

Penultimate version. September 2011.

Final version forthcoming In K. W. M. Fulford, Martin Davis, George Graham, John Sadler, Giovanni Stanghellini & Tim Thornton (eds.), *Oxford Handbook of Philosophy and Psychiatry* (Oxford: OUP)

AUTONOMY AND DEPRESSION

Lubomira Radoilska

I. Introduction

Autonomy is the focus of at least four major philosophical inquiries. One of them aims to establish the defining features of autonomous motivational states, such as identification, endorsement, or acceptance (Frankfurt 1998). It relates autonomy to the concepts of agency and freedom of will. Another approach explores the question whether autonomous choices ought to accord with particular values, such as self-respect (Hill 1991). It identifies covert forms of oppression and elaborates on corrective initiatives. A third inquiry concentrates on the links between responsiveness to reasons and effective control over one's life (Korsgaard 1996). It looks at rational agency and its interrelation with moral responsibility. Finally, autonomy is a central topic in bioethics. It is critically examined in connection to informed consent and decisional capacity (O'Neill 2002).

Arguably, the ensuing proliferation of autonomy conceptions points to the significance of the underlying concept. However, it may also prompt scepticism about its unity and theoretical appeal. From this latter perspective, the lines of inquiries mentioned above are deemed not to explore different conceptions of the same concept, but to sketch separate concepts with no necessary interconnections. If this view is correct, we should either replace 'autonomy' with narrower, qualified autonomy-concepts, such as autonomy as independence of mind, autonomy as responsiveness to reasons, etc. etc., or abandon 'autonomy' altogether, for it has become an overworked and misleading term (Arpaly 2003: 118).

In this paper, I shall not address this kind of general scepticism about autonomy, but aim to respond to a challenge faced by a specific conception of autonomy as an independent, though

limited in scope source of justification. In doing so, I shall make two related assumptions: firstly, that autonomy is primarily an agency concept and, secondly, that the conception I defend is central to it.¹ The challenge I shall address here questions the feasibility of a reliable distinction between autonomous and non-autonomous choices. This challenge is significant because, in the absence of such a distinction, a conception of autonomy as independent source of justification will run into various paradoxes, such as upholding non-autonomous choices in the name of autonomy. This line of inquiry is equally relevant to outlining a plausible conception of decisional capacity, for as we shall see later decisional capacity would not be able to fulfil its designated purpose – to protect certain self-regarding initiatives from interference – unless it points to an independent source of justification applicable to such initiatives. Depression offers a promising focal point for the discussion because it credibly substantiates the challenge faced by both decisional capacity and the related autonomy conception.

The argument proceeds as follows. In the next section, I outline two paradoxes about depression and dismiss an initially plausible solution to each of them. In the middle section, I sketch a logical reconstruction of the challenge stating that, in light of the earlier paradoxes about depression, decisional capacity and the related autonomy conception cannot be consistently defined in either value-neutral or value-laden terms. In the penultimate section, I propose a revised value-neutral solution, according to which depression hinders decisional capacity viz. personal autonomy in so far as it leads to paradoxical identification and thwarts the relationship to one's motives that typically conveys autonomy to ensuing initiatives. In the final section, I anticipate a possible objection suggesting that the account I propose has unacceptable consequences, such as being at odds with respect for depression sufferers as persons.

¹ I have argued for both claims in Radoilska (2012). In distinguishing a concept from its conceptions, I draw on Gallie (1956).

II. Two paradoxes about depression

Depression poses two important challenges to philosophical reflection about intentional agency. The first relates to a classical conception, according to which we always act under the guise of the good. The underlying claim is that, whenever we engage freely in some kind of pursuit, we do so because we conceive it as valuable in some respect. Immediate examples include finding an activity enjoyable or appreciating its effects. Although this conception has recently come under attack, it would be a mistake to underestimate its intuitive appeal.² For it offers a plausible way of explaining core, everyday cases of voluntary actions as opposed to coerced ones. Following this line of thought, voluntary actions could be fully accounted for by an agent acknowledging: ‘I did φ because I wanted to φ ’

in so far as this means:

‘I did φ because I like/ care about φ -ing’, or:

‘I did φ because, by φ -ing, I get [closer to] x, y, z that I like/care about’.

In contrast, coerced actions are not accurately explained by pointing to the fact that the agent consented to perform them. Even a first-person account, such as ‘I did φ because I wanted to φ ’ remains insufficient. In instances of coercion, this statement stands for:

‘I did φ because I was made to [want to] φ ’, or:

Unless I φ -ed, x, y, z that I like/care about, would have been lost or damaged. So, I did φ ’.

