

## Taking the Self out of Self-Rule

Michael Garnett

Accepted: 18 October 2011  
© Springer Science+Business Media B.V. 2011

**Abstract** Many philosophers believe that agents are self-ruled only when ruled by their (authentic) selves. Though this view is rarely argued for explicitly, one tempting line of thought suggests that *self-rule* is just obviously equivalent to *rule* by the *self*. However, the plausibility of this thought evaporates upon close examination of the logic of ‘self-rule’ and similar reflexives. Moreover, attempts to rescue the account by recasting it in negative terms are unpromising. In light of these problems, this paper instead proposes that agents are self-ruled only when not ruled by others. One reason for favouring this negative social view is its ability to yield plausible conclusions concerning various manipulation cases that are notoriously problematic for nonsocial accounts of self-rule. A second reason is that the account conforms with ordinary usage. It is concluded that self-rule may be best thought of as an essentially social concept.

**Keywords** Self-rule · Autonomy · Agency · Manipulation

We all have a reasonable idea of what it means to say that one person rules over another. This paper concerns what, exactly, it is for a person to rule over herself. One reason why this idea of individual self-rule, or of personal autonomy, is puzzling is that it seems that in order for one person to rule over another that person must possess some power that the other lacks. That is, to be a ruler one must be able to make and unmake laws that are binding on those who are ruled, and it is difficult to see how laws can be truly binding on anyone who possesses that same power to make and unmake those very laws. So it seems that the ruler and the ruled cannot have precisely the same powers. Yet in a case of self-rule the ruler and the ruled are precisely the same person.

---

M. Garnett (✉)  
Department of Philosophy, Birkbeck College, University of London,  
Malet Street, London WC1E 7HX, UK  
e-mail: m.garnett@bbk.ac.uk

This is perhaps the problem that Plato had in mind in the *Republic* when he observed:

But ‘master of oneself’ is an absurd phrase. For if you’re master *of* yourself you’re presumably also subject *to* yourself, and so *both* master *and* subject. For there is only one person in question throughout.

(1987: 430e)

No command one issues to oneself can be binding, since at any time one may choose to repeal that command and to issue oneself with a new one. One cannot simultaneously be both master and subject.<sup>1</sup> Given this, what can ‘self-rule’ possibly be?

The aim of this paper is to cast doubt on one very influential attempt to answer this question, and to make a case for taking seriously the claims of a less popular alternative. The solution to be challenged is the ‘deep self’ view of self-rule, according to which an agent is self-ruled only when she is ruled by her ‘deep’ or ‘authentic’ self. The solution to be advanced in its place is a negative social view of self-rule, according to which an agent is self-ruled only when she is not ruled by any other agent. I discuss these in turn.<sup>2</sup>

## 1 The Deep Self View of Self-Rule

After ridiculing the idea of ‘self-mastery’, Socrates continues:

What the expression is intended to mean, I think, is that there is a better and a worse element in the personality of each individual, and that when the naturally better element controls the worse then the man is said to be ‘master of himself’.

(1987: 431a)

This idea, that self-rule may be understood as rule by some particular part of oneself over one’s other parts, has been extremely influential. However, few take self-rule to require that one be ruled by a part of oneself that is *better* than one’s other parts. After all, a person may be alienated from his better nature, like the torturer who experiences his occasional pangs of guilt as moments of weakness; and if such a person were, despite his best efforts, to falter in trying to implement some intention, most theorists would regard this as a failure of self-rule.

What matters for the majority of autonomy theorists is not the quality of the ruling part but rather its authenticity: what is necessary for self-rule is that one’s ruling part be, in some sense, more *oneself* than one’s other parts. This way of thinking relies on the idea that a person is not to be identified equally with all parts of her mind: the idea that, while some of her attitudes are deeply expressive of who she is, others are simply alien forces at work within her. Hence it is theoretically possible on this view to divide a person up into those parts of himself with which he is identified and those parts of himself from which he is alienated. The thought is then that when he is ruled by an alien part of himself, he lacks

<sup>1</sup> C.f. Hobbes: ‘nor is it possible for any person to be bound to himself; because he that can bind, can release; and therefore he that is bound to to himself onely, is not bound’ (1996: 184). As Cohen (1996: 168–70) notes, there may be a thin sense in which one is bound by a law one can repeal, until one repeals it. Yet the significance of this fact decreases with the ease of repeal and, in the limiting case where one has only to will it for it to be so, the fact is of vanishing significance. For this reason I leave it aside.

<sup>2</sup> These are of course not the only possible accounts: for another, see e.g. the ‘positive’, ‘whole-self’ account suggested in Part I of Mele 1995. Indeed, nothing said here prevents such a view from being combined with the view proposed below.

self-rule, and that when he is ruled by a part of himself with which he is deeply identified, he enjoys self-rule.

