Routledge
Taylor & Francis Group

WINNER OF THE PHILOSOPHICAL EXPLORATIONS ESSAY PRIZE 2013

The relations between agency, identification, and alienation

Alec Hinshelwood*

Department of Philosophy, University College London, London, UK

This paper examines the relations between, on the one hand, accounts of the distinction between an agent's identifying with, as opposed to feeling alienated from, their attitudes; and on the other, metaphysical accounts of action. It claims that a commitment to an event-causal conception of agency, which would analyse agency in terms of the causal potency of psychological states and events, appears to render mandatory a particular style of account of identification and alienation – namely, the hierarchical model offered by Harry Frankfurt and Michael Bratman. It is argued that such accounts fall foul of a dilemma: the Authority Problem. The failure of attempts to avoid the Authority Problem is then used to motivate an attractive alternative style of account of the distinction, offered by Richard Moran. However, it is pressed that Moran's account rests on claims about agency which seem incompatible with the event-causal conception of agency. By making the links between the metaphysics of agency and accounts of identification and alienation more explicit, the paper allows us to better comprehend both the apparent need for and characteristic failures of some traditional accounts of identification and alienation, as well as make clear the action-theoretical debt incurred by those who would offer an alternative.

Keywords: agency; identification; alienation; Harry Frankfurt; Michael Bratman; Richard Moran

1. Introduction

It is possible for a person to harbour certain attitudes or experience certain emotions while yet feeling as if they are not, in some sense, really theirs. I shall refer to this phenomenon as *alienation*. Given that alienated mental states occur within a person's psychic life, there is a weak sense in which they are all states of theirs; nevertheless, there is a stronger sense of ownership that it is possible for a person to enjoy with respect to their mental life. Following Harry Frankfurt, I shall say that a person enjoys the latter kind of ownership over their attitudes if they are attitudes with which they *identify*.[1] It is this kind of ownership which is absent when an agent feels alienated from an attitude.

For the most part philosophers have discussed the notions of identification and alienation in connection with issues in moral psychology, for instance, in connection with questions about character and identity, as well as questions about moral responsibility. Whilst I do not challenge the idea that there might be important links between the aforementioned topics and the distinction between identification and alienation, in this paper I want to focus on the

*Email: a.hinshelwood.12@ucl.ac.uk

distinction itself – treating those other topics as downstream of my own. Moreover, my central contention is that we must re-assess the metaphysics of agency which is *upstream* of the distinction between identification and alienation. By making the links between the metaphysics of agency and accounts of identification and alienation more explicit, I hope to better comprehend both the apparent need for and characteristic failures of some traditional accounts of identification and alienation, as well as make clear the action-theoretical debt incurred by those who would offer an alternative. So, on the one hand, I claim that a commitment to an event-causal conception of agency, which would analyse agency in terms of the causal potency of psychological states and events, appears to render mandatory a style of account of identification and alienation which is beset by insoluble difficulties: the hierarchical model offered by Harry Frankfurt and Michael Bratman.[2] On the other, I claim that an attractive alternative style of account of the distinction – that of Richard Moran – rests on claims about agency which seem incompatible with the event-causal conception of agency. Whilst I favour the Moranian view, whether it can be fully defended and a non-event-causal conception of agency developed I leave for another occasion.

In Section 2, I give an example of alienation to get a fix on the distinction, and then in Sections 3–4 I introduce the hierarchical accounts of Bratman and Frankfurt. I argue that both accounts fall foul of what I call the Authority Problem, and in Section 5 I trace the need for a hierarchical account to the event-causal conception of agency. Rather than arguing against the event-causal conception of agency, however, I explore in Section 6 whether a modification of a later view of Frankfurt's might provide an account which is compatible with the event-causal view of agency, but which avoids the Authority Problem. However, difficulties with this modified view then motivate, in Section 7, an appeal to Moran's account of the distinction between identification and alienation, which I claim is incompatible with the event-causal conception of agency. I close in Section 8 by discussing the broader uses to which accounts of identification and alienation have been put by philosophers of action.

2. The distinction between identification and alienation

We can imagine a man who has received a strict religious upbringing and who later goes on to live in a monastery. Say that this person enters into a particularly severe order which encourages self-flagellation in response to thoughts or feelings of a sexual nature, but later becomes disillusioned and leaves the monastery. Now this person no longer believes in God, and neither does he maintain his previous evaluative outlook which condemned his erotic desires. Nevertheless, we can imagine that he still wants to hurt himself upon entertaining an erotic desire, even though he no longer considers there to be any reason for him to condemn them. When he experiences erotic desires he has to undertake various strategies to avoid or assuage his desires for self-harm. Perhaps the man knows that if he tries to solve difficult mathematical problems in his head, he can distract himself from his desires to self-harm and so prevent their occurring. In this case, the lapsed believer may experience his desire to hurt himself as an "alien intruder", or a force which moves him to act "in spite of himself", as we might say. There are mental goings on within his psychic life that – as both we and the sufferer are inclined to put it – are not his.

Important here is the man's strategic relation to his alienated desires. They feature as one more part of the world with which he has to contend, or perhaps better, as the demands of another to which he must listen but which do not seem legitimate to him. His wanting to hurt himself when he entertains erotic thoughts thus figures, in a particular way, as a fact about himself he has to take into account when considering what to do. The man may avert his eyes when walking past attractive people on the street, not because he

thinks finding people attractive is bad, but because the onset of his self-punitive desires may inflict costs on him he does not want to pay. He knows that once he starts desiring to hurt himself he will suffer the desire's force, and his desire then provides him with reasons to act because an instrumental value attaches to the avoidance of the desire. If this man *does* deliberate about whether to hurt himself, it is not because it is, for him, something he thinks it is good to do. Perhaps if he does not scratch himself a little he knows his desire will grow to quite unmanageable proportions, whereas if he nips it in the bud in this way it will disappear. Here scratching himself figures instrumentally as a means by which he can forestall the onslaught of more troubling desires. The man is thus forced to deliberate about the avoidance or satisfaction of his alienated desires in a way which renders them similar to facts about his height or even his external surroundings. His alienated desires constrain the projects he might sensibly undertake, or figure as obstacles which must be strategically overcome. Because the attitudes from which the man is alienated figure in his deliberations in much the same way as straightforwardly external matters, the idea that they are beyond his control seems compelling.[3]

