring to the developmental psychological facts. And the kind of interaction that children seem to be aware of is exactly the kind mental holists refer to. Indeed it is of the kind presupposed by externalist TT.

2.3 The threat of collapse and a different strategy to avoid it

The fact that Hutto says that children start being exposed to (folk-psychological) narratives in between the ages of 3 and 4 does not mean that his claim is that by the age of 4 children have a rudimentary grasp of FP. On the contrary, while TT-ists claim that when children can pass the false belief test by the age of 4 (see, however, Carpendale & Lewis, 2004) this marks their ability to wield a primitive version of FP, Hutto argues that it is precisely the fact that children are only able to understand reason explanations at later ages — approximately from the age of 5 onwards — that illustrates the fact that passing the test is not enough for FP. The continued exposure to narratives, apparently, is needed, according to Hutto, for true FP competence.
What, then, is the relation between the knowledge of simple belief-desire interactions that children possess by the age of 4 (see previous section) and full blown FP competence? Precisely because Hutto does not admit that children by the age of 4 have at least some understanding of FP principles, he does not answer this question. But from his book and his many articles an answer can be distilled: complexity and subtlety. Full FP competence differs from the simple belief-desire psychology I claim children of 4 years acquire knowledge of, by being infinitely more complex (e.g. in plotting the interrelations of the attitudes with all sorts of other mental states such as emotions and perceptions as well as with items such as character traits and moods) and subtle (e.g. in charting how circumstances and a range of other variables might be relevant).

If this is indeed the relation between the simple belief-desire psychology of the 4-year old and the full blooded FP of adults, then it seems Hutto has no argument against the external TT-ist who claims that the narrative structure of FP does not exclude the idea that FP implicitly is still a term-introducing theory. Once it is recognized that the simple belief-desire psychology of, say, the false belief test is precisely the sort of psychology that proponents of ETT wish to draw on when devising their notion of FP as a term-introducing theory (‘this is how the notion of a belief functions: if you want $x$, believe that $y$-ing will get you $x$, you will $y$, ceteris paribus’), complexity and subtlety introduced through narratives will not help to make full-blown FP non-theoretical. It will only serve to make full-blown FP a complex and subtle theory.

Here, the way out for the NPH might seem to be to argue that the simple belief-desire psychology of the 4-year old is already narrative by nature. Indeed in the passage I quoted at the end of the previous subsection, that is what Hutto appears to do. And indeed, the Sally-Ann example is an example of a story! But this move does not help. If simple belief-desire psychology is called ‘narrative’, this very much undermines the informativeness of the notion of ‘a narrative’. The NPH depends, as a theory distinct from TT, on the contrast between the notion of ‘a story’ and the notion of ‘a theory’. This contrast disappears (especially given the absence of a definition of ‘narrative’ that sets it apart from ‘theory’) when one of the prime examples of the belief-desire interactions that serves to illustrate the idea of a term-introducing theory is called a narrative. Calling simple belief-desire interactions narratives will only encourage those who suspect that the NPH is a variety of implicit ETT.
But note that this threat of collapse of the NPH into a form of TT hinges on the idea that full-blown FP reason explanation is continuous with simple belief-desire psychology. By rejecting the idea that full-blown FP explanation is just a more complex form of the same type of belief-desire psychology that ETT accounts for, (i) the apparent compatibility of the NPH with ETT as outlined in section 1, and (ii) the failure of Hutto’s anti-holism strategy against ETT can be shown not to amount to the collapse of the NPH into ETT. Compatibility does not amount to a collapse when full-blown FP explanation is discontinuous with simple belief-desire psychology in the sense that the narrative element in FP explanation that is stressed by the NPH is a further addition to belief-desire psychology that is beyond the reach of ETT. I will argue for such discontinuity, and hence against the collapse of the NPH into TT, in the next section.

3. The Division of Labour between Belief-Desire Interaction and Narrativity

In this section I will argue that the idea of a discontinuity between simple belief-desire psychology and full blown FP reason explanation is, despite being unorthodox, indeed highly plausible. Narrative FP explanations do involve belief-desire structures. But the contribution of narratives to the explanation of actions cannot, I claim, be understood as an extension in the direction of complexity, context-dependence and subtlety of belief-desire schemata.