This talk of parts, persons and selves is apt to become confusing, so let me nail down some terminology. I shall use the term 'agent' to refer to the person naïvely conceived—as identified, in a shallow sense, with the totality of her mental attitudes.<sup>3</sup> And I shall use the term 'deep self' to refer to that part (or those parts) of the agent that are more profoundly expressive of who the agent is, and with which the agent is identified in some deeper sense. The view before us may be stated as:

*The deep self view of self-rule* An agent is self-ruled just in case she is ruled by her deep self

For proponents of this view, the task of providing an account of self-rule is the task of providing an account of the deep self. Indeed, so widespread is the commitment to this view that for many years the search for a correct theory of autonomy has been virtually synonymous with the search for a correct theory of the self. Accordingly, accounts of the latter have proliferated: the self is variously identified with the agent's higher-order preferences (Frankfurt 1971, 1999), with its system of values (Watson 1975), with its psychologically most well-integrated attitudes (Ekstrom 1993), and with its faculty of practical reason (Velleman 1992; Moran 2002). However, I here take no stand on the identity of the deep self. Moreover, in arguing against the deep self view of self-rule I do not seek to dispute the claim that there exists such a thing as the deep self. Instead I shall argue that, even if we do have deep selves, this view does not represent a good way of thinking about self-rule.<sup>4</sup>

So eager are proponents of the deep self view of self-rule to uncover the nature of the deep self that they rarely pause to offer arguments in favour of the view itself. Thus Robert Noggle claims that 'whatever we think autonomy is, if one acts on an alien desire, one does not act autonomously' (1995: 57). Laura Ekstrom asserts that 'one's action is self-governed when it is directed by the true self' (2005: 155). Harry Frankfurt understands 'autonomy to entail subjection to the essential requirements of the self' (2002: 293). David Velleman states that a person's faculty of practical reason 'can be the locus of his autonomy, by virtue of being his essential self' (2002: 115). In all of these cases, the deep self view of self-rule is taken simply as an uncontroversial starting point. Indeed, it is typical for proponents of this view to write as though they regard it as self-evidently true, as something close to a conceptual truth. Admittedly, it is not hard to see why. On the deep self view of self-rule, a person is *self-ruled* just in case her *self* is doing the *ruling*. This seems like an extremely natural way of thinking about autonomy. What else could self-rule possibly be, one may wonder, if not rule by the self?

Yet this tendency of the view to look like something close to a conceptual truth is deeply misleading. In fact, it depends only on a kind of linguistic trick. 'Self-rule' is a reflexive.

<sup>3</sup> 'This is the sense in which everything is attributable to me that occurs in my conscious life or figures in the best overall explanation of my conscious life and behaviour' (Scanlon 2002: 170). Note that some theorists, such as Korsgaard (2010), sometimes deny that beings 'naïvely conceived' in this way are agents at all, reserving the term 'agent' as an honorific for beings ruled by their deep selves. However, this disagreement is merely terminological (even Korsgaard allows that lower animals are agents in an extended sense (2010: 81–108)).

<sup>4</sup> Thus in denying the deep self view of self-rule I do not thereby also deny the importance of rule by the deep self to other important phenomena, such as true agency or moral responsibility. Rather, my argument's implication is simply that insofar as these things require rule by the deep self, they do not thereby require self-rule.

Thus the 'self' in 'self-rule' does not serve to introduce any notion of the *self*, either deep or shallow. Compare 'self-ruled' with another reflexive, such as 'self-adhesive'. No one thinks that a self-adhesive envelope is an envelope that adheres to something called its 'self': a self-adhesive envelope is, like all envelopes, an entity without a self, adhesive or otherwise. Similarly, flour needs no self in order to be self-raising, a mattress needs no self in order to be self-inflating, and an electric saw needs no self in order to be self-adjusting.

Instead of introducing any notion of the self, reflexives typically serve simply to denote a relation that holds between an object and itself, i.e. between an object and that same object. Thus, something is self-adhesive when the thing that adheres is identical with that to which it adheres; something is self-raising when the thing that does the raising is identical with that which is raised; and so on. The same is true as regards self-rule: something is self-ruled just in case the thing that is doing the ruling is identical with the thing that is being ruled. So there is no more initial reason to suppose that 'self-rule' requires the activity of a self than there is to suppose that any other reflexive relation requires the activity of a self.

Of course, self-rule is standardly ascribed to agents, and agents do indeed have selves. This said, let us consider the case of the self-adhesive envelope more carefully. How does such an envelope adhere to itself? By virtue of some part of it adhering to some other part or parts of it. That is, the adherence of some part of the envelope to some other part of it is sufficient to render the envelope self-adhesive as a whole. So, by analogy with the envelope, an agent may be self-ruled by virtue of some part of that agent ruling over some other part or parts of that agent. Thus, we may say that an agent may be self-ruled by virtue of the part of her that is her self ruling over the rest of her.

This is true, but it does not get us to the deep self view of self-rule, for its proponent must still explain why self-rule is *exclusively* a matter of rule by the self. An envelope is self-adhesive no matter which part of it does the adhering; by convention, and for obvious reasons, the adhesive part is typically the underside of the flap, but were it instead the bottom right corner, the envelope would still qualify as 'self-adhesive'. (That is, one's grounds for returning such an envelope to its manufacturer would not be that it was not self-adhesive, but that it was useless.) In the same way, there is no reason to think that the self-ruled agent must qualify as such by virtue of the rule of any one particular mental part. So we have been given no reason why an agent cannot qualify as self-ruled in virtue of one of her 'alien' parts doing the ruling. After all, such an agent both rules (by virtue of the fact that a part of her rules) and is ruled (by virtue of the fact that a part (or some parts) of her are ruled). So the proponent of the deep self view has still not explained why an agent rules himself only when he is ruled by that part of him that is his deep self.