This man's self-punitive desires provide an example of attitudes from which an agent is alienated. In the first instance, I understand identification simply as the contrasting absence of any such phenomenology. There is some pre-theoretical reason for this. The language of alienation derives from ordinary language and how people naturally describe such experiences, whereas the notion of identification in ordinary language does not necessarily pick out the contrary of alienation. We can speak of a person's identifying with another person, an institution or a fictional character, but I take it that when Frankfurt introduced the term identification he meant to distinguish a peculiar sense of it which picks out a possible way in which an agent might relate to their mental life. The only pre-theoretical handle we have on this technical sense of identification is just whatever we mean when we think someone is not alienated from an attitude.

3. Frankfurt's first account and the Authority Problem

Frankfurt's initial account of the distinction between alienation and identification came in his paper "Freedom of the Will and the Concept of a Person". The keystone of Frankfurt's analysis is the concept of a second-order desire:

> Someone has a desire of the second-order either when he wants simply to have a certain desire or when he wants a certain desire to [move him to act]. In situations of the latter kind, I shall call his second-order desires "second-order volitions" or "volitions of the second-order". (1988a, 16)

Frankfurt goes on:

> [A person] identifies himself […] through the formation of a second-order volition, with one rather than the other of his […] first-order desires. He makes one of them more truly his own and, in so doing, he withdraws himself from the other. (1988a, 18)

So, to identify with a desire is for it to be the object of a second-order volition. Frankfurt is less explicit about alienation, but I take his thought to be that an agent is alienated from their desire to do something insofar as they want *not* to want to do that thing. Accounts such as this have become known as hierarchical accounts because they explain identification and alienation in terms of hierarchies of attitudes.

Frankfurt's initial account faces an insuperable difficulty, however, which was initially pointed out by Gary Watson.[4] The problem is this: second-order volitions are just desires

with desires as objects; in principle then, a person could fail to identify with them. Why then should they be privileged in "speaking for the agent", as it were, and constitute *the agent's* identifying with their psychic life as opposed to being simply more elements within it? It seems as if the mere presence of a second-order desire cannot be sufficient for an agent's identifying with an attitude, for it could "speak for the agent" only if the agent identified with *it*, as is the case with first-order desires. This means that in order for a second-order attitude to constitute an agent's identifying with a lower order attitude, according to the hierarchical account of identification, it must be the object of a still higher order attitude. But then a vicious regress ensues because an agent's identifying with an attitude presupposes an infinite hierarchy of higher order attitudes; identification then seems impossible.

This problem, which I call the Authority Problem, actually comes in the form of a dilemma. We may ask: in virtue of what does a particular psychic item represent *the agent*, and so constitute their identifying with an object attitude? As we might put it, what confers authority on psychic items of that sort?[5] The problem arises if the psychic item in question is one from which an agent could be alienated. For then either the item in question must be identified with in order that it represent the agent, and so explicate how the object attitude of that item is itself something the agent identifies with, or it need not be. If the former horn is taken, the account of identification is landed in a regress: in order for any given instance of the relevant psychic item to constitute an agent's identifying with an attitude, an antecedent item of the relevant sort's being related to it will be presupposed, and so on *ad infinitum*. If the second horn is taken, however, it will be unclear how the item can account for an agent's identifying with the object attitude. It will seem arbitrary to insist that *these* items "speak for the agent" if they may do so even if the agent does not identify with them. For if such items can "speak for the agent" even whilst the agent does not identify with them, then on what grounds do the *first*-order attitudes require being identified with in order to count as "the agent's own"? In failing to insist that these states must be identified with, this horn of the dilemma undercuts the idea that psychic items "speak for the agent" only if the agent identifies with them. For it allows that states with which the agent is not identified may well "speak for the agent" too.

4. Bratman's account

Bratman's account of identification, like Frankfurt's, is hierarchical: he seeks to elucidate the idea of an agent's identifying with an attitude in terms of some higher order attitude which takes the first as an object. The nub of Bratman's proposed solution to the Authority Problem is the claim that the psychic items constitutive of an agent's identification with an object attitude count as such because they constitute, in part, the agent's literal identity over time. These items "speak for the agent" because they contribute to an agent's persistence over time, given a broadly Lockean understanding of personal identity.

The relevant psychic items to which Bratman appeals are what he calls "plans and policies", which he says "induce organization and coordination by way of continuities and connections [amongst one's psychic life]".[6] So, he claims, they are well placed to play a constituting role in maintaining the kinds of psychic ties which create a Lockean identity of person over time. Some examples he cites are "a policy of developing and supporting a strong concern with honesty in writing, or of trying to be more willing to be playful or less inclined to be impatient with others", and he calls these "self-governing" policies.[7]

Of course, Bratman has a well-developed account of what is involved in a first-order attitude's figuring as the object of a self-governing policy. His idea is that self-governing policies support the to-be-identified-with desires as functioning in a particular way. The way he suggests is that of "end-setting", where this is characterized thus:

> [A] desire for *E functions as end-setting for practical reasoning* when that desire motivates by way of a process of practical reasoning that appeals to *E* as justifying end. (2002, 75)[8]

Bratman claims that it is possible for a desire to function as end-setting without an agent's identifying with that desire, but he insists that where a desire's so functioning is the object a self-governing policy, the agent will identify with it.[9] As with Frankfurt, Bratman is less explicit about alienation, but I understand him to take an analogous position: an agent is alienated from an attitude insofar as they have a higher policy in support of its *not* functioning as end-setting in practical reasoning.[10]