3.1 Some examples and their analysis

In order to distinguish the roles of belief-desire interaction on the one hand and narratives on the other in full-blown FP explanations of actions, it will be helpful to consider some examples of such explanations. Oddly enough, there are not that many examples of full-blown narrative explanations of actions to be found in Hutto’s book. A striking one, though, is to be found in the beginning of the book. It is about the explanation of some initially incomprehensible behaviour by Hutto’s wife. Hutto writes:

On the morning of my flight [my wife] agreed to drive me to Heathrow after I had first dropped off my car at the garage. So, we set off in our separate cars and she took the lead (…). As we came to the relevant intersection, she stopped at a set of red traffic lights, but uncharacteristically failed to signal. (…) To my amazement when the lights changed she did not turn, began driving towards the town center, straight past the garage at which she herself made the booking. (…) At this point, I was
faced with a rather tricky interpretive problem. Given that my wife is very competent and reliable, lacking any malicious streak or any known reason to treat me badly, I was at an utter loss to make sense of her actions (Hutto, 2008a, pp. 18–9).

Afterwards, Hutto’s wife’s explained her behaviour in narrative form:

(…) After the incident she explained that although it was true that she had phoned the garage to make the appointment herself, and she had used the number I had given her, she believed it was the number of our old garage, which is located in the next village (Hutto, 2008a, pp. 19–20).

This example typically shows what Hutto takes to be the point of narrative FP explanations: ‘Folk psychological narratives are used to make sense of (…) seemingly aberrant actions. Typically — or at least when they work — they help us to understand why someone has acted in a way that has strayed from our normal expectations (on the assumption that they have acted for a reason nonetheless’ (Hutto, 2008a, p. 37). This idea about the function of narrative explanation is shared by Alisdair Macintyre (MacIntyre, 1981). Consider the following example:

I am standing waiting for a bus and the young man standing next to me suddenly says: ‘The name of the common wild duck is *Histrionicus histrionicus histrionicus*.’ There’s no problem as to the meaning of the sentence he uttered: the problem is, how to answer the question, what was he doing in uttering it? Suppose he just uttered such sentences at random intervals; this would be one possible form of madness. We would render his action of utterence intelligible if one of the following turned out to be true. He has mistaken me for someone who had yesterday approached him in the library and asked: ‘Do you by any chance know the Latin name of the common wild duck?’ Or he has just come from a session with his psychotherapist who has urged him to break down his shyness by talking to strangers. ‘But what shall I say?’ ‘Oh, anything at all.’ Or he is a Soviet spy waiting at a prearranged rendez-vous and uttering the ill-chosen code sentence which will identify him to his contact. In each case the act of utterance becomes intelligible by identifying its place in a narrative (Macintyre, 1981, p. 210).

[4] Macintyre also seems to agree with Hutto (or, given the fact that Macintyre’s book appeared in 1981, Hutto with Macintyre) on the genesis of our narrative FP capabilities: ‘It is through hearing stories about wicked stepmothers, lost children, good but misguided kings, wolves that suckle twin boys, youngest sons who receive no inheritance but must make their own way in the world and eldest sons who waste their inheritance on riotous living and go into exile to live with the swine, that children learn or mis-learn both what a child and what a parent is, what the cast of characters may be in the drama into which they have been born and what the ways of the world are. Deprive children of stories and you...
These are helpful examples. But what do they show? Do they show that there are highly complex schemes of beliefs and desires at play in narrative explanations in which these beliefs and desires interact with other psychological states, situations, character traits, etc. They might. But I think there is a much more elegant and clear analysis.

Take Hutto’s own example. Behind the story, a fairly simple belief-desire structure is visible. Hutto’s wife:

\[\text{wants}\] to drive him to the airport after having dropped his car off at the garage.

She also

\[\text{believes}\] that the phone number Hutto gave her was the number of their old garage, so that the appointment she had made was an appointment with that old garage in another village.