Once this problem is appreciated, it is difficult to see how a defender of the view might respond. The problem is essentially one of arbitrariness: there appear to be no grounds on which to insist that an agent rules herself only when she is ruled by one (as opposed to any other) part of herself. Note again that this is *not* the problem of determining which part of a person constitutes her deep self. It may be that we lack grounds on which to decide between various possible accounts of the deep self, and that therefore the adoption of any particular account must be arbitrary. But this is not the arbitrariness problem I am pressing here. Instead the present argument is that, even once we know that some part of a person is indeed her deep self, there is no reason to think that she is self-ruled only when ruled by *that part*. What is arbitrary is insisting that self-rule requires that one be ruled by one part of oneself rather than any other part, even after it is agreed that that part constitutes one's deep self.

Perhaps it will be replied that the deep self view of self-rule simply describes how *we* (that is, philosophers) use the term 'self-rule'. Yet the deep self view serves not to stipulate a terminological definition but to advance a substantive conceptual claim, one about what it

is to be in charge of oneself, to govern oneself, and to be in control of one's own behaviour. Alternatively, it might be claimed that 'self-rule' is really just a stand-in for some other philosophical concept, such as true agency or moral accountability, and that the deep self view should be read simply as an account of this other thing and not of self-rule itself. Yet this means either denying that we have any independent concept of self-rule or autonomy at all, which seems unnecessarily radical, or else denying that the deep self view is an account of self-rule, in which case we are no closer to discovering when an agent is self-ruled.

## 2 'Self-Rule' as a Negative Concept

The deep self view of self-rule (as it has so far been characterised) might usefully be classified as a *positive* conception of self-rule, on the grounds that it requires the self-ruled agent actively to do something (i.e., *rule*). Yet not all reflexives serve to denote positive capacities: for instance, one is not *self-reliant* in virtue of something positive that one does to oneself, but simply in virtue of one's not relying on anyone else. This suggests that, in light of the problems raised in the previous section, we might do well to consider whether 'self-rule' is better construed negatively; that is, whether self-rule might require not that the self-ruled agent rule himself in some positive sense but, instead, that he not be ruled by anything else. Let me state this idea as:

*The negative view of self-rule* An agent is self-ruled just in case she is not ruled from without

To make progress with this thought, we must clarify what is meant by 'rule from without'. Later in the paper I shall argue that we should adopt a particular, social reading of this. But before that I must address a further issue concerning the deep self view. For many deep self theorists consider their view of self-rule to be consistent with the negative view as here stated. Indeed, it is not uncommon for such theorists to characterise their view in explicitly negative terms.<sup>5</sup> Moreover, the reason for this is straightforward: on the deep self view, an agent lacks self-rule when her behaviour is ruled by 'alien' motivations (that is, motivational forces that are external to her deep self), and this can quite naturally be described as 'rule from without'. Yet, as I shall now explain, it is in fact inadvisable for the deep self theorist to attempt to cast her view in this way.

To see why, we need to press the deep self theorist to answer the following simple question: when an agent's behaviour is determined by her deep self in the required way, which entity is it that thereby qualifies as self-ruled? There are two possible candidates. The first is the whole agent, naïvely conceived. Clearly, however, this cannot be what the deep self theorist has in mind, since a mere absence of rule by forces *external to the whole agent naïvely conceived* is consistent with that agent's behaviour being determined by its 'alien' motivations (since these motivations are internal to the agent considered in this expansive way). So applying the negative view to

<sup>5</sup> According to Laura Ekstrom, for instance, 'we all agree that autonomy... is opposed to rule-from-without' (2005: 155).

the whole agent fails to yield anything like the deep self view of self-rule. The second candidate is the deep self. Yet this is also problematic because, of course, the deep self *is* 'ruled by' forces external to it: it is not some isolated entity causally sealed off from its surroundings. The deep self is shaped by relevant features of the world, such as (on a popular view) the agent's perceptions of the reasons. So the deep self theorist still needs to explain why some of the forces to which the deep self is subject preserve its autonomy whereas others undermine it; which is to say, the deep self theorist needs to explain why rule by certain external forces does *not* qualify as 'rule from without'.

Now, it may appear that the deep self theorist has an obvious and familiar response to this. To the question 'how can I be self-ruled if I am in fact ruled by my perceptions of the reasons', a standard answer is: 'remember that your rational faculty is just you in the deep sense, and that you are self-ruled insofar as it governs you; so causal forces that pass through your deep self in this way are therefore not threats to your autonomy, but rather conduits through which your autonomy is manifested.' Yet—crucially—this explanation is apt only in contexts where 'self-rule' is applied in a positive sense to the whole agent (as in §1). This is because the aim of the reply is to explain why rule by the deep self does not compromise the self-rule of the whole, and the explanation is that, since self-rule is rule by the self (that is, rule of the whole agent by the deep self), rule by the reasons is no threat to self-rule. In the present context, however, where we are interested in the self-rule of the *deep self* (rather than in the self-rule of the whole), this simply misses the point. What we need is an explanation of how the *deep self* can remain (negatively) self-ruled despite being subject to external forces. And this is an explanation which the familiar response does not provide.

