



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Mindshaping and the intentional control of the mind

Citation for published version:

Vierkant, T & Paraskevaides, A 2012, Mindshaping and the intentional control of the mind. in F Paglieri (ed.), *Consciousness in Interaction: The role of the natural and social context in shaping consciousness.* Advances in Consciousness Research, vol. 86, John Benjamins Pub Co, pp. 105-124.

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Consciousness in Interaction

Publisher Rights Statement:

© Vierkant, T., & Paraskevaides, A. (2012). Mindshaping and the intentional control of the mind. In F. Paglieri (Ed.), *Consciousness in Interaction: The role of the natural and social context in shaping consciousness.* (pp. 105-124). (Advances in Consciousness Research; Vol. 86). John Benjamins Pub Co.

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Mindshaping and the Intentional Control of the Mind

Tillmann Vierkant, Andreas Paraskevaides

Tillmann Vierkant (t.vierkant@ed.ac.uk) is a lecturer at the philosophy department of the University of Edinburgh. He works currently on conscious will, metacognition, mental actions and self-regulation.

Andreas Paraskevaides (andreas.parask@gmail.com) has recently obtained a PhD in Philosophy from the University of Edinburgh. In his thesis, he examines social constraints on human agency. His research interests include self-knowledge, self-regulation and the mindshaping aspect of folk psychology.

Abstract. Understanding and controlling our minds is one of the most fascinating features of human cognition. It has often been assumed that this ability requires a theoretical understanding of psychological states. This assumption has recently been put under pressure by so called mindshaping approaches. We agree that these approaches provide us with a new way of self-understanding and that they enable a very powerful form of self-regulation which we label narrative control. However, we insist that there still is a crucial role for a theoretical understanding of psychological states in human cognition. We argue that this is because a theoretical understanding is necessary for all intentional control of the mind.

Keywords: Mindreading; Mindshaping; Narrative Control; Manipulative Control; Folk Psychology.

Introduction

In this paper, we focus on the nature of our understanding of human mentality and on the kinds of control this understanding enables us to exercise over our own psychology and actions. The nature of our familiarity with our own and others' minds has been a persistent puzzle, and for good reason. We routinely offer reasons for our actions as if we're in a position to know what was going on in our minds at the time of action. We also regularly rely on others' justifications of their own actions and their claims on their mental states, as well as offering explanations of our own of others' behaviour. This familiarity we share with each other also allows us to have certain expectations on what others will do and anticipate accordingly their resulting behaviour. Our explanations and predictions about others, which rely on our familiarity with them, are made with less confidence than in our own case and we also

make plenty of mistakes and are sometimes frustrated in our attempts to understand one another. Certain utterances and actions can leave us at a complete loss, unable to discern the reasons behind them. Despite this, we somehow remain able to efficiently communicate and collaborate in most social settings, which would seem impossible without a significant predictive and interpretive success in our interactions.

But where does this success come from? Why is it that we rely so much on our presupposed familiarity with each other's mind, which allows us to make sense of each other's actions and to trust each other's justifications? After all, most of us folk are not trained psychologists. We lack the kind of formal training that would give us an in-depth understanding of human psychology. This doesn't stop us from engaging in a constant exchange of reasons that consists in a stream of interpretations, inquiries, clarifications, evaluations and accusations. We do not simply defer to psychologists as the ultimate authority on our thoughts and actions, expecting them to unravel the tangle of our relations and offer us an objective interpretation that we can use as a point of reference. This seems like borderline arrogance on the part of us folk. What is the source of this confidence in our comprehension of human mentality? Isn't there a distinct possibility that this confidence is actually misplaced and that all of our social interactions are based on the mistaken premise that we are in a position to penetrate the pervasive mystery of human psychology?

A traditional response to these worries is to claim that even us folk who are not certified psychologists have a capacity to accurately represent our own and other's mental states. That is because we still have a systematic pre-theoretical understanding of what intentional states such as beliefs and desires are that we use in our interpretations, predictions and evaluations of human action. This understanding is pre-theoretical because while it does enable folk to appropriately categorize intentional states by reference to their common causes and effects and to the relations between these states, it is not explicitly formulated as a scientific theory. The advocates of this response are of course the advocates of theory theory as an explanation for what the practice of folk psychology consists in. The variety of arguments for this view and its juxtaposition with simulation theory are well documented and thus will not be discussed here¹.

¹ For comprehensive overviews of the debate between simulation and theory-theory see Davies and Stone (1996, 2000).

What we do want to stress is that the advocates of this view correctly emphasize that humans do have a capacity to mindread² that depends on our representation of intentional states such as beliefs as representational and psychological. We have the capacity to metarepresent, in other words, because we can understand mental states such as beliefs and desires as psychological states, which represent or misrepresent the world. However, we do think as well that theory theorists have attributed far too much explanatory weight to these abilities. We do not think that metarepresentation is required for most of the important functions that folk psychology has in understanding and shaping minds. Still, neither do we think that theory theory is simply completely mistaken and that these quasi-theoretical structures play no or just an insignificant role. On our view, metarepresentation does allow for a very specific form of understanding minds³ and more importantly it is crucial in the intentional regulation of the mind. Towards the end of this paper we will try to spell out exactly how metarepresentation is crucial here. But in order to work out exactly what the contribution of metarepresentation is it is essential to first separate it from a quite different way in which the practice of ascribing folk-psychological states to each other helps us to understand and shape minds which is entirely independent of metarepresentational abilities.

