

AUTONOMY AND THE PROBLEM OF SOCIALISATION

Suzy Killmister

University of Connecticut / Massey University

Forthcoming in *Social Theory and Practice*

1: The Problem

Autonomous persons are self-governing. That much, at least, can be agreed upon. What it means to be self-governing, however, is far from straight-forward. One of the most intractable problems in the debate over how to conceive of autonomy revolves around the problem of socialisation. Autonomous action should spring from the self.¹ Yet that same self often appears to be the product of external influences. This causes an inescapable tension in determining the conditions for an act to be considered autonomous. While many contemporary theories of autonomy have long since abandoned the quest for a pure noumenal self, and rejected the model of rugged individualism that it inspires, they still face the problem of how to conceptualise autonomy such that we can differentiate socialisation that is compatible with autonomy from socialisation that is not.² It is not enough simply to note that we are all the products of our environment, since some environmental factors seem very much to compromise autonomy; for instance hypnosis, brain-washing and perhaps oppressive socialisation. What is required, then, is a principled way to determine

¹ My concern in this paper is with 'local' as opposed to 'global' autonomy. That is, I am concerned whether in a given instance an individual acts, desires, or intends autonomously, rather than whether they count more generally as an autonomous agent.

² In this paper I assume a compatibilist framework.

the boundary between benign and pernicious socialisation.³ This requires having a plausible conception of what autonomy is, and ascertaining the mechanism by which some forms of socialisation, but not others, compromise that autonomy.

Given the centrality of the problem of socialisation to the issue of autonomy, it is perhaps surprising to discover the lack of progress that has been made towards resolving it. In fact, as I will show in Section Two, even those theories of autonomy which explicitly recognise and engage with the problem of socialisation either offer problematic determinations of personal autonomy, or if their determinations do seem plausible, fail to explain exactly *how* pernicious socialisation succeeds in compromising autonomy. This latter is vital, because for a theory of autonomy to successfully address the socialisation problem it must be able to show how the necessary and sufficient conditions of autonomy, as understood by that theory, are compromised by some forms of socialisation but not others.

Since the theories I find lacking cover the spectrum from purely procedural to strongly substantive, the discussion to this point may lead to scepticism over the potential for any theory to adequately deal with the problem of socialisation. Such a conclusion, however, would be overly hasty. In Section Three I sketch a theory of autonomy that can both differentiate benign from pernicious socialisation and, more importantly, provide an explanation for the *means* by which pernicious socialisation compromises autonomy. In Section Four I road-test the theory.

1.a Hypotheticals

³ This is not to imply that socialisation is *at best* benign. Certain forms of socialisation may in fact be necessary for the development of autonomy. My purpose in this paper, however, is in understanding how some forms of socialisation undermine autonomy, and so the relevant boundary is that between benign and pernicious socialisation.

To motivate this discussion, I begin by sketching two hypothetical cases which will serve as a reference point throughout this paper. I openly acknowledge that, as a heuristic tool, hypotheticals can be problematic. Intuition pumping presumes that our intuitions are the appropriate guide to determining the autonomy or otherwise of a given action, whereas they may in fact lead us astray. As such, these hypotheticals are not offered as definitive examples of autonomous and non-autonomous action. Rather, they serve as an entry point to consider the adequacy of each theory's response to the situations described. This adequacy does not turn on the diagnosis of autonomy they offer, but rather the strength of justification for that diagnosis.

Case A: Julie is a young woman who has recently entered a relationship with a significantly older man. One morning Julie's partner suggests that he would find her more sexually attractive if she were to have larger breasts. Julie reflects on her situation and resolves to end the relationship.

Case B: Meredith is a young woman who has recently entered a relationship with a significantly older man. One morning Meredith's partner suggests that he would find her more sexually attractive if she were to have larger breasts. Meredith reflects on her situation and resolves to undergo surgery for breast implants.

I am assuming that most people's intuitions would tend towards ascribing autonomy to Julie in Case A, whereas they would at least entertain some doubts about Meredith's autonomy in Case B. These scenarios are admittedly under-described. This is deliberate. One of the ways to assess a theory of autonomy is to consider what extra information would be required to determine whether or not a given scenario reflects autonomous or non-

autonomous action. One aspect of my critique of the range of theories of autonomy currently dominant in the field is that they would ask the wrong questions about Julie and Meredith's respective situations.

2. The State of Play

I will now consider four very different theories of autonomy, and show that each fails to adequately deal with the cases illustrated in Section One above. First I consider the reflective endorsement approach of John Christman. I argue that this approach, while more promising than is often believed, provides insufficient resources to fully address doubts about Meredith's autonomy. Second, I consider the externalist historical approach of Al Mele, arguing that his approach may have the unintended consequence of falsely attributing non-autonomy to Julie. Third, I consider the strongly substantive account of Natalie Stoljar. I argue that while her account has the virtue of mapping onto our intuitions, it fails to adequately explain *how* Meredith's autonomy may be compromised. Finally, I consider the weakly substantive accounts recently put forward separately by Andrea Westlund and Paul Benson, both of whom identify autonomy with the disposition to claim authority for one's actions. I claim that such an approach may offer a useful heuristic for judging whether an action is autonomous, but fails to capture the conceptual substance of autonomous action.

2a. Reflective Endorsement

Reflective endorsement theories are a subset of proceduralist theories, which define autonomy in terms of the psychological dispositions of the agent. Within the reflective

endorsement camp, John Christman's theory is perhaps the best situated to address the problem of socialisation. This is because his model of reflective endorsement was explicitly developed to overcome the inadequacies of some earlier models. Christman acknowledges that the second-order desire theories of Gerald Dworkin and Harry Frankfurt are blind to the autonomy-compromising nature of manipulation. Such theories require merely that an individual's second-order desires align with her first order-desires. However, effective manipulation can reach all the way to second-order desires – an individual can be manipulated 'all the way down'. To judge as autonomous individuals who are hypnotised, brain-washed, or otherwise manipulated into developing second-order desires, so the argument goes, is to misunderstand the very nature of autonomy.

Christman attempts to remedy this problem through the expansion of reflective endorsement to the history, rather than just the content, of the desire.⁴ For Christman, the necessary endorsement must be directed at the *process of desire-formation*, rather than the desire itself. The individual must not feel alienated from the desire, were she to reflect on it in light of its formation. For instance, we can take an individual who desires to eat a slice of cheesecake. Rather than asking, as Dworkin and Frankfurt would, whether the individual *desires to desire* to eat a slice of cheesecake, Christman asks how she would feel about her desire, were she to be aware of how the desire were formed. If the desire were implanted by a hypnotist, and knowledge of this process caused the individual to feel alienated from the desire, then that desire would not be autonomous.

⁴ John Christman, 'Autonomy and Personal History', *Canadian Journal of Philosophy*, 21/1 (1991); 'Liberalism, Autonomy, and Self-Transformation', *Social Theory and Practice*, 27/2 (2001); 'Autonomy, History, and the Subject of Justice', *Social Theory and Practice*, 33/1 (2007); *The Politics of Persons: Individual Autonomy and Socio-Historical Selves* (Cambridge: Cambridge University Press, 2009).