Nevertheless, it is possible to work out what the negative deep self theorist's account of autonomy-undermining external influences would have to look like. The account would have to be that the deep self is never ruled by any of its *inputs* (such as perceptions of reasons), being ruled only by *impediments to its outputs* (such as rogue desires that stand in the way of rational action). Yet as a specification of the negative view of autonomy, this is unattractive: other than the fact that it helps out the deep self view, there is little to say for it, and much to say against it. First of all, it cannot be generalised. We typically ascribe self-rule not only to deep selves but also to other entities, such as nations. No one, however, would hold that a nation lacks self-rule only when it is prevented from acting, and not also when its decisions are determined by the wrong sorts of causes (for instance, were a majority of a nation's parliament brainwashed by an enemy, so that its collective decision-making processes were effectively commandeered by the foreign power, we would not naturally deem the nation to be fully self-ruled). Second, it collapses the idea of *self-rule* into that of *power*: on this view, the self-rule of a thing is simply its ability to achieve its aims unimpeded. This means that, on this view, the concept of self-rule no longer has a distinct role to play in our conceptual economy, separate from others such as those of agential power or pure negative liberty. But while the mere ability of a thing to realise its aims may be a matter of its power, or of its rule over its environment, it is not, surely, a matter of its rule over *itself*.

Thus it is most likely a mistake for the deep self theorist to try to recast her view of self-rule in purely negative terms. Given both this and the problems faced by the positive version of the deep self view, I now leave it to one side. In its place I consider a different possible elaboration of the negative view: the idea that 'rule from without' may best be interpreted as 'rule by others'.

### 3 A Social View of Self-Rule

On the negative view of self-rule, a person is self-ruled when she is not ruled by external forces. In the remainder of this paper I make a case in favour of understanding such forces as concerning solely the activities of other agents. Specifically, I defend:

*The negative social view of self-rule* An agent is self-ruled just in case she is not ruled by other agents

In §§ 4 and 5 I present two arguments in favour of this view. Before that, however, it will be helpful to make some brief observations about the account. First, it should be noted that it represents a more radical break with traditional ways of thinking about autonomy than do many other attempts to rethink the concept along more ‘social’ or ‘relational’ lines. Indeed, there are two respects in which this is the case. First, it takes autonomy to be constitutively social, rather than merely social in genesis. On many relational views (such as Meyers’ (1989)), great emphasis is placed on the social conditions necessary for the development of autonomy, but autonomy itself is nevertheless analysed in terms of capacities that are in themselves nonsocial. On the present view, by contrast, autonomy is understood in terms of a condition that is itself inherently social. Second, and more importantly, even views that seek to understand autonomy in constitutively social terms typically aim to do so via some constitutively social account of the *self* (e.g. Friedman 2003). This is because it is assumed that such a social account of the self will generate a social account of autonomy, on the grounds (presumably) that autonomy is rule by the self. Thus these views typically remain within the broad theoretical framework of the deep self view, and so are vulnerable to the arguments presented above. By contrast, the social view proposed here involves a genuinely distinct analysis of the concept of self-rule.

This is a significant advantage for several reasons, one of which is the fact that it allows us to separate out judgements concerning a person’s autonomy from judgements concerning her authenticity. It allows us, for instance, to regard a person subject to ongoing coercion, such as a slave, as lacking in self-rule, without thereby having to regard him as also suffering from any kind of internal fracturing of his will. More significantly, it allows us to regard a victim of ideological domination, such as a woman who has been taught to place the needs of men always above her own, as lacking in self-rule in important respects, while also recognising that the problematic values on which she acts may nevertheless be authentic and genuinely her own. We thus avoid a politically problematic dilemma endemic to deep self views of all types, whereby we are compelled to choose between recognising the fact that oppressed people lack autonomy in a significant sense and the fact that they may nevertheless qualify as agents *par excellence*. This is a result of the fact that, on the negative social view presented here, one’s autonomy is purely a matter of the type of social control to which one is subject and need have nothing especially to do with the quality of one’s agency.

Note also that, as stated above, this proposed view is merely an account-sketch. Just as any full version of the deep self view of self-rule must include an account of the deep self, so any full version of the negative social view of self-rule must include an account of interpersonal rule, and different versions of the view will be generated by different accounts of the latter. Yet, again, just as my argument against the deep self view did not target any one particular version of the view, so my arguments in favour of the negative social view proceed at a similarly general level. That is, my present aim is not to argue in favour of any one version of the negative social view, but merely to argue that views of this general type deserve to be brought within the philosophical fold.