In order to do this, we wish to draw attention to a different interpretation of our understanding of human mentality, which relies on the insight that the reason we are such adept folk psychologists is not because we are particularly competent mind-readers, but because our interactions have shaped our expectations and understanding of what intelligible action involves and these interactions provide us with a set of norms that we rely on to make sense of each other's actions. The social framework wherein our actions take place is a constant influence on our minds and actions and our understanding of human mentality depends on our social interactions and the common expectations and commitments that constitute these interactions. An appropriate label for this approach to explaining our understanding of human minds is "mindshaping" (Zawidski, 2008). As the friends of mindshaping have it, "minds are as much made as discovered" (McGeer, 1996, p.512).

² We owe this use of the term "mindreading" and its subsequent juxtaposition to "mindshaping" to Zawidski (2008).

³ We do not want to commit ourselves to any position with regard to the question whether or not this way of understanding minds is necessary for predicting behaviour in special situations like the famous false belief task.

Minds in the making

The main idea behind mindshaping is that the folk-psychological practice of evaluating and inquiring into the reasons for each other's actions has a transformative effect on our self-understanding, that also carries over to the way we treat each other's actions and assertions aimed at justifying these actions. Throughout our mutual development in a common social framework, we develop various norms of intelligibility, that is, expectations and guidelines for what counts as an intelligible action and what counts as an intelligible reason for action in a given circumstance. The reason our interpretations and predictions are often successful in our social interactions is that we have learned to present ourselves in ways that fit this context, ways which can be understood in the context of the norms of intelligibility that we all share. To clarify this idea and its implications, we will provide some examples from the work of authors (see Hutto, 2004, 2007; McGeer, 2007; Zawidski, 2008) that have diverged from the traditional understanding of folk-psychological practices that treats them as primarily consisting of mindreading.

Victoria McGeer (2007) argues that the problem with focusing on just the predictive and explanatory aspects of folk psychology is that “we overlook the way folk psychology operates as a regulative practice, moulding the way individuals act, think and operate so that they become well-behaved folk-psychological agents: agents that can be well-predicted and explained using both the concepts and the rationalizing narrative structures of folk psychology” (McGeer, 2007, p. 139). In McGeer's view, learning to engage in folk psychology involves more becoming skilled in making ourselves intelligible from a folk-psychological perspective that depends on the common understanding and expectations we share of intelligible behaviour, and less becoming skilled at accurately representing the mental factors that influence human action.

Our ability to create and respond to these narrative structures of sense-making folk psychology, or the norms of intelligibility that guide our actions, has also been explored by Daniel Hutto (2004, 2007, 2008) and Tad Zawidski (2008). According to Hutto, there is often no need to engage in mind-reading in order to explain one another's actions, since we can tell stories about our actions that can reveal the reasons behind them without needing to have an understanding of intentional states *as*

intentional states, so without needing to metarepresent. It is our capacity to exchange these justifying narratives that underlies our social interactions, Hutto argues, and not any special affinity on our part for accurately predicting and explaining human behaviour by successfully representing human mentality⁴. For Hutto, “‘folk psychology’ is an instrument of culture, giving us the grounds for our evaluative expectations for what constitutes good reasons. This is not the same as merely providing a framework for disinterested explanation and prediction.” (Hutto, 2004, p. 559)

As such, what makes folk psychology unique for us is that it enables us to justify each other’s actions by reference to our reasons for acting and to hold each other responsible for offering such justifications and being able to live up to them⁵.

Tad Zawidzki (2008) also elaborates on this line of thought, by arguing that in order to explain how folk-psychological practices have persisted for so long we need to understand their function of setting a standard used as a guide for the ways in which we can intelligibly behave. One of Zawidzki’s main arguments for the prominence of this aspect of folk psychology is that if folk-psychological practices had the main function of identifying the mental states guiding our behaviour and anticipating their effects, the fact that they have persisted for so long and have not been eradicated during our evolution as a species seems particularly puzzling, considering the limits in engaging in such practices. To illustrate this point, he makes an analogy to traffic rules. If all drivers had to actively predict and interpret each other’s behaviour from moment to moment in order to cooperate, chaos would ensue before too long.

Fortunately, there is a framework wherein all drivers interact, which supports their attempts to anticipate and explain each other’s actions. This framework is established through the common understanding of these rules that competent drivers display and

⁴ For the idea that we use narratives to understand and justify each other’s actions, see also Bruner, 1990, whose work has influenced both McGeer and Hutto’s theoretical standpoints.