The objections to Christman's theory are well rehearsed.⁵ Nonetheless, it will be useful to sketch these standard concerns with reference to the hypotheticals described in Section One, in order to orient the discussion. We can assume that – absent any bizarre additional information – Christman would diagnose Julie's decision as autonomous. In order to determine whether Meredith were autonomous, however, Christman would need more information. In particular, Christman would need to ask whether Meredith would feel alienated from her desire for breast implants, were she to reflect on that desire in light of its history. It is here that the theory has been assumed to run into difficulties. We might predict that Meredith would feel alienated from her desire were she to learn that it is the product of patriarchal norms of feminine beauty that she has internalised. However, if the internalisation has been successful, then Meredith may in fact endorse the patriarchal socialisation that has shaped her conception of beauty.⁶ The problem for Christman's theory is taken to be that effective manipulation will include the development of pro-attitudes towards the process, as well as the content, of that manipulation.

While powerful, I believe that such criticisms are ultimately misguided. In refusing to accept the perspective of the deeply socialised agent as authoritative, these criticisms implicitly assume that there must be a true, or authentic, pre-social self lurking within. In defence of Christman, an individual who has only been partially indoctrinated (i.e. indoctrinated to believe a particular set of norms, such as patriarchal ones) would presumably be able to reflect upon, and critique, those norms from the perspective of her broader beliefs and

⁵ See, i.e., Paul Benson, 'Autonomy and Oppressive Socialization', *Social Theory and Practice*, 17/3 (1991); Catriona Mackenzie and Natalie Stoljar, 'Introduction: Autonomy Reconfigured', in Mackenzie and Stoljar (eds), *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self* (New York: Oxford University Press, 2000).

⁶ See Benson, 'Autonomy and Oppressive Socialization'.

values. In other words, she could undergo a process of reflective equilibrium, assessing the compatibility of the patriarchal norms with her other values and commitments. If such an agent reconciles the patriarchal norms with her wider set of beliefs, and incorporates them into her overarching worldview, then it seems plausible to assume that they have become authentic for her. If we try instead to imagine an agent *all* of whose beliefs and values have been indoctrinated, then we must either bite the bullet and accept that these beliefs and values are now *constitutive* of the self, and thus in principle consistent with autonomy, or search for a ‘true’ self hidden within. Plugging for the second horn of this dilemma, I take it, reveals an implicit commitment to the possibility of a pre-social self.⁷ While this may seem acceptable to some incompatibilists, it is a position that is for the most part explicitly rejected by those who have object most vociferously to Christman’s theory.

It must be stressed that taking the first horn of this dilemma, as I believe we should, is compatible with placing additional conditions on autonomy; there may still be a reason to deny that Meredith is fully autonomous, even if she reflectively endorses her desire for breast implants in accordance with the conditions laid out by Christman. Indeed, I will be arguing in Section Three that Christman’s conditions are insufficient in two senses: first, his criteria for reflective endorsement are not strong enough, and need to be supplemented by an additional condition; and second, reflective endorsement, even when strengthened, is itself insufficient, and must be supplemented.

It is worth noting at this point that Christman himself appeals to an additional, non-subjective, condition of autonomy: “The lack of resistance to the development of D did not

⁷ Thanks to an anonymous reviewer for this journal for helping me to see this point clearly.

take place (or would not have) under the influence of factors that inhibit self-reflection”⁸, with this later clarified to mean “freedom from those factors and conditions that we independently know effectively prevent minimal self-understanding”.⁹ Such a condition allows for there to be interruptions to autonomy that are not mitigated by the reflective endorsement of the individual.¹⁰

Perhaps because such conditions play a largely subsidiary role in Christman’s theory, they are not spelled out in detail. While he gives some examples of relevant factors (uncontrollable rage, hallucinogenic drugs, lack of minimal education), he also alludes more broadly to ‘manipulation’. However, he offers no further explanation as to what is to be considered a case of manipulation. Not only do we not know whether Meredith would count as manipulated, more importantly we do not know what additional information we would need to make that determination. While Christman’s theory is promising, it does not yet ask the right questions about an agent such as Meredith.

I have suggested that the key problem with Christman’s theory is not his inclusion of a historical version of reflective endorsement, as is commonly thought, but rather that this condition is insufficient to determine the autonomy of an agent. Something more is needed to make that determination. I will offer my own suggestions for additional conditions in Section Three. Before doing so, however, it will be useful to examine alternative solutions to the problems raised by Christman’s theory. First, I consider a theory that may be thought

⁸ Christman, ‘Autonomy and Personal History’, p.11.

⁹ Christman, *The Politics of Persons*, p.147.

¹⁰ It is admittedly somewhat difficult to see how this can be reconciled with Christman’s earlier claim that there are no external and objective criteria for autonomy: “Imagine any process which you stipulate as resulting in the subversion of autonomy, and I can imagine a person who (given a fantastical enough hypothetical situation) would want her choices formed that way, and this desire is itself an autonomous one”. ‘Autonomy and Personal History’, p.21.

to address the very problem Christman's supplementary criteria are aiming at, namely to identify processes of desire formation that are objectively incompatible with autonomy. Such a theory has been developed by Al Mele, and it is to that theory that I now turn.

2b. Historical Conditions

At the centre of Mele's theory of autonomy is a hypothetical counter-example to procedural theories (i.e. those theories which only consider the structure of an agent's dispositions and are thus purely concerned with the 'internal' profile of the individual's psychology).¹¹ Mele asks us to imagine two philosophers, Ann and Beth. While Ann enthusiastically dedicates herself to her profession, eschewing outside interests to commit long hours to her work, Beth values having a rounded life, and thus limits the hours she is willing to commit to philosophy. The Dean would prefer Beth to be more like Ann, and so he arranges for a process of covert brain manipulation to be conducted on Beth, at the end of which her values and interests are identically aligned with Ann's. From this point on Beth, too, enthusiastically dedicates herself to philosophy, and eschews her previous interests in the name of greater productivity. Mele observes that internalist theories of autonomy are unable to account for the intuition that Beth is no longer autonomous. Since Ann and Beth are, Mele claims, 'psychological twins', a view that only looks to what's happening inside their heads cannot distinguish the autonomy of Ann from the heteronomy of Beth.