For this reason I here attempt to remain mostly agnostic on the crucial question of how best to elucidate our ordinary notion of interpersonal rule. I note, simply, that any plausible philosophical account of interpersonal rule will have to be guided by certain relatively firm and widely shared intuitions. Most will agree, for instance, that to *rule* someone is more than to cause that person to do something, and more, even, than intentionally to get that person to do something. Thus, for instance, children taught to think critically, people rationally swayed by the better arguments, and recipients of attractive offers are not thereby subject to the rule of others. By contrast, victims of coercion, members of psychologically oppressive cults, and (most of) the characters from Huxley's *Brave New World* (1932) are paradigmatic examples of people subject to the rule of others. These are all instances of the data that any plausible theory of interpersonal rule must respect. For current purposes, therefore, I take the social view's reference to 'rule by other agents' to refer to something like our ordinary notion of interpersonal rule, a notion with conceptual contours that are broadly familiar to us all and with which any plausible philosophical account of the concept can be reasonably expected to conform.

However, although I do not here endorse any particular account of interpersonal rule, it is nevertheless necessary for me to reject one. Some readers may be attracted to the thought that to rule a person is to influence her in ways that undermine the rule of her deep self (and hence that undermine her self-rule, on the deep self view). Yet to adopt this account would be to collapse the negative social view back into the deep self view. So it is incumbent upon the social theorist to explain why interpersonal rule is not best analysed in terms of rule by the deep self. Accordingly, here is a simple argument to this effect. It is an uncontested fact about our ordinary concept of rule that coercion, and especially unauthorised coercion, is a means by which one person can rule another: no one would deny, for instance, that a government typically rules its citizens by virtue of the coercive structures that it imposes. Yet it has long been established that coercion need not undermine the control of one's deep self over one's behaviour (Frankfurt 1975): though the coerced party faces a worsened choice situation, the choice she makes may nevertheless be perfectly rational or wholehearted.<sup>6</sup> So ruling over another is not essentially a matter of undermining the position of her deep self.

If this paper succeeds in its task, we autonomy theorists will then have reason to turn our attention towards the development of a more promising philosophical account of interpersonal rule. First, however, I must make my case that such a project would be worthwhile. To this end I now make two arguments on behalf of the negative social view. The first is that it is better placed than the deep self view to make sense of our intuitions concerning manipulation cases. The second is that it allows for an account of autonomy that accords well with ordinary usage.

#### 4 Manipulation

Many people share a strong intuition that victims of severe and unauthorised manipulation are not fully autonomous.<sup>7</sup> As I shall argue in this section, however, this is an intuition that

<sup>6</sup> Sarah Buss (2005) makes similar points with respect to deception and manipulation.

<sup>7</sup> Many but not all people. In particular, some deep self theorists accept the main claim of this section, that manipulation need not undermine the rule of one's deep self, and conclude, not (as I do) that self-rule is therefore not a matter of rule by one's deep self, but instead that self-rule is consistent with manipulation (Buss 2005). For those who feel no (even *prima facie*) resistance to this latter claim the argument of this section will lack persuasive force, and, as regards such people, I must rest my case solely on the considerations presented elsewhere in this paper.



traditional views struggle to explain. By contrast, on any version of the negative social view such forms of manipulation pose obvious problems for self-rule. I focus on two types of manipulation: deception and global character manipulation.

Alfred Mele gives the following example of autonomy-undermining deception:

Connie is deliberating about how best to invest her money. A respected investment firm has provided her with detailed information about a wide range of options. Connie has good reason to believe that the information is accurate and no reason to be suspicious of the firm. She deliberates on the basis of relevant values and desires of hers together with the information provided and rationally concludes that a certain investment policy would be best for her at this time. As it happens, however, Connie was systematically deceived by the firm. Their figures were contrived, assembled with the design of leading any rational agent with Connie's interests to decide on an investment policy that would maximize benefit to the firm at the investor's expense.

(1995: 179)

The negative social view can easily yield the result that Connie lacks self-rule in making her decision, since on most plausible accounts of interpersonal rule her behaviour will count as subject to the rule of the firm. By contrast, it is unclear how the same result could be plausibly yielded by any account of autonomy that fails to recognise the relevance of such essentially interpersonal factors.

For instance, John Christman (2007: 12, 21 and 2009: 155) requires for autonomy that, were a person to reflect critically on a mental attitude in light of the historical processes by which it arose, she would feel no need to repudiate it or to resist its effects. Thus Connie is autonomous in her investment decision only if she would stand by that decision under knowledge of the historical process by which she came to make it. Now, if 'knowledge of the historical process' is interpreted broadly so as to include knowledge of the fact that her relevant beliefs are likely to be false, then Christman's view does indeed deliver the conclusion that Connie lacks autonomy. However, this is not how Christman interprets the requirement, and for good reason, since it would commit him to the implausibly strong principle that false belief is undermining of autonomy in general. Not every mistaken belief is a nonautonomous belief.<sup>8</sup>

Instead, Christman merely requires of these cases (cases, that is, in which a desire arises on the basis of new information) that the 'process of reasoning is internally consistent (in a weak sense)' (1991: 13). That is, he takes the relevant 'historical process' in these cases to be the generation of the decision from the belief, not the generation of the belief itself. So since Connie's reasoning is perfectly rational, Christman's view is blind to her non-autonomy.<sup>9</sup>

Moreover, this problem generalises. For it is possible to imagine an autonomous but still mistaken agent, such as one who has erroneously picked up an outdated copy of the firm's brochure, who differs from Connie *only* in point of not being a victim of deception. And unless we are willing to treat this irreducibly social difference as relevant to autonomy, it seems, we must deem either both or neither of these agents autonomous: our account must

<sup>8</sup> Note that not even Susan Wolf (1990) holds that autonomy requires true belief—for her, autonomy requires only the *capacity* to track the truth.