⁵ See Hutto, 2004, p. 565: “[T]he traditional picture is only attractive if we assume that in giving explanations we always occupy an estranged, spectatorial point of view. Yet, in ordinary cases the other is not at arms length. For this reason the *standard* way we come to determine the reasons for which others act is dramatically unlike that employed in forensic investigations that seek to locate the cause of particular events. We cannot use the same sorts of methods we would deploy in determining, say, the cause of a plane crash. Rather, we usually rely on the revelations of others. They explain their actions for themselves. Of course, their admissions are defeasible and often people are self-deceived about their reasons for acting. But we have fairly robust methods for testing, questioning and challenging such aims when it is important to do so, as in legal cases. We compare one person’s avowals with the accounts of others, uncovering lies or internal contradictions that will invalidate either their testimony or their credibility. Countless everyday conversations involving the explanation of actions in terms of reasons mimic this process to a greater or lesser degree.”

expect others to share with them. The fact that all drivers are expected to conform to this shared understanding of their situation makes anticipation and explanation of driving actions easier for them, assuming they are motivated to do their best in respecting the basic rules of traffic that constitute their shared understanding⁶. As drivers conform to a basic understanding of traffic regulations, so do human agents in general conform to the norms inherent in their folk-psychological practices. As Zawidski writes:

“[E]volution discovered simple mechanisms for shaping hominid behaviour so as to make it more predictable, or at least easier to coordinate with. Among these was the practice of ascribing propositional attitudes defined by normative relations to each other [...], to which, thanks to various mechanisms of socializations, hominids strive to conform. Because of this, solutions to coordination problems do not depend on reliably accurate predictions based on correct ascriptions of cognitive states- an epistemically intractable task. Rather, they depend on figuring out what the *normatively sanctioned* response to some problem is, and assuming that others do the same [...]. This assumption is justified by the efficacy of mechanism and practices of mind shaping.” (Zawidski, 2008, p. 199)

A view that makes mindreading the most essential ingredient for explaining our success in understanding the human mind also faces the problem that the practice of positively or negatively evaluating agents for their actions seems to be made redundant. When agents frequently fail to reliably conform to the predictions and explanations offered for their actions, for example, the mindreading view seems to imply that these predictions and explanations are somehow flawed. If failing to understand certain agents' actions was only a matter of displaying various inaccuracies in our folk-psychological attributions, for example, then it seems that the proper response to such failures would be to change our attributions so that they more accurately reflect the behaviour we are trying to account for. But then there would be no need to confront the objects of our folk-psychological understanding for failing to live up to this understanding, since it seems that the fault would be with our way of seeing things and not with their behaviour. Positive and negative

⁶ See Zawidski, 2008, p. 199: “There is no way that we can divine the cognitive states of fellow drivers, in the heat of traffic, with sufficient speed and accuracy to avoid catastrophe. Fortunately we do not need to. This intractable epistemic task is off-loaded onto our social environment. Legislatures pass laws and educators teach novices in such a way that the coordination problem becomes exponentially more tractable.”

evaluations seem out of place in such a context. This point can be further clarified through Zawidski's traffic analogy. As he points out, when a driver makes a mistake, the traffic regulations don't usually change to accommodate this mistake. Instead, the driver is confronted for failing to live up to the common expectations shared by all competent drivers⁷.

Paying closer attention to the contrast between mindshaping and mindreading brings us to another issue that is worth clarifying⁸. In our exposition of these approaches, we treat mindshaping as a competing approach to mindreading in order to highlight the fact that mindshaping need not involve the attribution of intentional states for predictive and interpretive purposes. Appropriately responding to the norms of intelligibility that guide our actions does not have to depend on possessing an understanding of intentional states as intentional states. However, we do not wish to argue that such an understanding is excluded by the process of mindshaping. It is possible that learning to think and act in intelligible ways also leads to developing an understanding of intentional states as intentional states, though this understanding is not necessary in order to explain every aspect of our capacity to act as intelligible folk-psychological agents⁹.

Now, the process of mindshaping, from all that's been said so far, can be viewed as being largely automatic. What we mean by that is that regulating our minds and actions so that they become intelligible from a folk-psychological perspective is

⁷ See also McGeer, 2007, p. 148: "This is one of the most telling features that differentiates folk psychology as a regulative practice from what it would be like if it were a mere explanatory-predictive practice, appropriately construed as a proto-scientific theory of behaviour. For in the case of a proto-scientific theory, failure in explanation and prediction should lead to some revision in the theory itself or in the way the theory is applied; it does not lead to putting normative pressure on the "objects" of theoretical attention themselves to encourage them to become more amenable to folk-psychological explanation and prediction on future occasions."

⁸ We are grateful to Fabio Paglieri for bringing this need for clarification to our attention.