This account of identification fails to avoid the Authority Problem, however. The question to ask is whether or not one could be alienated from one's self-governing policy in support of a given desire's functioning as end-setting. To avoid the problem, Bratman needs to show how one could not be alienated from one's self-governing policies. But even if self-governing policies can constitute an agent's identifying with an object attitude only if they contribute towards the agent's literal identity, it does not change the fact that the agent might still be alienated from them. Bratman has guaranteed only that they must contribute towards the agent's identity (in his sense), not that the agent identifies with them. So for example, say that I have a desire not to argue with my friends on any occasion: I am very meek. If Bratman is right, then I can appeal in deliberation to thoughts about not arguing with my friends as justifying whilst it remain possible that I fail to identify with the desire to do so. Perhaps upon reflection I see that they are taking advantage of me and generally not respecting me. If so much is possible, though, I do not see how adding a higher order policy in favour of the desire would amount to my identifying with it; even if it (allegedly) contributes towards my literal identity over time. The higher order policy in support of the desire may just be entrenched due to years of my being a meek person, such that it has become formative of my character (in some sense) to treat not arguing with my friends as end-setting in my practical deliberations.[11] Might I not then feel estranged from my policy to always support that desire? It seems possible that I could suffer from that policy as forever condemning me to unhealthy and exploitative friendships. The fact that it contributes to my Lockean identity over time does not seem to guarantee of itself that I shall identify with the policy. The only sense in which self-governing policies are inalienable, then, fails to speak to the concerns of the Authority Problem: that sense is the metaphysical one such that they (allegedly) contribute towards my literal identity, but that is consistent with an agent's feeling alienated from them.

Perhaps Bratman would insist that without identifying with a policy it could not constitute one's literal identity. But the foregoing example surely shows how the kind of dispositions and character traits which could constitute Lockean identity, if any could, are such that one might feel alienated from them. But in any case, if Bratman is helping himself to the idea that identification is partly constitutive of Lockean identity, then we cannot very well appeal to the fact that self-governing policies constitute Lockean identity in trying to avoid the Authority Problem. If we want to avoid the Authority Problem by appealing to items which constitute one's literal identity, but then specify that such items can do so only if one identifies with them, then we shall not be any closer to providing an analysis of identification.

5.   The link between the event-causal conception of agency and hierarchical accounts of identification and alienation

I think we should suspect that the Authority Problem is insoluble as far as hierarchical accounts of identification go. One might, of course, hold out hope that the needed inalienable psychic item could be found. But better, I think, would be to raise the question as to why it seems mandatory to explicate an agent's relation to their psychic life – as manifested in the phenomena of alienation and identification – in terms of the presence of higher order psychic states which "speak for them".

To bring out one line of reasoning which arguably leads to this conclusion, it is worth returning to a step of Bratman's account which I suppressed in the foregoing.[12] Now, it looks as if we only have two options when attempting to explicate identification and alienation: *either* we appeal to the activity of the agent directly – perhaps as undertaking special acts of endorsement and rejection with respect to their mental life; perhaps in some more nuanced way – *or* we analyse the phenomena in terms of the obtaining of certain psychic states and the relations between them.[13] Bratman, before settling on the notion of self-governing policies in support of a desire's functioning as end-setting, offers this account:

> [W]e should understand an agent's endorsement of a desire in terms, roughly, of a self-governing policy in favour of the agent's treatment of that desire as providing a justifying reason in motivationally efficacious practical reasoning. (2000, 54)

But Bratman worries that appealing to "an agent's treating something as a reason" in explicating identification unhelpfully presupposes the idea of identification itself.[14] This stems from his commitments in the philosophy of action. In order for us to be able to think of some process or occurrence as an *agent's doing something*, as opposed to merely happening in or to them, Bratman thinks it must be one with which the agent identifies. Thus, he writes:

> An agent moved by desires of which he is unaware, or on which is he incapable of reflecting, or from whose role in action he is, as we sometimes say, estranged, seems himself less the source of the activity than a locus of forces. (2001, 312)

He makes the same distinction again this way:

> When a person acts because of what she desires, or intends, or the like, we sometimes do not want to say simply that the pro-attitude leads to action. In some cases we suppose, further, that the *agent* is the source of, determines, directs, governs the action and is not merely the locus of a series of happenings, of causal pushes and pulls. (2001, 311)

Bratman's thought is that what makes the difference between a mere "series of happenings" and an instance of agency is that the agent identifies with the motivations which produce the event in which the action consists.[15] This means that to employ the idea of an *agent's* treating something as a reason when explicating identification unhelpfully presupposes the idea. As Bratman sees it, whatever process is involved in treating something as a reason, in order for it to be an activity on the part of the agent it must be one with which the agent identifies. We cannot then *analyse* identification in terms of an agent's treating something as a reason.[16]

In the background here is the idea that agency must consist in the causal links between states and events.[17] An action is an event – an arm's rising, say – which is suitably caused by some preceding mental occurrences (typically a belief and a desire); likewise, some process constitutes someone's thinking something if it is suitably produced by some

preceding mental goings on. Bratman's contribution here is the idea that only events caused by psychic states with which the agent identifies are such as to count as instances of their agency. In defence of this, one might add that if an action *just is* the causing of an event by some preceding mental occurrences, then arguably the causing of an event by mental occurrences from which someone is alienated could not constitute an action. If we take the language of suffering and passivity with respect to alienated mental states seriously, then it seems there is no room left for the agent when their movements are caused by alienated states.

Let me summarize the two important claims from the foregoing this way:

> The Event-causal Claim: the only viable account of agency is one couched in terms of causal links between states and events.
> The Identification Claim: (i) psychological states and events belong to the agent properly speaking only if the agent identifies with them and (ii) The causation of events by psychological states and events constitutes agency properly so-called only if the agent identifies with the causally efficacious states and events.