<sup>9</sup> Mele (1995: 187) requires, as part of a sufficient condition for autonomy, that 'the agent's beliefs are conducive to informed deliberation about all matters that concern him'. For the reasons just given, this is too strong to serve also as a necessary condition, and it is difficult to see how Mele could attenuate it in a plausible way without incorporating some inherently social requirement.

either be too weak or too strong. This is because what matters here is not the agent's epistemic situation considered simply in itself, but how that situation positions her in relevant relations of social power.<sup>10</sup>

I turn now to global character manipulation. The following case is adapted from Mele (2006: 184–9):

Diana, a goddess on the verge of death, will be resurrected only if a man dressed in blue spills the blood of another on a certain spot at a certain future time. Knowing this, and having also detailed and extensive knowledge of the current state of the universe and the laws of nature, she creates a man, Ernie, such that he will perform this precise action. So as to avoid worries about mental content in beings created *ex nihilo* (the ritual will work only if Ernie has mental content, and Diana has read Davidson), she does this by creating a zygote in Mary which will grow into Ernie. Ernie's constitution is such that, when placed in the external circumstances Diana knows will obtain, he is guaranteed to buy a house in a certain location, to wear blue pyjamas, and to defend his family one particular night by attacking an intruder with a kitchen knife.

It is natural to regard Ernie as here subject to Diana's rule. On plausible versions of the negative social view, therefore, he does not perform the specified action with full autonomy. By contrast, other accounts struggle to recognise Ernie's loss of autonomy.<sup>11</sup> This is because it is possible to fill in the details such that, again, the *only* difference between Ernie and an autonomous counterpart is the fact that, in Ernie's case, the relevant causal chains are initiated by a manipulator. So in order to explain Ernie's nonautonomy we must cite the presence of the manipulator—that is, an essentially social feature of the case. In light of these difficulties, Mele (2008: 278–85) urges us to bite the bullet and to insist that, despite appearances, Ernie may be fully autonomous.<sup>12</sup> On the negative social view, by contrast, there are no bullets to bite. Thus insofar as we are inclined to think that manipulation of this type undermines autonomy, we have reason to endorse this view.

It may be replied: how can the social view differentiate Ernie's relationship with Diana from a normal individual's relationship with his parents and educators? All of us, after all, have characters that have been shaped by others, and the negative social view's recognition of Ernie's lack of autonomy will count for little if it turns out that, on that view, there is no relevant difference between Ernie and you or I. Yet it is a fact about our ordinary notion of interpersonal rule that, in normal cases, adults are not subject to the rule of their parents and schoolteachers, even though their characters may have been intentionally shaped by them. By contrast, I expect many to share the intuition that Ernie is subject to the rule of Diana. So the simple response is that, judged by the standards of our ordinary notion, Ernie is subject to a form of external rule that normal adults are typically not.

<sup>10</sup> Notice also that the problem applies regardless of one's stance on determinism. No sane incompatibilist denies that an agent's choices should be influenced by her relevant beliefs, and that these, in turn, should be the appropriate effects of certain features of the world. (For more on the incompatibilist's trouble with manipulation, see Mele 2006: 138–44.)

<sup>11</sup> Again, they struggle regardless of their stance on determinism. Suppose that Ernie's decision to stab the intruder is not fully determined by prior causes, but that Diana has arranged both Ernie and his environment so as to make it extremely likely that he does so, and he does so. Ernie's autonomy is compromised all the same.

<sup>12</sup> To be precise, Mele urges compatibilists to bite the bullet, since this case enters his discussion simply as a possible argument against compatibilism.

Of course, this response assumes that it will be possible to develop an account of interpersonal rule that preserves these features of the ordinary concept. As some readers may be sceptical on this point, let me just very briefly mention a couple of ways in which an account of interpersonal rule might seek to do this. One strategy would be to require of rulers that they exercise more than some specified degree of control over their subjects. For although normal educators exercise some general influence over the future behaviour of their pupils, in normal cases these chains of control are rendered relatively weak by the open-ended and critical processes of intellectual evolution that people standardly undergo. Indeed, this is the reason why indoctrination aims to prevent its victims from questioning what they are taught: indoctrinators seek to achieve substantially greater degrees of control over their pupils' future behaviour by *shutting down* these unpredictable critical processes. With Diana, by contrast, a substantially greater degree of control is achieved by the alternative expedient of supernatural foreknowledge. Like the indoctrinator, then, Diana is able to make reliable predictions about the long-term effects of her interventions on her victim's behaviour, albeit by increased knowledge of the effects of her actions rather than by simplification of the relevant processes. Thus, on this first possible approach, Diana qualifies as a ruler because she exercises a degree of intimate control over her subject that ordinary, non-indoctrinating educators do not. Alternatively, a second strategy would be to require simply that, as regards non-persons (or proto-persons), exercises of control are ruling only when they are not in the objective interests of the being over which they are exercised. This would yield the result that we do not rule over a small child by developing its rational capacity or imparting it with knowledge. By contrast, it would plausibly entail that Diana rules Ernie by shaping him such that he will later perform her ritual.