⁹ We are also in agreement with Zawidski that the extent to which mindshaping involves mindreading is an empirical question. See e.g. Zawidski, 2008, pp.204-205:

"[I]t is possible that mind shaping functions to socialize individuals such that they are more likely to token the kinds of propositional attitudes, and engage in the kinds of behaviors that their typical interpreters expect. On this view, accurate descriptions of mental states supporting accurate predictions of behavior remain central functions of propositional attitude ascription. However, propositional attitude ascriptions succeed in realizing these functions only to the extent that they also succeed in prior shaping of individuals, to make 'abnormal' propositional attitudes and behaviors less likely in populations of interactants...Nothing in the foregoing is meant to suggest that human beings do not predict each other's behavior, nor even that they never use mental state ascription to this end. Once the use of mental state ascription to mind shape is reliable and prevalent, a derivative mind-reading use is possible, much as we predict that motorists will stop at red lights. The more effective mechanisms of socialization are at molding individuals capable of and willing to conform to the norms of folk psychology, the easier it is to predict individuals in such terms."

something that is already taken care of for us by the norms and expectations that are part of our social environment. As McGeer also notes,

“If we learn to govern our behaviour in ways that make us more readable to others, then their work as interpretative agents is greatly reduced. The same is true for us, if they learn to govern themselves likewise... When we develop as folk psychologists, we no doubt hone our interpretive skills; but, more importantly, we come to live in a world where the kind of interpretive work we need to do is enormously enhanced by how much meaning our interactions already carry for us and carry because of the way we habitually conform to norms that invest our actions with common meaning.” (McGeer, 2007, pp.149-150)

Our habitual conformity to the shared norms of intelligibility is not then something that often requires intentional effort on our part. A lot of the time, it seems that as long as we have properly developed as competent self-knowing agents that are in a position to offer intelligible reasons for their actions, we simply know what to say and do in order to coherently respond to the various social settings we find ourselves in. So as long as our understanding of human mentality has been the product of the normative pressures that underlie our mutual social development, figuring out the most intelligible responses to our environment so that we make sense of ourselves and of others will be something that comes naturally to us.

Be that as it may, we wish to argue that our understanding of human mentality that has been the product of a long process of mindshaping can still allow us to exercise different kinds of control over our minds and actions. There are two kinds of control that we wish to draw attention to. Those are narrative and manipulative control. In discussing narrative control, we will show how the narratives we rely on in our social interactions allow us to intervene on our actions when they fail to fit our self-understanding. In discussing manipulative control, we will show that the capacity for metarepresentation, or mindreading, while not being at the heart of our social development and understanding of human minds can still have important implications for the kind of control we can exercise on our intentional states and actions.

Narrative control

The two authors that will help us get to grips with the role narratives play in our actions are Daniel Dennett (1992, 1993, 1996, 2003) and David Velleman (2009).

Dennett, in his (2003), express the following worry: “Aren’t we learning from psychologists that we are *actually* a far cry from the rational agents we pretend to be?” (Dennett, 2003, p. 268). His response is that, actually, by pretending to be agents we make ourselves into agents. Having the ability to inquire into one another’s thoughts and actions led to an increase in the sophistication of our capacity to monitor and control our behaviour. The development of a language in which we could express our reasons for acting, making them publicly available objects of inquiry, was the decisive step in this process (see also Dennett, 1998).

As Dennett’s story goes, there was a significant change in the kind of control we could exercise on ourselves, that depends on our capacity to offer reasons for our behaviour and expect others to do the same:

“We human beings can not only do things when requested to do them; we can answer inquiries about what we are doing and why. We can engage in the practice of asking, and giving, reasons. It is this kind of asking, which we can also direct to ourselves, that creates the special category of voluntary actions that sets us apart.” (Dennett, 2003, p. 251).

Pretending to be an agent, in accordance with Dennett’s story, involves trying to live up to the reasons that would make it possible for one to be treated as an agent. Trying to live up to these reasons, which can be expressed publicly and thus be used in evaluating one another’s behaviour, leads to developing a degree of self-monitoring and self-control that is instrumental to being able to fit one’s behaviour to these reasons.

A further idea that is relevant for our discussion and that appears in various guises in Dennett’s writings (see especially Dennett 1992, 1993), is that we develop self-concepts which enable us to treat our actions as the manifestations of a stable, coherent point of view. A self-concept is essential for being able to live up to one’s reasons for acting, because, to put it in a way closer to Dennett’s terminology, the agent can tell stories about what he’s doing and why, stories that fit the attitudes expressed in this concept. One of Dennett’s more recent formulations of this argument goes like this:

“The acts and events you can tell us about, and the reasons for them, are yours because you made them-and because they made you. What you are is that agent whose life you can tell about. You can tell us, and you can tell yourself.” (Dennett, 2003, p. 255)

Velleman (2009) further develops the idea that we create narratives about ourselves, which he argues that we use in regulating our actions. For Velleman, the agent is a little like an actor improvising his role in a play.¹⁰ There's no set script guiding his every move, no ready-made role that he must enact. The behaviour he exhibits in the play, such as the lines he delivers, is being made-up by him on the spot. It also takes the form of a response to the actor's ever changing circumstances within the play he participates in. Other actors also improvise their lines and movements and the actor has to take those into account when creating the role that he enacts. He tells stories about himself and his circumstances and he acts as the protagonist of these narratives. In a sense, the actor's responses to his environment reveal his reasons for acting, because they reveal the motives that he takes himself to have. When the actor manifests these motives, he knows what he's doing because he's doing what the persona in the role he enacts would do and knows what his motives are because he has the capacity to manifest them in his actions. His behaviour makes sense to him as behaviour that arises from the stories he tells about himself and his circumstances.