The distinction between these two categories of actions is central to our thinking about intentional agency. In particular, this contrast helps to pin down the idea of an agent as the ultimate source of actions, which are free, intentional, and uncompelled.³ A straightforward link between motivation

² Critical approaches include: Stocker (1979), Velleman (1992) and, more recently, Setiya (2010).

³ The proposed sketch of intentional agency draws on Aristotle’s account of voluntary actions as opposed to the so-called mixed actions from *Nicomachean Ethics* 3.1–5; see also Radoilska (2007), 191–231. On the notion of a free, intentional, and uncompelled action, see Mele (1987).

and evaluation is the distinctive feature of such paradigm cases. Conversely, in coerced actions, this link becomes complicated. As a result, the agent cannot be seen as the ultimate source of such actions, for the relationship between motivation and evaluation that they substantiate is conditioned from outside. Hence, the intuition that we always act under the guise of the good enables us to achieve two related goals. The first is to appreciate the distinction between actions proper and actions performed under various degrees of undue influence or coercion. The second is to, nevertheless, conceive the latter category as something partly done *by* the coerced agent rather than merely done *to* her. Thus, intentional actions imply a robust, though not always direct link between motivation and evaluation.

This conception of intentional agency seems at odds with depression, which typically involves mental states where the underlying connection between motivation and evaluation is apparently severed. As Michael Stocker points out:

“Through spiritual or physical tiredness, through accidie, through weakness of body, through illness, through general apathy, through despair, through inability to concentrate, through a feeling of uselessness or futility, and so on, one may feel less and less motivated to seek what is good. One’s lessened desire need not signal, much less be the product of, the fact that, or one’s belief that, there is less good to be obtained or produced, as in the case of a universal Weltschmerz. Indeed, a frequent added defect of being in such ‘depressions’ is that one sees all the good to be won or saved and one lacks the will, interest, desire, or strength.” (1979: 744).⁴

At first sight, the underlying difficulty seems easy to resolve. Arguably, the preceding outline of intentional agency only requires that the motives for an action can be traced back to the agent’s appraisal of this action as good under the description, under which it was undertaken (Anscombe

⁴ Arguably, this outline matches the clinical description of depression proposed by ICD-10 and DSM-IV, listing, on the one hand, loss of interest and enjoyment and, on the other, a bleak and pessimistic view of the future as central diagnostic criteria.

1963). Thus, considering a course of action as worthwhile does not have to be immediately motivating. However, the challenge from depression reappears as soon as we take into consideration that, in depression, the course of action deemed to be worthwhile is typically not abandoned because of some interference from outside, making it, for instance, less likely to succeed. Instead, the root of the problem seems to be that commitments, the agent still considers as worthwhile, no longer motivate her to act one way or the other. The distinctive paradox brought by depression here is that the desirable is not desired and the choiceworthy does not get chosen.

The second challenge posed by depression calls into question our ability to reliably distinguish autonomous from non-autonomous choices. This affects primarily a key function of attributing autonomy to substantively self-regarding initiatives, which is to protect them from interference.⁵ Here, the category of substantively self-regarding is best understood negatively, in the sense that related initiatives require no further justification on moral or prudential grounds. Instead, their permissibility depends entirely on the issue whether a person autonomously commits to them or not. Since autonomy, in this context, stands for an independent source of justification, it seems natural to opt for purely procedural, value-neutral criteria when ascertaining whether an initiative should be protected from interference because it is autonomously undertaken rather than because it is morally commendable or prudentially required. The so-called hierarchical or higher-order

⁵ In this paper, I shall employ ‘initiatives’ as a general term, covering choices, decisions, and actions. The notion of substantively self-regarding as opposed to self-regarding initiatives that are equally other-regarding will be further clarified in the following section. See also Feinberg (1986, Ch. 17) and Scoccia (2008).

accounts of personal autonomy provide a plausible way of identifying autonomous versus non-autonomous choices, which does not implicitly rely on further moral or prudential reasons.⁶

This function of autonomy is particularly relevant to medical contexts. For treatment refusal looks like a central, if not *the* central case of a substantively self-regarding initiative: compared to a person's interest in her life and welfare, the interest of others in the matter can only be secondary.⁷ Moreover, the idea of autonomy as a justificatory alternative to moral and prudential reasons, applicable to this particular cluster of choices, often underpins current legal thinking about decisional capacity. The following excerpt of a recent U.K. high court ruling is an example:

“A mentally competent patient has an absolute right to refuse to consent to medical treatment for any reason, rational or irrational, or for no reason at all, even where that decision may lead to his or her own death.” (*Re MB* 1997)

Depression creates a major difficulty for the underlying approach to decisional capacity and personal autonomy. For recognised symptoms, such as low self-esteem and suicidality arguably motivate treatment refusal by people who suffer from depression, especially in cases where this is very likely to lead to their own death.⁸ The plausible causal link from symptoms to motives for

⁶ See Frankfurt (1998; 1999) and Dworkin (1988). For an informative survey of both developments and critiques of hierarchical conceptions of personal autonomy, see Christman (1988) and Taylor (2005).