It is these claims which make an account of identification and alienation in terms of psychic states and their relations seem mandatory. The Event-causal Claim pushes one towards the Identification Claim, which then precludes one's accounting for identification in terms of the agent's doing anything on pain of circularity. Any residual mention of the agent in one's account of identification involves helping oneself to the idea unanalysed, so the only illuminating account of identification available is one in terms of psychic states and the relations between them. Thus, instead of talking about the agent's treating their desires as reason-giving, Bratman claims that we must understand identification in terms of higher order policies in favour of desires' functioning in a certain way.

Now, one might attempt to resist the move from the Event-causal Claim to the Identification Claim.[18] Indeed, the Identification Claim seems to render the idea of alienated agency an oxymoron, which is arguably too strong a conclusion. But even if one rejected the Identification Claim, it still seems as if the Event-causal Claim on its own would mandate an account of identification couched solely in terms of psychic states. If one accepts that agency must consist in the causal links between states and events, then it does not seem available to one to claim that the mental phenomena which cause actions, when things go well, are either the objects or manifestations of some unexplained piece of agential activity. For an adherent of an event-causal conception of agency, then, only an account of identification and alienation in terms of psychic states and their relations is available.

6. A modified version of Frankfurt's later account

In the light of the Authority Problem, and in opposition to the hierarchical model, one might wish to make room for an account of the distinction between identification and alienation which makes a direct appeal to some conception of agential activity. And an argument of the following form might present itself. If we accept an event-causal conception of agency, then we must provide a hierarchical account of the distinction between alienation and identification. However, as the latter cannot be provided, as *per* the Authority Problem, we may argue by *modus tollens* for an alternative conception of agency.

Now, putting to one side whether there might not be other, moral-psychologically independent grounds for denying an event-causal conception of agency, the proposed conclusion is surely arrived at too quickly. For the conclusion of the last section was only

that, given an event-causal conception of agency, an account of the distinction between alienation and identification must be provided in terms of psychic states and the relations between them. And it might reasonably be thought that there is some leeway here, between that idea and the more fully articulated idea of a *hierarchy* of psychic states. But if there is some purely psychic account of the distinction available, compatible with an event-causal conception of agency and yet non-hierarchical, then one will not be able to use the Authority Problem as lever to dislodge the Event-causal Claim.

An attempt in this vein might be made by modifying a later account of Frankfurt's, in which he proposed a way of avoiding the Authority Problem that appealed to the notion of "satisfaction". Frankfurt claims that being satisfied with one's attitudes requires "no adoption of any cognitive, attitudinal, affective, or intentional stance". Nor does it

> Require the performance of a particular act; and it does not require any deliberate abstention. Satisfaction is a state of the entire psychic system – a state constituted just by the absence of any tendency or inclination to alter its condition. (1999, 104)

"Being satisfied", Frankfurt said, "is a matter of simply *having no interest* in making changes",[19] although it must not be "unreflective"; satisfaction "develops and prevails as an unmanaged consequence of the person's appreciation of his psychic condition."[20]

Frankfurt then claims that identification "is constituted neatly by an endorsing higher-order desire with which the person is satisfied."[21] So, satisfaction with a psychic item at least seems to entail the absence of some still higher order, countervailing attitude. If that is right, then identifying with a desire would not generate the regress horn of the Authority Problem: one could identify with a desire just by its being the object of a second-order desire concerning which one had no still higher order attitudes.

Now, there are two issues I want to raise for this account. One concerns the fact that it still insists that identification with an attitude depends upon a higher order attitude with the first as its object; considering this will lead to the modified view which avoids the Authority Problem without challenging the Event-causal Claim. The second issue concerns the notion of satisfaction itself; considering this will lead me to introduce a view which once again threatens the Event-causal Claim.

To begin with the first: if I can be satisfied with a desire without any higher order desire for it, but identification with an attitude consists in its being the object of a higher order desire with which I am satisfied, then Frankfurt is insisting that I do not need to identify with my higher order desires in order for them to constitute my identifying with first-order attitudes. But that forces Frankfurt to say that I may not identify with those attitudes which are to constitute my identifying with my first-order attitudes - however we are to understand satisfaction. But if one can fail to identify with these higher order attitudes, then can we insist that such attitudes are better placed than first-order attitudes to "speak for the agent"? We seem to be back onto the second horn of the Authority Problem.

Of course, it may be that if an agent is satisfied with an attitude, then it can "speak for them" – regardless of whether the agent identifies with it. But then with satisfaction in play, one need not insist on cashing out identification in terms of higher order attitudes at all. For why not claim instead that there is generally a presumption in favour of an agent's satisfaction with their first-order attitudes, identification consists in that, and therefore that identification is simply the "normal case" for first-order attitudes? Frankfurt himself later said

> Being identified with the contents of one's own mind is a very elementary arrangement. [...] It is so natural to us, and as a rule it comes about so effortlessly, that we generally do not notice it at all. In very large measure, it is simply the default condition. (2006, 8)

Such an account would concede to the second horn of the Authority Problem that higher order attitudes are not privileged in "speaking for the agent". But hierarchies could then enter as explaining alienation – desires in favour of *not* being moved by some desire, say – where there is, once again, simply a defeasible presumption in favour our being satisfied with, and so identifying with, those higher order negative attitudes. Frankfurt would not then have to say that an agent need not, by the theory's own lights, identify with the attitudes which constitute their identifying with lower order attitudes: here there are no higher order attitudes which are so constitutive, and the higher order attitudes which constitute an agent's being alienated from lower order attitudes are not dependent on still higher order attitudes for the agent's identifying with them.