The next step taken by Velleman is to remove the line between the actor and the persona that he's playing. The agent is playing himself according to the narratives created in response to his circumstances. His narratives about himself, created in response to his circumstances, feed back into his behaviour as a guiding influence.

The agent's narratives of himself influence his behaviour because the agent wants to achieve an understanding of himself through his behaviour. Returning to the actor analogy, if the persona the actor was playing was facing its archenemy then the actions that would make sense as expressing this persona would be to treat the enemy as such, perhaps by attacking him. If the actor went on with the play by treating his enemy as his best friend then this kind of behaviour would presumably make no sense either to the actor or to his enemy. Excluding special explanations of what the actor is up to, the spectators of his performance would agree that he has no idea what he's

¹⁰ See e.g. Velleman, 2009, p. 14:

“Imagine an actor who plays himself, responding to his actual circumstances and manifesting the occurrent thoughts and feelings that the circumstances actually arouse in him, given his actual attitudes and traits. This actor improvises just as he did when portraying a fictional character, by enacting his idea of how it would be understandable for his character to manifest his thoughts and feelings under the circumstances. But now the character is himself, and so what would be understandable coming from the character, given the character's motives, is what would be understandable coming from him, given the motives he actually has. Thus, he manifests his actual thoughts and feelings, as elicited from his actual makeup by his actual circumstances, in accordance with his idea of what it makes sense for him to do in light of them.”

doing or what his motives are. In the same way, an agent acts in ways that are intelligible, because they fit the narratives he creates about his circumstances and so they fit his own self-understanding.

Velleman's story can fit the context of our own view of the role of mindshaping in our social development, as he extends his analogy with self-improvisation to our interpersonal relations. We don't only enact a role for ourselves, he writes, but we also need to present ourselves in a coherent manner to other self-improvisers. We not only need to make sense of our actions as manifesting a coherent point of view but we also have to present these actions as such to others, in order to engage in coherent interactions¹¹. Because of this aspect of our nature, the enactment of our self-understanding depends on a mutual understanding of what it means to engage in intelligible enactments of this sort. Like a troupe of collaborating self-enactors, we depend on one another's cues for understanding the context in which we find ourselves, and the actions that would be available within this context.

In this framework, Velleman uses the example of crying performed as an action. We will use this example to illustrate the kind of action that we take to be the paradigm of narrative control. Maintaining the analogy with self-improvisation, Velleman describes his case in the following manner:

“Imagine that the improviser is terribly sad and disposed to cry. Maybe he starts to cry involuntarily, in an uncontrolled outpouring of emotion. If he doesn't already know that he is sad, and why he is sad, then crying won't fit his conception of the character he is playing, and so it will strike him as an unlikely thing for the character to do. He will therefore try to stop crying, so as not to spoil his rendition of the character. But if he knows what he is crying about, then crying may strike him as the very thing that his character would do under the circumstances, and so it will strike him as the very thing to enact, drawing on the tears supplied by the emotion. He will therefore cry out of genuine sadness but also under the guidance of his own conception of crying as what it would make sense for his character to do.” (Velleman, 2009, p. 15)

Applying this analogy to our case when acting as agents, Velleman distinguishes among three forms that the behaviour of crying can take. To one extreme is the case of crying as an emotional outburst outside our control, as when something

¹¹ For an empirically informed insight into the need for coherence of conduct and the sense in which such coherence supports our social interactions, see Miceli and Castelfranchi's article in this volume.

unexpectedly causes us great pain and we react to it by releasing tears. On the other extreme is an entirely deliberate manifestation of crying, as when we start crying without letting ourselves get carried away by our emotions. However, in between these extremes there are the cases when we do let ourselves cry and get carried away by our emotions, while still remaining in control of our action because we have the capacity to stop if at any point this action stops being a coherent response to our circumstances; if at any point, that is, this action does not fit our self-understanding. These latter cases, we maintain, are cases in which we exercise narrative control over our actions.¹² Our actions are controlled by narratives which help us to interpret what it is that we are doing. Crucially this interpretation is constantly up for grabs. We can give reasons for and against our narrative evaluations of the situation, others can easily make us aware of important reasons for or against our evaluations and we can easily reevaluate. In other words, in the narrative world there are constantly new narratives, which allow us to constantly reaffirm or reevaluate our prior narrative understanding of the world.

In this sense narrative control can be easily ‘meta’, which means here that narrative control allows for a reflective stance towards other narrative. It enables us to think about the norms that guide our actions and shape our minds.¹³

Nevertheless, we argue this is a different sense of ‘meta’ than the one employed in metarepresentation as discussed at the beginning of the paper. Therefore, it is now time to look at the limits of narrative control and at other forms of control which require metarepresentation.

Intentional control of the mind: The limits of narrative control

In the previous section we described an extremely powerful mechanism of human self-understanding and self-regulation. In fact this mechanism is so powerful that

¹² See Velleman, 2009, pp. 24-25, footnote 16:

“There is also an intermediate stage between losing oneself in an activity and consciously putting it into action. Even when letting oneself get carried away by a behavior such as crying, one can retain enough self-awareness to pull up short if the behaviour becomes discordant with one’s thoughts. In this third case, one’s thoughts and one’s behavior proceed in parallel, connected only counterfactually by one’s readiness to stop if the two should diverge... This ability to think along with oneself in this way, with thoughts that neither follow nor lead one’s behavior, depends on a degree of self-knowledge that can be attained only through long practice in the more deliberate, thought-first mode of action.”