There are, of course numerous background beliefs (about traffic lights, traffic rules about when to signal, etc.) and desires (wanting to abide by the rules when driving, etc.) at play. But that is also the case in the most simple belief-desire explanations (think about Sally’s background belief that balls do not change place by themselves and usually stay where they are put). So these background beliefs and desires are not at play because of any narrative involvement; which is exactly why Hutto omits them too in his recounting of this particular narrative explanation. Where, then, does narrativity enter in this example?

Well, look at the individual belief and the individual desire cited above. The belief that ‘the phone number Hutto gave his wife was the number of their old garage, so that the appointment she had made was an appointment with that old garage in another village’ itself contains a narrative. And similarly for the desire ‘to drive Hutto to the airport after having dropped his car off at the garage’. That is what the narrative element does in this particular example: it fills in the individual belief and desire of an otherwise simple explanatory belief-desire scheme, rendering the belief and the desire themselves very complex. But not by breaking them down into further beliefs and desires. Even if that were a possibility, it would be cognitively much more labourious than the narrative procedure of inserting a compressed sequence of events relevant to the action to be explained, leaving all the attitudes that might have been involved aside (note that the narrative contents of the above belief and desire do not contain further

---

leave them unscripted, anxious stutterers in their actions as in their words’ (Macintyre, 1981, p. 216)
attitudes). Compare this with the same simple belief-desire scheme in the Sally-Ann case: Sally wants her ball and believes it is in the basket. That’s it. The difference with Hutto’s example, the difference that narrativity makes, is in the complexity, not of the belief-desire scheme, but in the belief and the desire themselves.

The outcome of an explanation in which the beliefs and the desires themselves have a narrative form is that the action that is explained is made intelligible not merely as the logical result of attitudes immediately preceding the action, but as a part of a longer story, i.e. as coherently fitting into the agents autobiography. This diachronic element in the explanation is not a part of the belief-desire interaction that 4 year-olds are to some extent capable of understanding, it is a part of the narrative contents of the belief and the desire themselves that 4 year olds are not yet capable to entertain. The work that is done by narrativity in Hutto’s example is different from the work that is being done by the ‘formal’ belief-desire interaction. The simple belief-desire structure in Hutto’s narrative explanation is the skeleton of the explanation. The real work in the explanation is done by the narratives that have the form of a belief and a desire. To use a different metaphor, the belief-desire structure is the canvas on which the actual narrative explanation is painted.

What about Macintyre’s example? In the first reading of the example, the person standing next to Macintyre (or the ‘I’in the example) at the bus stop believes that he has met Macintyre in the library yesterday when Macintyre (or the ‘I’) asked him what the Latin name of the common duck is while he failed to provide the answer he now has ready at hand. He desires to help Macintyre (or the ‘I’) by providing him with the answer, even if it is one day late. Again, there is a simple belief-desire structure (again with a huge background of other beliefs and desires not relevant to the narrativity at play) which provides the skeleton of the explanation. And again the individual belief and desire are filled-in by the narratives. The same analysis applies to the other two readings, but I take it that the point is clear.

What is the role of narratives relative to the belief-desire structure of an explanation? I would say it is the chunking — in the psychologist’s sense of ‘compressing’ — of large amounts of diachronically ordered information relevant to the explanation of the action. The narratives that ‘fill’ the beliefs and desires of full-blown FP explanations do not merely summarise more beliefs and desires, they do something more. They introduce a diachronic ‘logic’ to the events that are either believed or desired. On the one hand this diachronic logic makes a large amount of information highly tractable by highlighting only
relevant events, where ‘relevant’ means ‘relevant to the story’. On the other hand this diachronic logic makes it possible to make an action intelligible by letting it fit into this logic from the point of view of the agent.

So, this is the hypothesis: children by the age of 4 have mastered simple belief-desire interaction (to some extent at least), but they have not yet acquired the ability to chunk large amounts of diachronically ordered information in the specific way that is relevant to mature FP explanations. For that they need to acquire narrative competency. And they start to acquire that, roughly, from the age of 5 onwards. In this respect it is interesting to note (as Hutto does in a footnote, 2008a, p. 254) that

whereas pre-school-aged children tend to connect events in a linear fashion, 5-year olds begin to use complex syntax to impose a causal temporal hierarchy that contrasts important events with background information (Capps et al., 2000, p. 201).