What are the implications of this account? Importantly, the search for the inalienable psychic item needed to forestall the Authority Problem would not be necessary. One *can* be alienated from one's desires; but there is a standing, though defeasible, presumption in favour of our identifying with them – a presumption extended to higher order attitudes when they are invoked to explain alienation. In order to challenge this account – or others with its shape – it seems one would have to examine, case-by-case, whether the psychic items offered as explaining alienation really are necessary and sufficient for that. There would no longer be a single, apparently knock-down argument against psychic hierarchies, as *per* the Authority Problem.

So far so good, it might be thought: the Authority Problem has been avoided and we have not yet seen grounds for moving beyond psychic states and their connections; we have not, therefore, found a reason for rejecting an event-causal conception of agency. At this point, however, I think we have to examine the notion of satisfaction itself: my second concern. Frankfurt insists that satisfaction is "neither deliberately contrived nor wantonly unselfconscious",[22] but it is far from clear what this amounts to, nor, and importantly, why we should presume something like it to hold for our attitudes as the "normal case". For we might worry with Bratman that one might have no countervailing attitudes towards one's attitudes due to "enervation or exhaustion or depression",[23] and that seems to undermine the thought that mere satisfaction, for all Frankfurt tells us about it, could be sufficient for an agent's identifying with an attitude. The conditions for satisfaction, as merely negative, might be met in ways which intuitively are incompatible with identification.

Given the conception of desire which is operative in the background here, it should not be surprising that satisfaction appears not to suffice for identification. Indeed, with such a conception, it is hard to see how identification could be the norm. So Frankfurt writes:

> Our most elementary desires come to us as urges or impulses; we are moved by them, but they do not as such affect our thinking at all. They are merely psychic raw material. A desire provides us not with a reason but with a problem – the problem of how to respond to it. (2002a, 184)

Desires here are conceived as a kind of brute psychic given, the agent's relation to which is left undecided by the mere presence of the desire itself. So conceived, it is difficult to see how the mere presence of a collection of desires within an agent will, just as such, suffice for the agent's identifying with them – nothing about that rules out failures of identification, as Bratman worries. But then it is difficult to see how it could simply be the normal case that an agent identifies with their first-order attitudes. Surely something further is needed, beyond the mere presence of the attitudes, in order to secure the idea that an agent identifies with those attitudes just where they have no countervailing higher order attitudes against them?

Perhaps more could be said in favour of the satisfaction proposal, but clearly there can be no recourse at this stage to higher order attitudes to provide the needed extra – we would be back with the first horn of the Authority Problem. Now, a presumption in favour of an agent's identifying with their first-order attitudes is certainly an attractive idea, so it would be natural at this point to challenge Frankfurt's conception of desire. It seems to me, then, that Richard Wollheim is right when he urges:

> [Our desires] do not present themselves to us, their owners, as simply being what they are for, and all – innocently, one might say – begging for satisfaction as vociferously as their strength determines. (1993, 59)

Rather, it is somehow internal to desires that we identify with them – though of course that can fail. So we should not think of identification with a desire as something extra, conferred upon it in virtue of its relations to other attitudes; we should find a way of conceiving desires which builds in an agent's identifying with them, at least ordinarily. It is here that I suggest the work of Richard Moran is helpful. As I read him, he understands desires in terms of a basic a kind of activity inherent in holding them, which breaks down in alienated cases. The nature of desire, cashed out in agential terms, secures the presumption in favour of an agent's identifying with their first-order desires. But then this appeal to an unreduced conception of agential activity jeopardizes, once again, the event-causal conception of agency.[24]

7. Moran's account

Moran argues that when an agent's attitudes are responsive to their perception of the reasons for those attitudes we can then see the agent as actively holding those attitudes and that the agent thereby identifies with them.[25] Here the agent's attitudes are, in a sense which obviously calls for greater elucidation, dependent on their engagement with their reasons. By contrast, if an agent harbours an attitude for which they can see no rationale and which persists even if the face of the agent's taking their situation to not warrant it, then the agent is alienated from that attitude. In such a situation, the agent's attitude can no longer be taken as a response on their part, but has rather become an obstacle, as it were, with which they must contend. The crucial point is that an agent identifies with a desire insofar as they see *some* value in what the desire is for and the agent's holding the desire is responsive to their sense of its being worth holding.

So, for example, when I think I see a dear friend across the street and want to greet them, usually I identify with the desire. In such a case, the desire would persist just insofar as I took the approaching person to actually be my friend, and so warrant my wanting to greet them. If it turned out, in fact, that I had made a mistake and that I did not know the approaching person, then I would cease, *ceteris paribus*, wanting to speak with them. Were I to persist in wanting to greet the stranger, however, even when viewed as such and in the absence of my taking there to be any good in talking to them, I would feel alienated from that desire. Likewise, the lapsed believer whom I used to introduce the notion of alienation suffers from attitudes for which he cannot find a rationale, even if he can offer some explanation of them in terms of his psychic history. No reflection on the qualities of his wife, and no reflection on the baselessness of his self-punitive desires make any difference to whether or not he wishes to harm himself.[26]

This picture does not involve seeing the agent as undertaking special acts of endorsement and rejection with respect to their attitudes, but it arguably does require us to

understand our attitudes as *themselves* manifestations of agency - exercises of our rational capacities. As I read Moran, we identify with attitudes *not* in virtue of some preceding act; nor because they are either the object of or caused by some further mental state, but rather because the attitudes with which we identify are themselves manifestations of our rational agency. With desires, and in the good case, we should think of them as instances of an agent's *holding* some end as worth achieving or to-be-done.[27] On this picture, our attitudes are themselves understood as exercises of our rational capacities when we identify with them; and the reasoning agent and their capacities are taken as basic in accounting for their activity. This then sits at odds with the event-causal conception of agency, for that conception of agency refuses to take as basic the agent, their powers and then to see certain states or events as the exercises of those powers. Such a conception would not rest content with the idea that attitudes themselves can be active: whether attitudes enjoy that status would have to depend on their being related in the right way to further states and events. Within such a picture, an agent's identifying with an attitude is factorized into an attitude plus its relation to something further; as I understand it, the Moranian alternative refuses such factorization.