Mele's solution is to appeal to an external, historical, criterion: for an agent to be autonomous with respect to a pro-attitude P, it must not be the case that P is compelled*. 'Compulsion*' differs from 'compulsion' for Mele in that, by definition, 'compulsion*

¹¹ Alfred Mele, *Autonomous Agents: From Self-Control to Autonomy* (New York: Oxford University Press, 1995) pp.145-46.

cannot have been arranged by the agent. (This can be contrasted with Odysseus being ‘compelled’ by his crew to remain on board when passing the sirens.) Compulsion*, for Mele, refers to a process in which a pro-attitude is instilled in an agent through a process which bypasses the agent’s capacities for self-control, and results in a pro-attitude the agent is practically unable to shed. More precisely, compulsion* identifies situations in which an agent (i) comes to have a pro-attitude because of an external force, rather than via the exercise of her skills for critical reflection and evaluative judgment; and (ii) the instilled pro-attitude is one she is unable (in the absence of radical counterfactuals) to eradicate or attenuate; and (iii) she did not arrange the bypassing herself; and (iv) she does not, nor did she earlier, possess other pro-attitudes that would support her endorsing the instilled pro-attitude (unless those endorsing pro-attitudes were themselves compelled*).¹²

There is an immediate and decisive problem with utilising Mele’s condition as a guide for identifying autonomy-inhibiting socialisation: it would rule out much of early childhood education, including education directed at developing the very capacities necessary for autonomy.¹³ According to Mele, the possession of a pro-attitude is compelled* if its acquisition by-passes capacities for self-control. If the relevant capacities are inoperative or not yet developed, as they will be in very young children, Mele takes them to have been by-passed.¹⁴ Mele observes in passing that such by-passing will be common in young children, but notes that this is not sufficient for compulsion* unless the pro-attitude is also practically unsheddable. The example Mele gives to illustrate a pro-attitude that is compelled* in early

¹² Ibid. p.175-176.

¹³ Stefaan Cuypers develops a similar argument against Mele, though he does not fully engage with Mele’s further criterion that the pro-attitude in question has to be practically unsheddable. Stefaan Cuypers, ‘The Trouble with Externalist Compatibilist Autonomy’, *Philosophical studies*, 129/2 (2006).

¹⁴ Mele, *Autonomous Agents*, p.167-168.

childhood is a case of religious indoctrination, where the child develops deep-seated religious beliefs that cannot be shaken due to the parent's graphic portrayals of the Hellish fate of those who question the existence of God. In such a case, the child's abilities to assess and believe on the basis of evidence are circumvented – these abilities are not utilised in the adoption or retention of her religious beliefs.

While I agree with Mele's diagnosis that this example should count as autonomy-inhibiting socialisation, my concern is with how wide Mele's net will need to be cast. Many of us have pro-attitudes that we are practically unable to shed, and that can be traced to early childhood socialisation: our love for our parents; our trust in scientific and mathematic proof; our respect for the needs of others. While some of these may be innate, others are more plausibly thought of as the product of socialisation. Of most relevance here is the fact that attitudes towards reasons, evidence, and reflection are both practically unsheddable and presumably inculcated. They are unsheddable – as Mele understands that term – because they are fully integrated into our broader attitudes towards ourselves and the world around us. An individual who has internalised the minimal demands of rationality in decision-making, for instance, will be no more able to shed that attitude than a parent can shed her love for her children. While some external event may precipitate such shedding (we might imagine some kind of spiritual epiphany, perhaps), the agent cannot bring such an event on herself. She is, in Mele's terms, 'stuck with the values'.

The problem here is that one of the central roles of socialisation is to inculcate the very pro-attitudes that are required for the kind of self-control Mele sees as necessary for

autonomy.¹⁵ To be self-controlled is to believe and desire on the basis of an assessment of evidence. This will rely upon having the appropriate pro-attitudes towards evidence and reasons, which must at some point have been instilled in the child. While it may be the case that the child will later come to endorse those evidentiary pro-attitudes, they will still count as compelled* on Mele's account because the later process of endorsement *must rely on pro-attitudes that were themselves compelled**. To explain: Mele provides something of an 'escape clause' for early-childhood socialisation in the fourth clause of his definition of compulsion* outlined above, namely that the compelled* pro-attitudes *are not endorsed by other non-compelled pro-attitudes*. This may be thought to lead to the conclusion that the pro-attitudes necessary for autonomy are not compelled, because they could have been endorsed by the maturing child. However, the process of endorsement itself relies on certain attitudes towards belief and evidence, which must at some point have been inculcated in the not-yet-reasoning child, and must thus count as compelled*.¹⁶

The kind of self-control that Mele sees as necessary for autonomy has to be cultivated, and one of the key ways it is cultivated is through a process of socialisation that instils pro-attitudes towards reasons and beliefs in children, before they are yet able to reason and believe on the basis of evidence (indeed, the whole point is to *enable* children to reason and believe on the basis of evidence). We can see this problem at work if we return to the case of Julie. For argument's sake, let's imagine that Julie acts on the basis of her commitment to the claim that a woman's worth is not determined by her physical appearance. This is a commitment that her parents took great care to instil in her when she was too young to

¹⁵ Thanks to an anonymous reviewer for this journal for pushing me on this point.

¹⁶ This line of argument undercuts Mele's response to Cuypers, which focuses on Cuypers' obfuscation of the unsheddability condition. See Mele 'Manipulation, Compatibilism, and Moral Responsibility', *Journal of Ethics*, 12 (2008).

reflect on it on the basis of evidence. Moreover, this commitment is now, for her, practically unsheddable. The only way that this pro-attitude could count as noncompelled*, on Mele's criteria, is for it to be endorsed by other noncompelled* pro-attitudes. An obvious candidate would be Julie's pro-attitudes towards reasons and evidence – when a mature Julie critically assesses the merits of the claim that a women's worth is not determined by her physical appearance, she finds it to be highly reasonable. The problem is that this process of critical endorsement is tainted by the compelled* attitudes to reason and evidence that Julie's parents carefully cultivated in her as a child. The conclusion Mele would have to draw about Julie is thus problematic. Insofar as Julie acts on the basis of reasoning capacities that were instilled in her as a child (and thus bypassed her capacities for self-control), she is not acting autonomously. If a theory determines that an individual's autonomy is undermined *just because* she is utilising her capacities for critical reasoning, then it cannot resolve the problem of socialisation.

2c. Substantive Constraints

Christman's theory proved unhelpful in the quest to explain autonomy-compromising socialisation because it did not provide sufficient conditions for autonomy, and would thus classify as autonomous agents who are not. Mele's theory proved unhelpful because compulsion* over-generalised, classifying as non-autonomous agents who are. Since neither our attitude towards our histories, nor our lack of participation in the process of value-acquisition, resolve the problem of autonomy-inhibiting socialisation, a third pathway

could perhaps be seen in the actual content of socialisation. Such a theory is put forward by Natalie Stoljar.¹⁷

Stoljar's conception of autonomy is initially very promising for our purposes. She explicitly acknowledges that "the question for all theories of autonomy is what kinds of socialization are incompatible with autonomy"¹⁸, and she promises to provide such a theory through rejecting the content-neutrality of proceduralist approaches, and replacing it with a substantive criterion that "places restrictions on the content of agents' preferences".¹⁹ This should presumably be taken to mean that we can identify autonomy-inhibiting socialisation through reference to the inculcation of specified preferences. Such a theory, if successful, would provide just the kind of concrete, principled guidelines necessary to solve the problem of socialisation. For instance, it could explain which kinds of preferences Meredith would need to be motivated by for her actions to be non-autonomous. Unfortunately, Stoljar's theory does not fulfil this promise.