¹³ For a more detailed analysis of what is ‘meta’ about language based narrative self regulation and the contrast to metarepresentation see Vierkant (forthcoming).

some of the authors supporting the mindshaping approach (Hutto, 2008) have argued that there is no need to postulate a quasi-theoretical understanding of folk psychological states in order to understand human self-regulation and self-understanding. This section is supposed to demonstrate that this is too strong a claim. The main reason for this is that the discussed mindshaping mechanisms and narrative control do not allow for the intentional control of mental states.

This may seem puzzling on first sight. Did we not just argue that the agent in the Velleman example cries intentionally, because crying expresses her feelings of sadness and her beliefs about the appropriateness of crying in the situation and did we not argue that her describing the situation in this way provides a powerful normative incentive for her to conform to these descriptions? Does that not mean that the agent by crying intentionally is intentionally reinforcing her emotions and beliefs? In other words, is this not an example of what we want to show not to be possible for purely narrative control, i.e. the intentional control of mental states?

It is natural to interpret the situation in this way and indeed it seems to us that many authors do at least sometimes speak as if examples like intentional crying are about the intentional control of mental states, but this interpretation is nevertheless quite wrong.

Obviously, in this situation the agent is acting intentionally and her actions do bring about a change in her mental states, but this does not mean that she intentionally brings about a change in her mental states. A deer might well intentionally turn its head to check whether the wolf it noticed earlier is still where it used to be and by doing so, might well bring about a change in its belief about the location of the wolf, but obviously this does not mean that the deer had any intentions concerning its mental states. The deer is concerned with wolves, not with beliefs about wolves. More generally, it is an almost trivial fact that intentional actions will normally not only have consequences in the outside world, but also in the agent's mind. That this is so does clearly not entail that any such actions are directed at bringing about a change in a mental state.

Now, the crying human is clearly a more complicated case than the head turning deer. The human is not concerned with updating facts about any old external object she happens to have beliefs about, as in the case of the deer, but instead by crying intentionally she expresses her belief that her behaviour is appropriate in the situation she finds herself in. She cries intentionally, because she thinks crying is the right

thing to do. She is aware of a norm and intentionally follows that norm. However, even though she intentionally follows the norm, she is not intentionally reinforcing her belief in the norm. She follows the norm, because she believes in its validity and her intentional following the norm leads to her believing it more strongly, but she is not intentionally doing anything to her belief. She is not interested in her own psychology and the way in which the norm she uses to guide her actions might be represented in her mind.

Now one could object here that even if it is true that the agent in the Velleman scenario is not intentionally controlling mental states, this does not mean that she has no control over her beliefs. The agent follows the norm, because she judges that the norm is valid. She believes that she has reasons for crying, that it is the right thing to do etc. She would not act the way she does, if she did not judge the way she does, and she only judges the way she does because she has deliberated about the world and has come to the conclusion that in her situation it is appropriate to do the thing she does. Clearly her deliberation controls the judgment she reaches. That seems entirely right, but still there is a crucial ambiguity here. It seems right to say that the agent controls her judgments because they are sensitive to her reasons and her judgments would instantly change if in deliberation it turned out that her reasons do not lead her anymore to the judgment she had made earlier. On the other hand though, it is not at all obvious that this form of control over judgments really is intentional control.

In this context it might be helpful to remind the reader of the debate about the nature of mental actions (see e.g. the excellent O'Brien 2009 collection) and the claim that one can never intentionally judge or intend anything. The argument for this claim is that our judgments and intentions etc. are not about what we want to judge or intend but about whether or not we take a proposition to be true or not in the case of judgments and what we take to be the best thing to do in the case of intentions. This evaluative part at the heart of any judgment etc. is not up to us (Hieronymi 2009, Mele 2009, Strawson 2002, Moran 2002).

If we find this argument convincing, then it is now no longer surprising that the control structures described so far can not achieve the intentional control of mental states, but it seems that this should not worry us too much. If it is impossible to intentionally control mental states in general, then it is not very surprising that narrative control is not able to do it either.

However, even if we do accept the argument that intentional judgments are impossible (as we think one should) this does not mean that one cannot intentionally control one's judgments. As Pamela Hieronymi (2009) has pointed out, there are two ways in which an agent can come to a judgment. They can either evaluate a proposition and that will lead automatically and unintentionally to a judgment, and this is the ordinary way which we discussed so far (Hieronymi calls this evaluative control); or they can intentionally manipulate their psychology in such a way that they will acquire a specific judgment (Hieronymi calls this managerial or manipulative control). The crucial difference between the two is that in the latter but not in the former case it is entirely up to the agent whether they will end up judging the proposition to be true or false.