How does this relate to the presumption in favour of identification which I mentioned in the last section? The idea of identification as responsiveness to reasons takes as the base, or core case that of an agent's identifying with their attitudes. Rather than starting with some first-order attitudes and trying to locate the needed extra in order to "build up" to a case of identification – as hierarchical theorists do – this picture sees identification as part of the typical well-functioning of first-order attitudes and then "builds down" to alienation, by saying what is missing in those cases. The norm is reason-responsiveness; that is what alienated cases lack. This picture is available once we allow the rational agent to take centre stage and recognize their role in the formation and maintenance of first-order attitudes. It is because first-order attitudes are typically manifestations of the agent's activity that the agent identifies with them; here the activity involved in an agent's holding attitudes *is* the agent's responding to reasons, which *is* the agent's identifying with those attitudes. In the normal case then, we do not need to add something to first-order attitudes to get to identification; it is when the activity constitutive of the normal case breaks down that something is missing.

From this point of view, we can diagnose why hierarchical theorists have to "build up" to identification, as well as explain the theories' characteristic flaw: susceptibility to the Authority Problem. Given their commitments in the philosophy of action, principally to the event-causal picture, hierarchical theorists have to view the appeal to the agent's role with respect to first-order attitudes with suspicion. In denying the agent's role with respect to first-order attitudes, then, these theorists have to go in search of something extra which will constitute the agent's relation to their psychic life, as that is manifested in identification and alienation. This is what makes a presumption in favour of identification hard to understand on such a view. For our first-order attitudes then appear as if they formed a self-standing array of psychic objects, the agent's relation to which must – on pain of appealing to special acts of identification and rejection – be captured in terms of further psychic items. Talk about the agent's activity with respect to attitudes comes to seem as if it must be translated into talk about further, higher order attitudes. Higher order attitudes then need be inalienable to play the agent's role, and the Authority Problem ensues.

On the view I favour, the agent's relating to their psychic life is secured as soon as we have an agent who responds to reasons. On this model, there is no need to think of the agent as an inner judge, endorsing or rejecting psychic items as they come up for review. The

254 Alec Hinshelwood

agent's eye is, in the good case, directed towards the world; and in virtue of the agent's attitudes' responsiveness to reasons, the agent's consideration of their reasons is *ipso facto* an assessment of their attitudes. For the agent's forming, maintaining or revising their attitudes simply follows from the agent's engagement with their reasons, which engagement is the agent's relating to their psychic life in the form of identification.[28] It is only when things have gone wrong that the agent's attitudes must be approached as self-standing psychic objects, about which they have to do something.

Finally, it is worth noting that on this account alienation from a desire involves the agent's failing to find any worth in the desire's object. Some explicit thought will therefore be involved in a way which is not mirrored in an agent's identifying with a desire, which an agent may hold in response to reasons without any explicit process of reasoning on their part.[29] This asymmetry reflects the satisfaction account's invocation of higher order desires to explain alienation without, however, necessarily appealing to such states.[30]

8. The role of the distinction between identification and alienation within the philosophy of action

Before concluding it will be helpful to consider a fairly broad objection to the foregoing remarks, in order to further bring out the general thrust of the Moranian picture. Some hierarchical theorists have hoped that the notion of an agent's identifying with her attitudes can be used to mark differences between degrees of agency. Perhaps: there are mere happenings, some kinds of sub-intentional activity – arguably instanced by animals and even humans when at their more thoughtless or alienated – and finally there is full-blown intentional agency. (Perhaps, too, there is some still higher grade: some kind of perfectly moral agency.) The idea of an agent's identifying with an attitude on which she acts could then be marshalled to mark the difference between, say, animal or alienated agency and the more fully intentional sort. Thus, Bratman claims that "the idea of ownership of desires earns its place" by helping us to provide an analysis of human beings' "more complex forms of agency".[31] But, so the argument goes, if that is right, then surely a question will similarly arise about grades of *mental* agency, and won't one then need the notion of identification to explicate whatever sort of activity one thinks is involved in holding an attitude for reasons?[32]

Something like this, it should be recalled, was a worry of Bratman's (Section 5). The point I want to emphasize here is an assumption it makes about the order of explanation with respect to identification and agency. Roughly, for Frankfurt et al. our understanding of agency and the degrees in which it comes is in some way philosophically problematic; identification is then used to explain what agency, or at least one of its forms, *is*. In slogan form: *identification explains agency*. But this puts things back to front, from the point of view on identification and alienation I favour. On that view, agential capacities are taken for granted and then themselves employed in accounting for the difference between identification and alienation; the order of explanation is reversed. In slogan form: *agency explains identification*.[33] It is true that at a suitable level of abstraction we can see the Moranian account and the Frankfurt-style account as in agreement. As we might incautiously put it, ignoring many essential caveats: in both accounts *agency and identification co-vary* – at least with respect to mental life. For on a responsiveness-to-reasons account, when an agent's attitude is not a manifestation of their rational agency, it is thereby one from which the agent is alienated. But again, an appeal to an agent's rational agential powers is made to explain the distinction between alienation and identification, and not vice versa.