According to Stoljar, proceduralist theories of autonomy cannot account for what she calls the 'feminist intuition'. This amounts to the intuition that those who have been subject to oppressive socialisation are not autonomous. Stoljar cashes this intuition out in terms of oppressive socialisation leading to the internalisation of false beliefs. Drawing on Paul Benson's early work, the claim is that decisions made on the basis of false beliefs are not autonomous: "Women who accept the norm that pregnancy and motherhood increase their worthiness accept something *false*."²⁰

¹⁷ Natalie Stoljar, 'Autonomy and the Feminist Intuition', in Mackenzie and Stoljar (eds.), *Relational Autonomy*.

¹⁸ *Ibid.*, p.97.

¹⁹ *Ibid.*, p.95.

²⁰ *Ibid.*, p.109.

Admittedly, there is some ambiguity in Stoljar's paper as to whether the problem is exclusively with the *falsity* of these women's beliefs, or rather includes their *oppressiveness*. For instance, in describing the shifting intuitions she predicts towards the attribution of autonomy in smokers, Stoljar points out that "when smokers [...] do opt to smoke on the basis of false *and oppressive* norms [...] the intuition that they are not autonomous is triggered." Likewise, she explains the intuition that Luker's pregnant subjects are not autonomous on the grounds that "the internalized norms motivating the decision to take a contraceptive risk *have criticizable content*".²¹

Although it is unclear whether the problematic beliefs must be *both* false and oppressive, or if either conjunct is sufficient, the distinction is negligible for my purposes.²² When applied to our two cases it seems that on either interpretation, Stoljar's theory would account for our intuitions. Plausibly, Julie is acting on the basis of true and non-oppressive norms. Equally plausibly, Meredith is acting on the basis of false and oppressive norms. One of the benefits of Stoljar's theory, moreover, is that it does not necessitate the conclusion that Meredith's decision was non-autonomous. It is at least *conceptually* possible (though admittedly unlikely) that Meredith's decision could be motivated by norms that are neither false nor oppressive. By appealing to the reasons behind a decision, Stoljar's theory does not, contrary to the worries of many of her critics, pre-emptively rule out certain decisions or life-style choices as non-autonomous.²³

²¹ Ibid., pp.99-100, emphasis added.

²² I address the question of whether false beliefs alone can undermine autonomy in (...).

²³ Diana Meyers' worry is typical: "Restrictive, value-saturated accounts of autonomy are troubling because they promiscuously stigmatize women as victims and because they homogenize authentic selves and autonomous lives". *Gender in the Mirror: Cultural Imagery and Women's Agency* (New York: Oxford University Press, USA, 2002) p.160.

However, Stoljar's theory does not fulfil the desiderata set out at the beginning of this paper. This is because, as it stands, it leaves open the question as to *how* false and/or oppressive beliefs affect autonomy. Such an explanation is necessary to provide a principled justification of our intuitions. In order to fulfil this demand, we would need to know both what the constitutive features of autonomy are, and the way in which the internalisation of false/oppressive norms undermines or compromises those features. Stoljar's theory, however, provides no way to do this. On her account, the possession of such beliefs *per se* constitutes non-autonomy. This simply begs the question: oppressive socialisation undermines autonomy because autonomy is constituted by the absence of oppressive socialisation.

What Stoljar's theory is missing is an analysis of autonomy that explains *how* the belief in question undercuts autonomy, which in turn requires specification of some feature of autonomy that the belief can undercut. Once more, a theory of autonomy that promised to resolve the problem of socialisation has fallen short. In this case, we can attribute the failure to Stoljar's reductive strategy: if autonomy simply reduces to being free of oppressive socialisation, then it cannot provide a principled explanation as to *how* such socialisation undermines autonomy.

2.d The Authoritative Stance

The final view to consider is that put forward, in subtly different forms, by both Andrea Westlund and Paul Benson. Westlund and Benson maintain that autonomy is not to be understood as a structural feature of the agent's psychology, nor in terms of the agent's history. Rather, autonomy is constituted by a particular kind of disposition: a disposition of

responsiveness to normative pressure, which expresses itself through a willingness to stand behind, or speak for, one's actions.

According to Westlund, if an individual is unwilling or unable to give reasons in defence of her choices or actions, she fails to be autonomous. As she puts it, "Readiness to answer for oneself is, on this view, no mere *sign* that one is self-governing with respect to one's relevant practical reasoning. It is (at least a crucial part of) what *constitutes* one as self-governing."²⁴ Westlund motivates her view through exploration of the character of the Deferential Wife (DW). DW is characterised as subordinating herself to her husband in a very specific way: she does what he wants, because he wants it; and moreover, when pushed to explain her actions can only do so via reference to *his* reasons.

Westlund sees her approach as offering a more plausible diagnosis for why DW lacks autonomy than can be offered on either a procedural or a historical account. When applied to our original hypotheticals, Westlund's account does indeed seem at first glance to be able to provide a more plausible diagnosis. If Meredith can offer reasons of her own for her decision to get implants, then she is autonomous. If, instead, her only response is to defer to the reasons of her partner, she is not. Westlund's account thus promises an explanation of what autonomy is, and provides scope for exploring how certain forms of socialisation may undermine it. Unfortunately, the account cannot withstand further scrutiny, for two reasons.

First, Westlund's stress on being able to offer reasons for one's choices or actions threatens to rule out too much. Why becomes clear when we consider a further character to whom

²⁴ Andrea Westlund, 'Selflessness and Responsibility for Self: Is Deference Compatible with Autonomy?', *The Philosophical Review*, 112/4 (2003), p.503.

Westlund introduces us: Anti-Feminist (AF). AF shares DW's attitudes towards subservience; however, if pressed, she is able to articulate her own reasons for adopting this position. The problem occurs when Westlund considers what would happen if AF's reasons simply defer to the authority of a text such as the Bible. In such cases, Westlund deems her non-autonomous: "Such an interlocutor does not stand ready to be engaged in a potentially open-ended exchange of reasons, but instead, offers what amounts to a verbal expression of her resistance to such engagement."²⁵ What Westlund also refers to as 'pat answers', however, may be resistant to justification for good reason. If these are an agent's foundational beliefs, it may well feel to her that there is nothing more that she can do other than repeat or rephrase them. Foundational beliefs such as that God exists, or that all life is precious, do not always allow for deeper justification (or at least they may not present themselves as doing so, particularly to the less disputatious among us). These kinds of beliefs may manifest precisely the attitudes that Westlund takes to be constitutive of non-autonomy: an unwillingness to entertain challenges to the view, and consequently a lack of the necessary "disposition to be engaged by critical challenges to the commitments that underlie those judgments."²⁶

Westlund's theory commits us to the stance that actions underpinned by faith or foundational belief are non-autonomous. In fact, on Westlund's account it seems that the deeper the justification for an action lies within our motivational structure, the less that action is autonomous. To return to our initial hypotheticals, we can now see that Westlund confronts a similar problem to Mele. If Julie responds to a request for reasons with

²⁵ Ibid. pp.513-14.