Clearly these brief thoughts are no more than that. My motive in mentioning them here is simply to help ward off premature scepticism about the prospects of a philosophical account of interpersonal rule that does at least minimal justice to our intuitions. For on any account that does, I have argued, the negative social view is better placed to make sense of manipulation cases than rival views.

## 5 'Autonomy'

The second argument in favour of the negative social view of self-rule is that it allows for an account of autonomy that accords with ordinary usage. This claim may strike many as surprising, since 'autonomy' and its cognates are widely thought to be no more than technical philosophical terms. Thus Diana Meyers: 'the term "autonomy" is not in everyday usage, and, it must be admitted, it is a philosophical term of art' (2000: 151). In addition, Michael McKenna writes that 'autonomous agency seems almost exclusively a term of art largely unrecognised outside of philosophical discourse', continuing: 'it is nearly impossible to pick up the Sunday paper and find an article devoted to the autonomy of some agent's conduct' (2005: 206). Yet it is in fact not difficult to find newspaper articles concerned with autonomy: a brief survey shows that the term is alive and well.

Today, for example, we are told that a proposed consumer protection agency would have 'so much autonomy that it might as well be an independent agency', despite being based within the Federal Reserve (Hoover 2010); that the general secretary of the NSCN (I-M) has rejected an offer by the Indian government of 'greater autonomy' for the Naga people (Chishi 2010); that a Kentucky bill intended to ease the transition of students from community colleges to four-year schools now 'accommodates concerns of higher education officials about their autonomy' (Blackford 2010); and that a report on the future of UK

secondary schools recommends that all schools should become 'autonomous within five years', though only 'if they can handle independence' (Williams 2010).

In every one of these cases the term 'autonomy' is used to denote an absence of rule by others. Moreover, other academic disciplines tend to use 'autonomy' in this same way. Thus, in robotics and artificial intelligence research, an 'autonomous' agent is an agent not in need of on-line human control (Haselager 2005: 518). In business and management studies, employees may be more or less autonomous depending upon the level of direction they receive from management, and a local branch of a company may be more or less autonomous depending upon its independence from head office.<sup>13</sup> In educational theory, learning is autonomous to the extent to which it is independent of the direction and control of the teacher.<sup>14</sup>

Of course, such purely linguistic considerations need not dictate how we as philosophers go about using our own theoretical terms.<sup>15</sup> But there is a good methodological reason for deferring to them in this case. Most autonomy theorists seek to shape their theories around our intuitions about autonomy; that is, theories of autonomy are standardly tested by imagining different agents and considering which of them we would be inclined to regard as autonomous. Clearly, it is essential that we focus only on relevant intuitions: we will not get anywhere by testing our theory of  $x$  against our intuitions concerning  $y$ . Yet we are at risk of doing precisely this in the case of autonomy, since in addition to whatever esoteric meaning a philosopher has given the term, it also has an established exoteric meaning (as has just been shown). Our intuitions are therefore in permanent peril of being confused or distorted by extraneous concerns.

For example, suppose that we are attempting to elucidate a notion of authenticity, which we call 'autonomy', and that our method involves consulting our intuitions about which agents are autonomous. Given that in ordinary language 'autonomy' has a social meaning quite distinct from our target notion, introducing the term to the investigation is likely to complicate matters unnecessarily. Unless we are exceedingly careful, we may find ourselves mistakenly trying to accommodate intuitions that are motivated by the social notion—intuitions about manipulation cases, for instance—and strictly irrelevant to the concept under investigation. For this reason it may therefore behove us to cleave to common usage by adopting a negative social account of autonomy. After all, different terms—'authenticity', 'agential competence', 'moral responsibility'—are available to pick out the other notions of theoretical interest, and eliminating 'autonomy' from these discussions may help to focus and to clarify our investigations.

In this paper I have argued that the idea of self-rule is not obviously best understood in the usual way, as requiring that one be ruled by one's 'authentic' self. Instead, it may be that self-rule is better understood negatively, as requiring only the absence of rule from without. Moreover, there are good reasons for adopting a particular instance of the negative view, on which self-rule requires merely the absence of rule by other agents. This is a negative social view, and it represents a promising approach to the idea of individual self-rule.

<sup>13</sup> Note the use of the term, for instance, in Spector 1986 and in Datta and Grant 1990.

<sup>14</sup> See, for instance, Boud 1981: 'A collection of essays that examines ways in which teachers in higher education can enable students to become more autonomous in their learning: that is, how students can learn without the constant presence or intervention of a teacher.'

<sup>15</sup> And some may wish to challenge the lesson I draw from these considerations. For instance, it might be expected that few would be willing to accord full autonomy to a distressed lunatic screaming in the street (as we must on a negative social account). But consider this: when the new Director of a mental institution declares that he supports an increase in 'patient autonomy', this is not normally just a roundabout way of saying that he wants his patients to get better. Rather, it more likely means that he thinks that patients should have more of a say in how they are treated, and should be less subject to the rule of the doctors and nurses.