    The philosophical source of these differing attitudes to human agency must be traced, I think, to different conceptions of nature and what materials we allow ourselves – in line with some form of naturalism – when accounting for such agency. As Velleman rightly insists, "[o]ur concept of full-blooded human action requires some event or state of affairs that owes its occurrence to an agent and hence has an explanation that traces back to him".[34] But then he goes on to claim that there is an "obstacle to reconciling our conception of agency with the possible realities":

> [O]ur scientific view of the world regards all events and states of affairs as caused, and hence explained, by other events and states, or nothing at all. And this view would seem to leave no room for agents in the explanatory order. (2000b, 129)

Identification, cashed out in terms of psychic states, is then understood by Frankfurt et al. to be part of the project of fitting agents back in to nature, as it were. However, it conceives nature, though, it seems to me that a Moranian view of identification and alienation must assume that we can make sense of, and find the place in nature for, human agency *without* our having to reconstruct it out of event-causal resources.[35] For that view draws on a conception of the activity of mental life which resists reconstruction in terms of the bare existence of psychic states and the relations between them; in Moran's words "the concept of the person as a reasoning agent is [...] ineliminable".[36] Indeed, from this perspective, it will be hard not to agree with Jennifer Hornsby's verdict that when the event-causal picture is "the base line for questions in moral psychology, a shape is imposed on those questions that they should never have been allowed to take on".[37]

9. Conclusion

More would need to be said to be said about the details of the Moranian account to offer a full defence of it. My aim in the foregoing has rather been to emphasize the conflict between what seems to be a promising account of the distinction between identification and alienation and the account of agency which has been taken for granted by other theorists in this area. Whether, and how, an event-causal account of agency can be replaced by an alternative I must leave an open question. Furthermore, how such an account of agency might be related to the claims about specifically mental agency must also be left open. It goes without saying that these are hardly untroubled philosophical waters: the traditional account of agency opposed to an event-causal one is that of agent-causation, which has many detractors; and the idea of mental agency has been challenged on grounds that are arguably independent of any particular conception of agency. The problems there might be thought to be particularly acute.[38] Clearly much more needs to be said about the nature of mental agency, more overtly physical agency and the relation between them.

    However, what I want to have shown is that there is good reason to look for a non-hierarchical account of identification, and that doing so requires a serious re-assessment of the metaphysical ground on which much previous moral psychological debate has been conducted. If I am right, the Authority Problem is endemic to hierarchical analyses of identification and alienation, and the opposing account I considered sits at odds with an event-causal conception of agency. I am optimistic both that an alternative to an event-causal conception of agency is available and that this can be extended to mental agency, thereby illuminating the phenomena of identification and alienation in something like the way Moran suggests. That task, however, must await further treatment.

Notes

1. Cf. Frankfurt (1988a,1988b,1988c).
2. I only discuss Frankfurt's first and last accounts: (1988a), (1999); and Bratman's (2000, 2001, 2002) later account. For others, see Frankfurt's (1988b, 1988c) and Bratman's (1996).
3. See Harcourt (1998) for a fuller treatment of this aspect of alienation.
4. Watson (2004a).
5. This way of putting things derives from Watson's question: "[w]hat gives [higher order] volitions any special relation to "oneself"? (2004a, 29). Cf. Bratman (2000), (2001) who also speaks of agential authority in this context.
6. Bratman (2000, 47).
7. Bratman (2000, 47–48).
8. In fact, the initial specification Bratman offers is in terms of a self-governing policy in favour of the agent's treating a desire as a reason, but he worries that this would be circular. For ease of exposition I have jumped straight to Bratman's final formulation. I shall return to his circularity worry in Section 4.
9. Cf. Bratman (2002, 75). It is questionable whether one *could* be alienated from a desire which functioned as end-setting as Bratman characterizes that. I briefly return to this below, see Note 26.
10. For Bratman's argument in favour of desires functioning as end-setting, as opposed to Frankfurt's idea of their merely moving one to act, see his (2000, 54).
11. That we can use "character" in this sense might complicate any quick route from the notion of identification to an elucidation of the concept of character.
12. Another line of reasoning turns on the thought that someone's identifying with an attitude is bound up with their ability to reflect on it, and that the formation of higher order attitudes naturally captures this. Whilst I agree with the point about reflection, I disagree that it is best understood in terms of hierarchies of attitudes. For helpful discussion see Hieronymi (2009, Section 5).
13. Cf. Bratman (2000, 38–39).
14. Cf. Bratman (1996, 9), (2000, 55, Note 52), (2001, 323), (2002).
15. Cf. Bratman (2000, 38–40), (2001, 312).
16. As noted, Bratman claims that it is possible for a desire to function as end-setting without an agent's identifying with that desire, hence avoiding the supposed circularity. Against this one might ask: Could one appeal to some desired end as justifying, i.e. it function as end-setting, without one's treating it as reason-giving? "No", it seems to me. I relate this to the potential non-alienability of desires which function as end-setting in Note 26.
17. "States and events" are the usual *relata* postulated by the so-called "standard story" of action which has its origins in Davidson's (1963). For discussion, see Steward (1997).
18. This, in effect, is Mele's move. Cf. Mele (2003, ch. 10).
19. Frankfurt (1999, 105).
20. Frankfurt (1999, 105).
21. Frankfurt (1999, 105).
22. Frankfurt (1999, 105).
23. Bratman (1996, 7).
24. I should mention two other historically important accounts.
    First, Velleman (2000a, 2000b). I do not discuss this because it is not clear to me whether Velleman wants to offer an account of identification or preclude the need for one. Cf. Velleman (2000b, 137).
    Secondly, there is Watson's (2004a) initial counter-proposal to Frankfurt's (1988a). I do not discuss this because he repudiated the account in his (2004b). I think the spirit of Watson's view is maintained in Moran's, however. See Section 6 below.