²⁶ Ibid. p.508.

reference to her unshakeable belief that women should not mutilate their bodies, and is unwilling or unable to engage with further challenges to that commitment, she must be considered non-autonomous. The more committed she is to that stance, the more it forms an authentic part of her self-understanding, the less Westlund's account will deem her autonomous. Such a counter-intuitive conclusion raises grave doubts about Westlund's account.

Nonetheless, a very similar account may have the resources to fare better. Benson has developed a view that is structurally related to Westlund's.²⁷ He, too, holds autonomy to be constituted by a self-regarding attitude that expresses itself in a willingness to take responsibility for one's actions. However, for Benson that willingness is disconnected from the giving of reasons for our actions, and instead focuses on the act of *claiming authority* for our actions. Claiming authority involves standing behind, and speaking for, actions that we have performed. Since Benson accepts that we can claim authority even for actions that are trivial, or over which we are deeply ambivalent, we presumably do not need to be willing to engage in an open-ended justificatory dialogue. Our authority is claimed when we answer for the action, not when we have attempted to justify the action. In other words, it seems that on Benson's account it should be sufficient merely to claim 'because I wanted to' in response to an interlocutor. Claiming the action as one's own, and expressing one's willingness to stand responsible for it, is (at least part of) what constitutes autonomy.

While Benson's variation promises to overcome the limitation of Westlund's view explored above, it is still faced with a formidable obstacle. The problem for Benson – and for

²⁷ Paul Benson, 'Taking Ownership: Authority and Voice in Autonomous Agency', in John Christman and Joel Anderson (eds.), *Autonomy and the Challenges to Liberalism: New Essays* (Cambridge: Cambridge University Press, 2005).

Westlund too, it should be noted – is that individuals can be willing to stand behind actions even when these actions are paradigmatically non-autonomous. This can best be illuminated through Albert Moll’s famous study of hypnosis. Moll relates the following anecdote:

I tell a hypnotized subject that when he wakes he is to take a flower-pot from the window, wrap it in a cloth, put it on a sofa, and bow to it three times. All which he does. When he is asked for his reasons he answers, ‘You know, when I woke and saw the flower-pot there I thought that as it was rather cold the flower-pot had better be warmed a little, or else the plant would die. So I wrapped it in the cloth, and then I thought that as the sofa was near the fire I would put the flower-pot on it; and I bowed because I was pleased with myself for having such a bright idea.’²⁸

Clearly, the individual in this anecdote is not acting autonomously. Yet he is willing to not only claim authority for his actions, but also provide extensive justifications for them. It should be noted that even if this anecdote is not entirely reliable, it points towards a phenomenon that is widely accepted: we humans are very good at inventing post-facto explanations of our actions, even to ourselves.²⁹ If Meredith is acting on the basis of hypnosis, or perhaps an undiagnosed physiological condition, when she agrees to a breast implant, then provided she is willing and able to conjure up some kind of explanation then it looks like both Benson and Westlund would have to deem her autonomous. This strongly suggests that self-authorization alone is insufficient for autonomy.

²⁸ Albert Moll, *Hypnotism* (Kesinger Publishing, 2004) pp.153-54.

²⁹ See Cordelia Fine, *A Mind of Its Own: How Your Brain Distorts and Deceives*, (W.W. Norton, 2008)

To his credit Benson has acknowledged this, and (in contrast to Westlund) seeks to supplement his self-authorization requirement with a range of objective conditions, ruling out circumstances such as substantial lack of information or mind control.³⁰ The problem with this solution is that it implies that the question of what constitutes autonomy starts to look like it is better answered with direct references to these objective considerations, rather than the act of claiming authority. If agents' autonomy can be determined separately from their claiming of authority – either because they claim authority when it is not due, or fail to claim it when it is, both of which Benson acknowledges as possible – then it is hard to see precisely what work is left to be done by the act of claiming authority. In other words, once objective criteria are invoked to supplement the account, these criteria threaten to replace the core of the theory. Moreover, the onus is now on the theorist to explain exactly what justifies these additional criteria: what, precisely, constitutes mind-control? How much information is required for an action to be autonomous? We are given no reason to think that the answers to these questions will make reference to any facet of the claiming of authority, suggesting that this phenomenon is in fact peripheral to the concept of autonomy.

What appeal there is in Benson's theory perhaps reflects the fact that such attitudes are commonly correlated to autonomy. In the concluding section I suggest some factors which may explain the at least partial reliability of Benson's theory as a determinant of autonomy, but also explain why his theory is not directly tracking autonomy.

³⁰ Benson, 'Taking Ownership', p.118.

3: An Alternative Framework

We have now examined four starkly different theories of autonomy, each of which attempts to resolve the problem of socialisation, and each of which falls short. This may leave us feeling somewhat pessimistic about the possibility of solving this problem. What is required, to recap, is a theory which can lay out the necessary and sufficient conditions for autonomy, and in doing so show how some forms of socialisation support those conditions, while others compromise them.

In the remainder of this paper I will sketch an answer to the initial question of what autonomy is, and then suggest some ways certain forms of socialisation may compromise it. I argue for an understanding of autonomy as comprised of five conceptually distinct dimensions, each of which can be measured according to the extent it facilitates the realisation of the agent's will in a given action.³¹ This is not presented as a conclusive and comprehensive theory, but rather a starting point which provides sufficient content to enable the differentiation of autonomy-supporting and autonomy-compromising socialisation. I conclude by assessing how successfully my account can overcome the problems faced by the four theories outlined above.

3a. The Dimensions of Autonomy

To be autonomous is to be self-governed, it is to be *in control of* one's interactions with the world, and it is to have those interactions *correspond to* one's self. Paradigmatically

³¹ This is a preliminary account of what will be a much larger and more complex theory. My aim here is to develop the theory to the point where it is clear how the problem of socialisation may be solved. For a more detailed discussion of this theory, see (...) This understanding of autonomy as comprised of distinct dimensions shares some features with competency theories of autonomy. See, i.e., Diana Tietjens Meyers, *Self, Society, and Personal Choice* (New York: Columbia University Press, 1989), and Marilyn Friedman, *Autonomy, Gender and Politics*, (New York: Oxford University Press, 2003).

autonomous action, than, can be said to involve 'realising one's will in the world'. I suggest that this can best be understood in terms of an idealised three-step decision-making process: first, comprehension of the prospective action; second, critical reflection on whether to undertake the action; and third, resolution in intentional action. So for Meredith's act of getting breast implants to be ideally autonomous, she would need to comprehend what breast implants were; she would need to reflect on whether this act was in accordance with her values, commitments and desires; and this reflection would need to be resolved through the formation and execution of an intention to get breast implants. These steps are mutually interdependent, in the sense that we cannot resolve to act without first reflecting on the act, and we can do neither without first comprehending the existence of the prospective action. From this idealised three step process it is possible to discern five distinct dimensions of autonomy, each of which affects the degree to which a given action is autonomous.