As an example of manipulative control, Hieronymi cites the case of Pascal's wager. Pascal believes that it is better to believe in God than not for prudential reasons, but cannot see any evidence for his existence. As judgments are nothing else than the evaluation of the available evidence, Pascal finds it impossible to acquire his desired belief in God by normal means. It is simply not up to him whether he believes in God or not, but it depends on the evidence. As according to Pascal very little points to the existence of God, Pascal can't simply decide to believe in his existence anyway, because his judgments involuntarily follow the evidence. But Pascal really wants to believe in God, because he thinks that even though it is unlikely that God exists, the belief in him won't harm him, but in the unlikely case that God does exist the penalties for not believing are incredibly high. Pascal gets out of this bind by going to mass and praying rosaries, because as an astute observer of human psychology he notes that following the practice will bring it about that he will sooner or later start to believe. In other words, Pascal knows that it is possible to condition himself into believing in God, even though no rational deliberation would ever allow him to acquire that belief.

More generally, what the Pascal case shows is that even though it is impossible to control mental states intentionally by means of evaluation, it is still possible to intentionally bring it about that one acquires the desired mental states. One can achieve this if one directs attention not at the content of the state, but at the psychological state itself. These psychological states are then ordinary objects that one can intentionally manipulate. Obviously, though, this form of control can only be possible if one thinks of psychological states *as* psychological states.

This then is where the limits of narrative control lie. Narrative control is the ability to be governed by norms; it is not the ability to intentionally instantiate beliefs about norms (or anything else for that matter) in psychology. In order to be able to intentionally control one's psychology one needs to be able to have a psychological goal. One has to understand that there is such a thing as the psychological instantiation of norms and that these instantiations can be manipulated independent of the validity of the relevant norm. In other words, in order to exercise intentional control over one's mental states one needs to be a metarepresenter, i.e. someone who understands that there are mental states and that those states are representational. This is almost trivial; it seems difficult to understand how one could intentionally manipulate mental states, if one didn't know that there are such things.

How important is the intentional control of the mental?

Even if this argument were to show that manipulative control is the only way in which we can intentionally control our mental states, one might still wonder whether this is very important. It seems obvious that mindshaping and narrative control are centrally important for human self-regulation. If it is true that both mechanisms can work even though they do not contain any intentional control of mental states, then one might wonder why we should be worried about whether we have this ability or not.

The reason why this would be the wrong conclusion to reach comes down to one central issue: flexibility. The intentional manipulation of mental states allows for targeted psychological reconstruction and maintenance in a way that is impossible to reach with mindshaping.

Future directed self-control for example benefits enormously from metarepresentation. The simple reason for this is that a first-order deliberator will find it difficult to compute the fact that her norms will change over time or in certain contexts. Only an understanding of psychological states as states will allow an individual to understand that it is quite possible that she will not believe in a valid norm any longer if there are the right influences on her psychology. It is for example for a rationally deliberating person quite difficult to see why one might believe that it is permissible to drive a car while being very drunk. Only if one understands that drinking changes beliefs does it become understandable why one might have to guard

oneself against doing something, which with the current mindset one would not do under any circumstances.¹⁴

But is this intentional control not Cartesian?

We have argued that metarepresentation is important for self-regulation, because only metarepresenters can control their mental states intentionally. At this point it becomes legitimate to wonder whether the emphasis of the intentional control of the mental is not a serious limiting factor in terms of how social our understanding and control of our minds is.

In the first part of the paper we emphasized how the attribution of norms in our narrative practices leads to the shaping of our minds. It seems now as if the ability to intentionally shape one's mind gives individuals greater control to be masters of their own psychology and to achieve greater independence of societal norms. It is now up to the individual whether or not she wants to believe a certain proposition, acquire a set of intentions, desires etc.

It looks as if the ability to metarepresent has brought back the fully independent individual who decides of her own free will which norms she wants to endorse. This is a very tempting position to hold and it might seem as if we want to support the claim that the intentional control of the mental is some special capacity of a mysterious self or executive centre. This would not be very unusual either, after all the ability to intentionally shape one's mind has traditionally been strongly associated with human autonomy in the face of societal pressures, but it is not at all our position. In fact we would argue the very opposite. Intentional control of the mental does not diminish the importance of society for the individual, but instead enhances the influence of the social.

This is because the intentional acquisition of mental states will only get off the ground, if the subject evaluates that intentional control of the mental is the right thing

¹⁴ Our description of the drink driver is influenced strongly by Holton's (2009) account of weakness of the will. Holton argues that weakness of the will should be understood as reevaluation of what is the right thing to do. Like Holton we interpret the drink driver case as one, where the weak willed person under the influence of the situation changes her judgment about the permissibility of drink driving. We find Holton's account of weakness of the will compelling, but our point would equally be compatible with an account where what the agent needs to manipulate is not their belief (because one might think that this will normally remain constant in a weakness of the will scenario), but their motivational states (because otherwise their desire to drink drive might simply overpower their belief that this is the wrong thing to do, all things considered). Like in the belief case, only if one knows about one's motivational states will one be able to control them effectively. Thanks to Fabio Paglieri for pointing this out.