25. Moran (2001, 2002). Raz (1997) and Mayr (2012) argue for similar positions.
26. Further to Notes 9 and 16 then, it seems to me that an agent identifies with an attitude to the extent that they treat its object as reason-giving, which seems to me the same as its functioning as end-setting. I therefore think that Bratman is wrong to claim that one could fail to identify with an attitude that so functioned, but also that this counts against his theory more generally. For a desire's functioning as end-setting just *is* the attitude's being actively held by the agent in response to reasons. The circularity Bratman adverts to is an expression of the fact that the good account of identification he discovers requires agential resources that go beyond an event-causal conception, and so could hardly be put to use in providing such an account.
27. This is to adapt Boyle's understanding of Moran's claims about the role of agency in self-knowledge to Moran's claims about identification and alienation. Cf. Boyle (2011a, 2011b, 2011c).
28. Such attitudes would not be "wantonly unselfconscious", as Frankfurt claimed desires with which one is satisfied must not be. Cf. Moran (2001) and Rödl (2007).
29. Cf. Moran (2002, 198).
30. I think it is plausible that alienation from an attitude will almost always result in a higher order negative attitude towards it, but this would be *explained by* its persistence in the face of the agent's recognising the worthlessness of its object.
31. Bratman (2003, 222, Note 3).
32. Frankfurt appears to make an argument of just this sort against Moran in his response (2002b) to Moran's (2002).
33. Compare Steward (2012, 60, 148).
34. Velleman (2000b, 127).
35. There have been some recent views of agency which take an event-causal conception of nature to task. Cf. Lowe (2010), Mayr (2012) and Steward (2012).
36. Moran (2001, 193).
37. Hornsby (2004, 2–3).
38. Cf. Strawson (2003).

Notes on contributor

Alec Hinshelwood is a PhD student at University College London. His interests are in the philosophy of mind and ethics, but especially in their intersection in the philosophy of action. He is also interested in Aristotle, Kant, Marx, and Freud.

References

Boyle, M. 2011a. "Transparent Self-Knowledge." *Aristotelian Society Supplementary* 85 (1): 223–241.
Boyle, M. 2011b. "'Making Up your Mind' and the Activity of Reason." *Philosophers' Imprint* 11 (17): 1–24.
Boyle, M. 2011c. "Active Belief." In *Belief and Agency*, edited by D. Hunter, 119–147. Calgary: University of Calgary Press.
Bratman, M. 1996. "Identification, Decision, and Treating as a Reason." *Philosophical Topics* 24 (2): 1–18.
Bratman, M. 2000. "Reflection, Planning, and Temporally Extended Agency." *The Philosophical Review* 109 (1): 35–61.
Bratman, M. 2001. "Two Problems About Human Agency." *Proceedings of the Aristotelian Society New Series* 101 (1): 309–326.
Bratman, M. 2002. "Hierarchy, Circularity, and Double Reduction." In *Contours of Agency*, edited by S. Buss and L. Overton, 65–85. London: MIT Press.
Bratman, M. 2003. "A Desire of One's Own." *The Journal of Philosophy* 100 (5): 221–242.
Davidson, D. 1963. "Actions, Reasons, and Causes." *The Journal of Philosophy* 60 (23): 685–700.
Frankfurt, H. 1988a. "Freedom of the Will and the Concept of a Person." Chap. 2 in *The Importance of What We Care About*. Cambridge: CUP.
Frankfurt, H. 1988b. "Identification and Externality." Chap. 5 in *The Importance of What We Care About*. Cambridge: CUP.

Frankfurt, H. 1988c. "Identification and Wholeheartedness." Chap. 12 in *The Importance of What We Care About*. Cambridge: CUP.

Frankfurt, H. 1999. "The Faintest Passion." Chap. 8 in *Necessity, Volition, and Love*. Cambridge: CUP.

Frankfurt, H. 2002a. "Reply to Thomas Scanlon." In *Contours of Agency*, edited by S. Buss and L. Overton, 184–188. London: MIT Press.

Frankfurt, H. 2002b. "Reply to Richard Moran." In *Contours of Agency*, edited by S. Buss and L. Overton, 218–226. London: MIT Press.

Frankfurt, H. 2006. *Taking Ourselves Seriously and Getting It Right*. Stanford, CA: Stanford University Press.

Harcourt, E. 1998. "Mill's 'Sanctions', Internalization and the Self." *European Journal of Philosophy* 6 (3): 318–334.

Hieronymi, P. 2009. "Two Kinds of Agency." In *Mental Actions*, edited by L. O'Brien and M. Soteriou, 138–162. Oxford: OUP.

Hornsby, J. 2004. "Agency and Actions." In *Agency and Action*, edited by J. Hyman and H. Steward, 1–24. Cambridge: CUP.

Lowe, E. J. 2010. *Personal Agency: The Metaphysics of Mind and Action*. Oxford: OUP.

Mayr, E. 2012. *Understanding Human Agency*. Oxford: OUP.

Mele, A. 2003. *Motivation and Agency*. Oxford: OUP.

Moran, R. 2001. *Authority and Estrangement: An Essay on Self-Knowledge*. Oxford: Princeton University Press.

Moran, R. 2002. "Frankfurt on Identification: Ambiguities of Activity in Mental Life." In *Contours of Agency*, edited by S. Buss and L. Overton, 189–217. London: MIT Press.

Raz, J. 1997. "The Active and the Passive." *Proceedings of the Aristotelian Society, Supplementary Volumes* 71: 211–227.

Rödl, S. 2007. *Self-Consciousness*. London: Harvard University Press.

Steward, H. 1997. *The Ontology of Mind: Events, Processes, and States*. Oxford: OUP.

Steward, H. 2012. *A Metaphysics for Freedom*. Oxford: OUP.

Strawson, G. 2003. "Mental Ballistics or the Involuntariness of Spontaneity." *Proceeding of the Aristotelian Society New Series* 103 (1): 227–256.

Velleman, J. D. 2000a. "Introduction." Chap. 1 in *The Possibility of Practical Reason*. Oxford: OUP.

Velleman, J. D. 2000b. "What Happens When Someone Acts?" Chap. 6 in *The Possibility of Practical Reason*. Oxford: OUP.

Watson, G. 2004a. "Free Agency." Chap. 1 in *Agency and Answerability*. Oxford: OUP.

Watson, G. 2004b. "Free Action and Free Will." Chap. 6 in *Agency and Answerability*. Oxford: OUP.

Wollheim, R. 1993. "The Good Self and the Bad Self." Chap. 3 in *The Mind and Its Depths*. London: Harvard University Press.