There are two crucial caveats that must be stressed before this schema is explored in more detail. First, it would be a mistake to assume a thoroughly intellectualised theory of the decision-making process. Comprehension can be affective³²; reflection can be guided by the emotions and the imagination³³; and resolution may take place below the level of conscious awareness. It is no part of my theory that an action is non-autonomous because the agent did not consciously sit down and reason through the pros and cons of action.

³² Robin S. Dillon, 'Self-Respect: Moral, Emotional, Political', *Ethics*, 107/2 (1997), p.239.

³³ Karen Jones, 'Emotion, Weakness of Will, and the Normative Conception of Agency', *Philosophy*, 52 (2003); Catriona Mackenzie, 'Imagination, Identity, and Self-Transformation', in Catriona Mackenzie and Kim Atkins (eds.), *Practical Identity and Narrative Agency* (New York: Routledge, 2008).

Second, this approach requires rethinking autonomy in scalar terms, rather than as a threshold.³⁴ On my theory, any of the five dimensions can be realised to a greater or lesser extent, and its level of realisation can alter over time and according to context. The central implication of this is that, if we are examining cases such as Julie and Meredith, we must rethink the questions we are asking. Instead of demanding to know ‘is she autonomous?’, we must instead ask ‘how autonomous is she?’ Or, to put it in more accurate but convoluted terms, ‘to what extent are each of the five dimensions realised, in the decision at hand?’

With these caveats in mind, it is time to identify the five dimensions that I claim constitute autonomy, and which are discernable through reference to the three-stage process of comprehension, reflection, and resolution. It is important to bear in mind that each of these dimensions can be achieved to a greater or lesser extent. My claim is simply that, all else being equal, increasing the level of any of these dimensions will augment the agent’s autonomy with respect to the action at hand, and conversely, lowering the level of any of these dimensions will diminish the agent’s autonomy.

1. Knowledge of the World

Information affects the degree to which an action can be autonomous. If an agent has false beliefs about the act that she is performing, then her autonomy is compromised.³⁵ This can be seen with a basic example: Sally believes the liquid in the mug in front of her is coffee. Since she wants to stay awake to finish the paper

³⁴ Though there is not space to fully defend the point here, I also reject the idea that there is some point on the scale at which the agent becomes ‘fully’ or ‘properly’ autonomous. More plausible, I believe, is the claim that the purposes to which we put autonomy (i.e. as the basis for moral respect, or as a counter to paternalism) each appeals to a different level of autonomy. For further discussion, see (...)

³⁵ See Mele, *Autonomous Agents*, pp.179-182. I defend and extend Mele’s position in (...)

she is working on, she decides to drink it. Unbeknownst to her, the liquid is actually decaf, to which she has a strong and principled aversion. Since Sally's reasons for action do not map onto the action she performs, the action is less autonomous than it would have been if her information had been correct. Incorrect information reverberates through the decision-making process, undermining reflection (the agent reflects on an irrelevant, unavailable act) and resolution (the act she intends to perform is not the one she does perform). Extrapolating from this, we can say that the more correct and relevant information the agent has about the action she is contemplating performing, including the effects that will follow from that action, the more autonomous the action will be. (Relevance is determined through reference to the agent's desires, values, plans and commitments – her motivational attitudes. Information is relevant if it would play a role in enabling the agent to bring her actions in line with her motivational attitudes. I say more about this below.)

2. *Self-Knowledge*

If paradigmatically autonomous action is action that corresponds to the agent's self, then the agent is going to be more autonomous to the extent that she knows herself. More precisely, an agent's autonomy will be augmented to the extent that she knows what her motivational attitudes are (i.e. her desires, values, commitments and plans), and to the extent that she knows how she ranks those attitudes. Self-knowledge enables an agent to bring her actions into alignment with her motivational attitudes.

3. *Self-Definition*

Knowing oneself is a crucial component of autonomy, but even full transparency does not ensure that the motivational attitudes one finds within are reflective of one's will. As many have noted, an agent's autonomy is reduced if the motivational attitudes that move her to action are inauthentic; therefore, it is important that the agent's motivational attitudes are somehow made her own. On this matter I build on Christman's account of reflective endorsement. A motivational attitude is self-defined to the extent that the agent endorses retaining it in her motivational set, and this endorsement would persist in the light of full knowledge of the attitude's historical origin, *and* full knowledge of relevant facts about the world around her.

This is stronger than Christman's account in two senses. First, it requires positive endorsement rather than Christman's weaker non-alienation. This is in line with the scalar nature of my account – full endorsement would indicate greater autonomy than mere non-alienation, though both would register as augmenting rather than diminishing autonomy (unlike, say, alienation or disgust). Second, my account incorporates an additional knowledge condition regarding the external world. This latter is important to guard against forms of socialisation/indoctrination that embed acceptance of their own processes, for instance hypnosis that instils a positive attitude towards hypnosis. Knowledge of the external world would include knowledge of the motivation of the hypnotist, as well as knowledge of the role that hypnotism had historically played in manipulating agents into reshaping their desires. This would enable the agent to critically reflect on the validity of her endorsement – while she may still find herself feeling positively about hypnosis, she would be able to access reasons not to endorse that feeling.

4. *Internal Self-Realisation*

The dimensions outlined above correspond to the first two of the three-step process identified at the outset of this section, namely comprehension and reflection. Since autonomy refers to *self-governance*, and not merely self-analysis, the process thus far is still insufficient to fully capture the meaning of autonomy. If an agent reflects endlessly, without that reflection terminating in the formation of an intention, she has fallen short of the goal of realising her will in the world. In light of this, I propose that autonomy is augmented to the extent that the agent forms an intention that responds to and reflects her motivational set. Failure to form an intention, forming an intention that does not accurately reflect the agent's motivational set, or forming an intention that only accidentally reflects the agent's motivational set, will each indicate a reduction in the agent's autonomy.

5. *External Self-Realisation*

The final dimension of autonomy is one that is commonly overlooked. Since much philosophical focus tends to be on the authenticity of the agent's motivational attitudes, scant attention is paid to what happens once the agent's intention has been formed. As such, this dimension requires a little more in the way of motivation than some of the above dimensions. Self-governance must surely be taken to mean both governance *by* the self, and governance *of* the self. While concern with the former explains the focus on authenticity, concern with the latter should extend this focus to include the agent's navigation of a path through life. If the agent's authentic values and desires can't be implemented in action, then it is

hard to see how she is governing herself at all. To borrow the common analogy of captaining a ship, it would be as if the captain were busily plotting a course and issuing orders, and all the while the ship remained moored to the dock.