to do. Self control for example happens because the agent evaluatively forms the belief that she has to do something intentionally to her mind now in order to prevent herself from doing something in the future which is not in line with what she considers to be the right thing to do. This evaluation itself is clearly not, as Hieronymi correctly observes, done intentionally, but evaluatively. Obviously though, our evaluations are socially shaped in the many subtle ways described in the first half of the paper. So, when we intentionally manipulate our mentality, we do it on the basis of evaluations that have been shaped by social context. We just do it more effectively now. If we find the content of a specific norm not convincing, but have been shaped by social forces to value norms not only because of their contents, but also as norms, then we are now in a position to remake our psychology in such a way that we will come to accept the norm, which we could not have accepted otherwise.¹⁵

Conclusion

The ability to represent representations as such or metarepresentation is at the heart of the theory theory of mindreading. In the last couple of years, mindshaping approaches have developed powerful alternative models that seem to show that our ability to understand each other could not be due to some theory, but due to our embeddedness in normative narrative practices. These practices have been used as well to understand how we regulate our own minds. They control our behaviour and shape our minds, because we strive to fulfill our normative self-interpretations. However there is an important limit to what such narrative control can achieve. On a closer look it turns out that narrative control cannot achieve the intentional control of mental states. This type of control does require metarepresentation. It was shown that this form of control matters, because it enables a greater degree of flexibility in e.g. future directed self-control. Finally this greater degree of flexibility does not lead us back to a Cartesian homunculus, but simply adds another layer of control to our fundamentally social mind.

ACKNOWLEDGMENTS

¹⁵ Obviously, having this ability does not mean that it always will successfully achieve its goal. In addition to the fact discussed in the text that all intentional control is dependent on the involuntary formation of the intention to manipulate, it is obviously true as well that many attempts to intentionally manipulate one's mind go wrong. Thanks to Fabio Paglieri for pointing this out.

This paper would not have been written had it not had the support from the AHRC, under the ESF Eurocores CNCC scheme, for the CONTACT (Consciousness in Interaction) project, AH/E511139/1.

REFERENCES

- Bruner, J. (1990). *Acts of Meaning*. Cambridge: Harvard University Press.
- Davies, M. and Stone, T. (2000). Simulation Theory. *Routledge Encyclopaedia of Philosophy Online*.
- Davies, M. and Stone, T. (1996). The Mental Simulation Debate: A Progress Report. In P. Carruthers and P.K. Smith (Eds.), *Theories of Theories of Mind* (pp. 119-137). Cambridge: Cambridge University Press.
- Dennett, D.C. (2003). *Freedom Evolves*. New York: Penguin Books.
- Dennett, D.C. (1998). Reflections on Language and Mind. In Carruthers, P., & Boucher, J. (Eds.), *Language and Thought: Interdisciplinary Themes*. Cambridge: Cambridge University Press.
- Dennett, D.C. (1996). *Kinds of Minds*. New York: Basic Books.
- Dennett, D.C. (1993). *Consciousness Explained*. London: Penguin.
- Dennett, D.C. (1992). The Self as a Center of Narrative Gravity. In F.Kessel, P.Cole and D.Johnson (Eds.), *Self and Consciousness: Multiple Perspectives*, Hillsdale, NJ: Erlbaum.
- Hieronymi, P. (2009). Two Kinds of Agency. In O'Brien, L. & Soteriou, M. (Eds.), *Mental Actions* (pp. 138-162). Oxford: Oxford University Press.
- Holton, R. (2009). *Willing, wanting, waiting*. Oxford: Oxford University Press.
- Hutto, D. (2008). *Folk Psychological Narratives: The sociocultural basis of understanding reasons*. Cambridge MA: MIT Press.
- Hutto, D. (2007). Folk-Psychology without Theory or Simulation. In D. Hutto, M. Ratcliffe (Eds.) *Folk Psychology Re-Assessed* (pp.115-135). Dordrecht: Springer.
- Hutto, D. (2004). The Limits of Spectatorial Folk Psychology. *Mind and Language* 19(5), 548-573.
- McGeer, V. (2007). The Regulative Dimension of Folk Psychology. In D.Hutto and M.Ratcliffe (Eds.), *Folk Psychology Re-Assessed* (pp. 138-156). Dordrecht: Springer.
- McGeer, V. (1996). Is Self-Knowledge an Empirical Problem? Renegotiating the Space of Philosophical Explanation. *Journal of Philosophy*, 93, 483-515.
- Mele, A. (2009). Mental Action: A Case Study. In O'Brien, L. & Soteriou, M. (Eds.), *Mental Actions* (pp. 17-38).Oxford: Oxford University Press.
- Moran, R. (2001). *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press.
- Strawson, G. (2003). Mental Ballistics or The Involuntariness of Spontaneity. *Proceedings of the Aristotelian Society*, 103(3), 227-256.
- Velleman, J.D. (2009). *How We Get Along*. Cambridge: Cambridge University Press.
- Vierkant, T (forthcoming) What Metarepresentation is for. In J. Brandl, J. Perner, J. Proust (Eds.), *Metacognition*. Oxford: Oxford University Press. Scheduled for 2012.
- Zawidski, T. (2008). The Function of Folk Psychology: Mind Reading or Mind Shaping? *Philosophical*

Explorations, 11(3), 193 – 210.