That, I hypothesize, is precisely what narratives do: compress diachronically ordered and structured information, highlighting the relevant bits. That information, thus compressed, gets fed into the belief-desire structures children already are familiar with.

3.2 Persons

But there is more that narratives do than just compress information. They structure diachronically related events in a specific format (the MP3 of FP), i.e. along the axes of embodied persons as diachronically existing unified characters. The point I want to make in this short subsection is that the acquisition of narrative abilities by children after they have acquired some insight into belief-desire interaction can be characterized by the development of the notion of ‘a person’ or ‘a self’ as a diachronically coherent agent.

Unlike Hutto, Macintyre makes a point of highlighting the fact that the notion of an agent, person or self as a diachronically unified being on the one hand and the notion of a narrative, story or history on the other require each other. For Macintyre it is even impossible for us to construe the idea of personal identity without our narrative capabilities:

Empiricists (…) tried to give an account of personal identity solely in terms of psychological states or events. Analytical philosophers, in so many ways their heirs as well as their critics, have wrestled with the connection between those states and events a strict identity (…). Both have failed to see that a background has been omitted, the lack of which
makes the problems insoluble. That background is provided by the concept of a story and the kind of unity of character which a story requires (Macintyre, 1981, p. 217).

We need not go as far as Macintyre does with respect to the problem of personal identity in order to appreciate the relevance of this to our present discussion: persons-as-diachronically-unified-selves are an inalienable part of the FP narratives Hutto’s NPH is about. So, if narrative competence starts to emerge around the age of 5, that is when children start to grasp the idea of people as psychologically diachronically unified agents. Of course smaller children also have a grasp of, say, their parents as diachronically continuous beings. But just like the abilities underlying what Hutto calls ‘unprincipled bodily engagements’ get transformed through the use of language into belief-desire psychology (Hutto, 2008a, ch. 5–7), there is a further transformation marked by the acquisition of narrative capacities: the primitive embodied notion of ‘person’ becomes a diachronically unified psychological entity, a self. Note again that we need not go as far as Macintyre does: it may be the case that narratives constitute selves, so that the ability to grasp narratives allows for the emergence of the FP notion of ‘a person’. But it may just as well be that by the age of 5 children start to acquire the capacity to grasp the idea of the diachronic unity of psychological beings as such. And that may parallel the emergence of narrative capacities, or even explain it. It is even more likely, to my mind, that there really is no fact of the matter here about which side is right. Keeping the quote by Capps et al. at the end of the previous subsection in mind, I would say that it is most likely that both narrative capacities and the concept of persons or selves arise from developments in the syntactical abilities of 5-year olds as two sides of the same coin. Thus, the idea is that just as unprincipled bodily engagements get transformed through language into belief-desire psychology by the age of 4, a further development of linguistic abilities allows for the introduction of the notion of persons by the age of 5; not so much as a term that is used as frequently as ‘believes that’ or ‘wants to’, but as the characteristic ‘compression format’ of narrativity.

3.3 Conclusion

The functions of narrativity in FP explanations — compressing information, introducing a diachronic logic to events from the viewpoint of the agent, introducing the notion of a diachronically coherent person in reason explanations, etc. — cannot be reduced to the functions of
belief-desire interaction. That, at least, was the point of this section. Since ETT is merely about the interaction of propositional attitudes — however sophisticated and complex — it follows that mature FP explanations fall outside the scope of ETT to the extent that they involve narrativity. The compatibility of the NPH with ETT (Section 1) and the improbability of the anti-holism attack on ETT (Section 2), then, do not and cannot amount to a collapse of the NPH into ETT. Moreover, Hutto’s claim that TT doesn’t explain real-life FP (or FP stricto sensu) remains intact. Not, however, because all versions of it are crucially wrong, but because all versions of it leave something crucial out.

References