It is thus worth noting that if an agent's intention 'misfires', then it is odd to say that her autonomy is unscathed.³⁶ For instance, let's return to Sally and her cup of coffee. Sally intends to drink the (now correctly identified) coffee. However, when she tries to reach out her arm to grasp the cup, she finds that instead her arm flails wildly above her head. Even though the intention was formed in just the right way, in accordance with motivational attitudes that were correctly identified and endorsed in light of their history, Sally's autonomy cannot be considered unimpaired. This dimension thus tracks the extent to which the action that is performed corresponds to the intention that was formed. The highest level of this dimension will be achieved when an action successfully realises the precise intention; the lowest level reflects the absence of any action corresponding to the intention; and in between will be those actions that only partially or inaccurately realise the intention.

3b. Autonomy-Compromising Socialisation

With a sketch of our conceptual account in hand, the key question that remains is: what forms of socialisation compromise autonomy? In this section I will not attempt to

³⁶ Amongst those who would reject such reasoning is Michael McKenna, who explicitly states that "[autonomy's] scope comes to an end at the moment in which an agent performs simple mental actions, such as deciding or choosing". See his 'Autonomous and Morally Responsible Agency', in James Stacey Taylor (ed), *Personal Autonomy: New Essays on Personal Autonomy and its Role in Contemporary Moral Philosophy* (Cambridge: Cambridge University Press, 2005), p.229. Christman makes a similar claim in *The Politics of Persons*: "the competence conditions for autonomy merely refer to the effective ability to form intentions to act but not to complete such actions." (p.155)

definitively resolve this question of what the necessary preconditions for autonomy are. To do so would require sustained investigation of empirical questions of human psychology, which I don't propose to undertake here. Instead, I indicate the general shape that a theory of this kind would take, focusing exclusively on the first dimension of knowledge-of-the-world.

So what kinds of socialisation would compromise knowledge-of-the-world, as that dimension contributes to the phenomenon of autonomy? There are two directions from which this question can be approached. First, consideration must be given to forms of socialisation that directly instil false beliefs in a certain domain, and thus compromise autonomy in that domain. Second, consideration must also be given to forms of socialisation that undermine the background psychological conditions that facilitate the development and utilisation of knowledge-of-the-world.

In terms of the first dimension, it will be useful to start with a relatively non-controversial example. We could imagine, for instance, a cult that indoctrinated its members to believe that the leader was a God. Insofar as interactions with that person were pursued in light of believing her to be a God (i.e. obeying all of her commands because of her believed omniscience, donating all income to her in thanks for her divine interventions, etc.) my theory would deem the agent's autonomy to be compromised, and thus brand the socialisation as autonomy-undermining. The theory thus provides a principled way to differentiate autonomy-undermining from autonomy-supporting socialisation. Insofar as individuals are socialised to believe falsehoods about particular domains of action, their

autonomy will be undermined in those domains, and to that extent the socialisation can be classified as autonomy-undermining.

Of course, there are likely to be difficult cases. Does a refusal to teach children about sexual education later compromise their autonomy in sexual relations? Or does the teaching of intelligent design compromise the autonomy of individuals over a range of domains? These questions require investigation of the extent to which the (by hypothesis) false information interacts with the agents broader desires, values, commitments and plans, and as such the theory allows for there to be different answers for different agents. Insofar as misinformation prevents the agent from aligning her actions with her motivational attitudes, to that extent it compromises autonomy. If an individual has a commitment to, say, understanding the nature of the physical world around her, then teaching her only intelligent design may well undermine her autonomy. However, in the absence of such a commitment, such teachings may not affect the agent's autonomy at all.³⁷

The next question we must ask is whether socialisation could compromise autonomy, and in particular the dimension of knowledge-of-the-world, indirectly. An important attitude worth considering in this context is trust, both with respect to oneself and to appropriate others. To see why this is so, we must first briefly consider the nature of trust. For an agent to have an attitude of trust towards another individual involves two distinct assessments: first, that the trusted party is competent relevant to the entrusted activity, and second that

³⁷ More broadly, one may think that *all* agents have some level of commitment to truth seeking. If this is the case (and I make no judgment here one way or the other), then it would follow that *all* false belief would compromise autonomy, to some degree.

the trusted party regards the agent with benevolence.³⁸ For instance, I trust someone to catch me if I assess that they are sufficiently strong and co-ordinated to do so (competence), and in addition I assess that they will not deliberately drop me (benevolence). *Self-trust* requires that an agent take herself to be competent in the relevant domain, and in addition believe that she will not self-sabotage.

It should not be controversial to observe that certain forms of socialisation can effectively undermine trust, both towards oneself and towards others. All that is required is that individuals are taught to believe themselves or others to be either incompetent in the relevant domain, or malevolently motivated. Forms of socialisation that position certain social groups as irrational, or childlike, or malevolent, will decrease the likelihood of individuals trusting members of those groups. When the agent herself is a member of the targeted group, her trust in herself will likewise tend to be compromised.

This raises the question of whether lack of trust – either in oneself or in relevant others – could compromise knowledge-of-the-world. I believe it could. It is important to recall that knowledge in and of itself is insufficient – this dimension is relevant to autonomy insofar as the agent is able to put that knowledge to use in her reflection and resolution. Crucially, this requires self-trust. Benson's observations regarding the process of Gaslighting, though intended for a slightly different purpose, can be informative here.³⁹ If an individual believes herself to be delusional, failing to trust either her own perceptions or her rationality, then

³⁸ See Karen Jones, 'Trust as an Effective Attitude', *Ethics*, 107(1), 4-25; and Carolyn McLeod, *Self-Trust and Reproductive Autonomy*, (MIT Press, 2002).

³⁹ Paul Benson, 'Free Agency and Self-Worth', *The Journal of Philosophy*, 91/12 (1994). Gaslighting refers to an attempt to destroy another's perception of reality, and derives from an Ingrid Bergman film in which the lead character is made to doubt her own sanity, as her conniving husband manipulates her environment and discredits her perceptions. For further discussion of the relationship between self-trust and autonomy, see Trudy Govier, 'Self Trust, Autonomy, and Self Esteem', *Hypatia*, 8/1 (1993).

she is unable to use the information she is receiving to reflect on her prospective action. In the absence of self-trust, the necessary connection between comprehension and reflection is disrupted. As such, if a form of socialisation teaches individuals that they lack rationality in a certain domain, or are inclined towards false perceptions, then the resulting lack of epistemic trust will undermine the role that knowledge-of-the-world plays in autonomy.⁴⁰

Less commonly discussed, at least in the context of autonomy, is the importance of trust in relevant others. Knowledge-of-the-world depends upon access to reliable information. While some of that information may be obtainable through first-person observation, it is widely recognised that knowledge largely depends upon testimony.⁴¹ At least in some domains, autonomous action will depend upon trusting those who are able to provide us with the information needed for adequate reflection and resolution. An (admittedly not unproblematic) example can be seen in the domain of medical decisions. According to my theory, an decision regarding a particular medical treatment would be more autonomous the greater the understanding of what that treatment involved. This is information that can only plausibly be obtained by trusting those with greater knowledge of the treatment in question. If an individual were socialised to distrust all medical professionals, say, it would be very difficult for her to autonomously choose how to respond to an illness.⁴² As such,

⁴⁰ On this model, then, some forms of oppressive socialisation will count as autonomy-undermining. However, the relationship is contingent. Insofar as oppressive socialisation compromises one or more of the five dimensions, then it is autonomy-undermining. It is at least conceptually possible, though, that an agent undergo oppressive socialisation and yet her autonomy remain intact. C.f. Sonya Charles, 'How Should Feminist Autonomy Theorists Respond to the Problem of Internalized Oppression?', *Social Theory and Practice*, 36/3 (2010).