**Acknowledgments** Early versions of this paper were presented at both the Birkbeck work-in-progress workshop and the Birkbeck postgraduate seminar: I thank all those in attendance for the helpful discussions. Many thanks are also due to various anonymous referees for invaluable comments on previous drafts.

## References

- Blackford LB, (2010) Panel Approves Compromise Bill to Make College Transfers Easier, *Lexington Herald-Leader*, March 19.
- Boud D (1981) ed. *Developing Student Autonomy in Learning*, Nichols Publishing Company
- Buss S (2005) Valuing Autonomy and Respecting Persons: Manipulation, seduction, and the basis of moral constraints. *Ethics* 115(2):195–235
- Chishi H (2010) NSCN-IM Rejects Autonomy. *The Calcutta Telegraph*, March, 19
- Christman J (1991) Autonomy and Personal History, *Canadian Journal of Philosophy* 21.
- Christman J (2007) Autonomy, History, and the Subject of Justice, *Social Theory and Practice* 33.
- Christman J (2009) *The Politics of Persons: Individual autonomy and socio-historical selves*, Cambridge University Press.
- Cohen GA (1996) Reason, Humanity, and the Moral Law. In C. Korsgaard, *The Sources of Normativity*, Cambridge University Press.
- Datta D, Grant J (1990) Relationships Between Type of Acquisition, the Autonomy Given to the Acquired Firm, and Acquisition Success: An empirical analysis, *Journal of Management* 16.
- Ekstrom LW (1993) A Coherence Theory of Autonomy. *Philos Phenomenol Res* 53:599–616
- Ekstrom L (2005) Autonomy and Personal Integration, in *Personal Autonomy*, ed. J. S. Taylor, Cambridge University Press.
- Frankfurt HG (1971) Freedom of the Will and the Concept of a Person. *J Philos* 68(1):5–20
- Frankfurt HG (1975) Three Concepts of Free Action. *Proc Aristot Soc* 49:113–25
- Frankfurt HG (1999) Autonomy, Necessity, and Love, in *Necessity, Volition, and Love*, Cambridge University Press.
- Frankfurt HG (2002) Reply to Jonathan Lear, in *Contours of Agency: Essays on themes from Harry Frankfurt*, ed. S. Buss and L. Overton, Massachusetts Institute of Technology Press.
- Friedman M (2003) *Autonomy, Gender, Politics*, Oxford University Press.
- Haselager WFG (2005) Robotics, Philosophy and the Problems of Autonomy. *Pragmat Cogn* 13:518
- Hobbes T (1996) *Leviathan*, ed. R. Tuck, Cambridge University Press.
- Hoover K (2010) Consumer Bureau at Federal Reserve would have Autonomy, *Phoenix Business Journal*, March 19
- Korsgaard CM (2010) *Self-Constitution: Agency, identity, and integrity*, Oxford University Press.
- McKenna, 2005. The Relationship between Autonomous and Morally Responsible Agency, in *Personal Autonomy*, ed. J. S. Taylor, Cambridge: Cambridge University Press.
- Mele AR (1995) *Autonomous Agents: From self-control to autonomy*, Oxford University Press.
- Mele AR (2006) *Free Will and Luck*, Oxford University Press.
- Mele AR (2008) Manipulation, Compatibilism, and Moral Responsibility. *The Journal of Ethics* 12(3–4):263–86
- Meyers DT (1989) *Self, Society, and Personal Choice*, Columbia University Press
- Meyers DT (2000) Intersectional Identity and the Authentic Self? Opposites Attract!, in *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*, ed. C. Mackenzie and N. Stoljar, Oxford: Oxford University Press.
- Moran R (2002) Frankfurt on Identification: Ambiguities of activity in mental life, in *Contours of Agency: Essays on themes from Harry Frankfurt*, ed. S. Buss and L. Overton, Massachusetts Institute of Technology Press.
- Noggle R (1995) Autonomy, Value, and Conditioned Desire. *Am Philos Q* 32(1):57–69
- Plato (1987) *Republic*, trans. D. Lee, Penguin Books.
- Scanlon TM (2002) Reasons and Passions, in *Contours of Agency: Essays on themes from Harry Frankfurt*, ed. S. Buss and L. Overton, Massachusetts Institute of Technology Press.
- Spector, P. E. 1986. Perceived Control by Employees: A meta-analysis of studies concerning autonomy and participation at work, *Human Relations* 39.
- Velleman JD (1992) What Happens When Someone Acts? *Mind* 101:461–481
- Velleman JD (2002) Identification and Identity, in *Contours of Agency: Essays on themes from Harry Frankfurt*, ed. S. Buss and L. Overton, Massachusetts Institute of Technology Press.
- Watson G (1975) Free Agency. *J Philos* 72:205–20
- Williams R (2010) State Schools Accused of Turning Pupils into ‘Well-Drilled Automaton’, *The Guardian*, March 19.
- Wolf S (1990) *Freedom Within Reason*, Oxford University Press.