⁴¹ See, i.e., Tony Coody, *Testimony: A Philosophical Study*, (New York: Oxford University Press, 1995), and Lorraine Code, *What Can She Know?*, (Cornell University Press, 1991).

⁴² This shouldn't be taken to mean that autonomy requires deference to all professionals. This is particularly so when professionals lack the necessary knowledge to inform us of consequences. For instance, on my theory women's autonomy would have been diminished rather than augmented when they historically followed doctor's orders for bed-rest in response to a diagnosis of hysteria.

some forms of socialisation can diminish autonomy by inculcating attitudes of distrust towards those whose knowledge is necessary for informed critical reflection.

4. Putting the Theory to the Test

I have claimed that the five-dimensional model outlined above would be more successful in analysing the autonomy of Julie and Meredith than the four theories criticised above. It is time to put that claim to the test. What, then, would the theory say about Julie and Meredith? The first thing to note is that it would resist attempts to classify either woman as either autonomous or non-autonomous, and would instead lend itself to assessing the degree to which each is autonomous, in the given scenario. This would be determined in accordance with the five dimensions: knowledge-of-the-world; self-knowledge; self-definition; internal self-realisation; and external self-realisation. Since each scenario will present a different pattern across these dimensions, it allows for a nuanced analysis of precisely *how* each agent's autonomy may be compromised in the given situation.

As noted at the outset, one of the desiderata for a theory of autonomy is that it ask the right questions when confronted with an agent like Meredith: what more would need to be known in order to make an assessment of her autonomy? The five-dimensional model gives a clear answer. We would need to know whether there were any factual errors in her understanding of the operation she was contemplating; we would need to know whether she understood her own desires, values, commitments and plans; we would need to know whether those desires, values, commitments and plans were ones she would endorse in full knowledge of their history and of relevant external facts; we would need to know whether

her intention to get the breast implants accurately responded to and reflected those desires, values, commitments and plans; and we would need to know whether the operation she underwent accurately reflected the intention that she had formed. Admittedly, in most cases gaining access to this information will be epistemically challenging, to say the least. Nonetheless, it is no slight on a theory of autonomy that it does not give easy answers to the question of how we make determinations of autonomy.

How does the theory I have presented here compare to the theories examined in Section Two? Unlike Christman's theory, my theory would not pin the determination of Meredith's autonomy merely on her lack of alienation from desires induced by patriarchal socialisation. It requires further that she would endorse her desire, in the additional light of full knowledge of relevant facts about the world. In this case, this would include knowledge of the history of patriarchy, and the effect it tends to have on women raised in it. It is only if the endorsement would persist in light of *this* knowledge that it would signal high levels of autonomy. Moreover, it would look beyond endorsement to ensure that Meredith understood the implications of her surgery, and that the act of undertaking surgery was in line with the intention that followed from her reflection. It thus places more stringent requirements on an agent like Meredith, without going so far as to say that her autonomy was inevitably compromised because of the content of her action.

Next I examined Mele's historical theory of autonomy. The problem here was his definition of compulsion*, which resulted in individuals being deemed non-autonomous merely for being inculcated as children with the beliefs necessary for critical reflection. I argued that the inculcation of a belief in childhood, which then becomes unsheddable, cannot be

enough on its own to generate a diagnosis of non-autonomy. My theory again avoids this implication, because it relies on the informed endorsement of the agent, rather than demanding that a pro-attitude have an origin that is somehow external to socialisation. We can thus assess Julie's autonomy according to how well she meets the five dimensions outlined above, without worrying about the origin of her reflective capacities.

The third theory I considered was Stoljar's substantive account of autonomy. My theory follows Stoljar in holding that (at least some) false beliefs compromise autonomy. Yet unlike in Stoljar's theory, my inclusion of this condition is embedded within a conceptual account of autonomy that justifies its existence. Taking the condition in this context also sets its boundaries. Not all false beliefs will undermine autonomy; only those that are relevant to the agent's attempt to match her action to her endorsed desires, values, plans and commitments. When we compare how my theory would approach Meredith's situation, we can see this difference clearly. For Stoljar, what matters is whether the reasons that motivate Meredith to get the breast-implants are false and oppressive. On my account, what matters is whether Meredith those false beliefs compromise her attempt to align her actions with her endorsed motivational attitudes.

This raises the important the important question of whether the dimension of knowledge-of-the-world that I have put forward includes normative facts. For instance, if Meredith falsely believes that breast implants enhance a woman's self-worth, or that they bear no relation to patriarchy, does this mean her autonomy is compromised?⁴³ At the risk of sounding the same old refrain, the answer is going to be – yet again – that it depends. If Meredith has a background commitment to opposing patriarchal norms, or even a more

⁴³ Thanks to an anonymous reviewer of this journal for pushing me on this point.

abstract commitment to acting in accordance with just principles, then being mistaken about the patriarchal nature of breast implants will compromise her autonomy. In such a case, her actions would come apart one of her deep commitments. We could imagine a Meredith, though, who was not committed to any of these things. In such a case, the mistaken normative belief would not affect her ability to act in accordance with her motivational attitudes, and thus would leave her autonomy unscathed.

Finally, I examined Westlund and Benson's accounts, which, as we saw, required the addition of externalist conditions to make them plausible. The account I have presented can nonetheless explain the *prima facie* appeal of these views. If an individual cannot find reasons for what she has done, or does not feel that she is responsible for her actions, then this is a strong indication that she has not undergone the three-step process of comprehension, reflection and resolution. Nonetheless, this test is not infeasible. Such dispositions may well be correlated to autonomy, but they do not constitute it. Since the dispositions and the constitutive features of autonomy can come apart – as they do when a hypnotised individual claims authority for an action she has not freely committed – then these dispositions can at best function as a quick and semi-reliable guide to determining an agent's autonomy. When we look to Julie and Meredith, we see again that my theory asks more appropriate questions of their situations than does Benson's. Rather than asking whether they are willing to stand behind their decisions, my theory asks to what degree the five dimensions of autonomy were realised in those decisions.

The account that I have presented defines autonomy in terms of five distinct dimensions: knowledge-of-the-world, self-knowledge, self-definition, internal self-realisation and

external self-realisation. This framework provides a principled way for distinguishing autonomy-supporting from autonomy-compromising autonomy: insofar as socialisation reduces the level of one or more dimension of autonomy, over one or more domains of human life, then it is autonomy-compromising. Such an account, I suggest, holds out the best hope of satisfactorily resolving the problem of socialisation.