⁷ See however the notion of ‘garrison threshold’ in Feinberg (1986: 21–23) specifying the conditions under which this is not the case.

⁸ It is helpful to distinguish this kind of treatment refusal from others, where a causal path between depressive symptoms and effective motivation is rather doubtful. Examples are cases where a person objects to a particular treatment, e.g. a course of antidepressants rather than to treatment per se. Moreover, not all treatment refusals per se could be plausibly linked back to depression. For instance, a person may be unwilling to undergo any kind of medical treatment for

action casts the related decisions as clear-cut cases of non-autonomous initiatives, on a par with coerced actions, since undue influence onto the agent is present in both kinds of instances. Yet, if we follow the criteria proposed by the hierarchical accounts of autonomy, we seem compelled to acknowledge such depression-induced treatment refusals as autonomous. The paradox comes from the fact that a depressed person could not only identify with this kind of decisions, but also endorse them on reflection.⁹ For instance, she may still consider that her life is no longer worth living, whilst being aware that this kind of attitude toward oneself is often interpreted as a symptom of a particular mental disorder. In so doing, she would satisfy both weaker and stronger versions of a hierarchical conception of autonomy. This outcome seems counterintuitive, for, in the context of depression, it requires that we uphold some apparently non-autonomous choices in the name of autonomy.

It is tempting to try and resolve this second paradox about depression by abandoning the underlying, value-neutral approach to decisional capacity viz. personal autonomy in favour of a substantive or value-laden strategy. The latter strategy suggests that, in order to be autonomous, individual choices ought to accord with particular values, in addition to fulfilling purely

depression because she takes depressive episodes to offer her an opportunity to further her self-knowledge and resolve the underlying conflicts by introspection.

⁹ In Frankfurt (2002) the notions of identification and endorsement are both clarified in terms of acceptance that involves neither evaluative judgment, nor a cognitive process, such as deliberation. In contrast, Dworkin (1988) insists on the role of second-order reflection in order to ensure that the first-order motivations under scrutiny are the agent's own in the required sense, that is, independent from undue influence.

procedural constraints.¹⁰ This additional requirement would enable us to accommodate the compelling intuition that symptom-related initiatives cannot be considered as autonomous. In particular, we would be in a position to override depression-induced treatment refusals as non-autonomous on grounds that they are at odds with the crucial value of self-respect.¹¹ This tempting solution, however, comes at a price, for substantive or value-laden accounts of autonomy seem to undermine its possible function as a sufficient rationale for substantively self-regarding initiatives. This is because they tie up the issue of personal autonomy to that of whether a person effectively commits to worthwhile or rational projects.¹² As a result, autonomy can hardly provide a justificatory alternative. Instead, it appears as a mere appendage to the moral or prudential reasons in favour of a particular self-regarding initiative.

III. The challenge to decisional capacity: a logical reconstruction

A promising way forward is to bring together the two paradoxes about depression so that we are able to sketch out an account of depression clarifying its impact on both autonomous and broader intentional agency. As a first step in this direction, let us consider the following logical reconstruction of the challenge to convincingly define decisional capacity in either value-neutral or value-laden terms.

¹⁰ See Culver and Gert (2004) and Charland (2008) arguing in favour of such a move. Both papers take instances of depression-induced treatment refusals in the sense specified above as decisive counterexamples to a purely procedural understanding of decisional capacity.

¹¹ See Hill (1991: 4–18). In the following, I shall use the terms self-respect and self-esteem interchangeably. This is consistent with the contrast between self-respect, on the one hand, and both deference and servility, on the other, which is central to Hill's thesis.

¹² Culver and Gert (2004) provides an example of the latter strategy.

1. An important function of attributing autonomy to substantively self-regarding initiatives is to protect them from interference. The primary aim is to shield individuals from unwarranted uses of authority, such as the tyranny of majority. This function is associated with and often assimilated to the principle of respect for personal autonomy in bioethics.

1.1. The distinctive feature of substantively self-regarding initiatives is that, *ex hypothesi*, they could be solely justified by virtue of being autonomously chosen. If this condition is satisfied, such initiatives should be recognised as permissible in their own right, independently of further values or principles.¹³

1.2. Notwithstanding, the scope of substantively self-regarding initiatives is, to some degree, context-dependent. The relatively recent decriminalisation of suicide attempts is an example. Whilst earlier this kind of self-regarding initiative was not deemed eligible for protection from interference, it is now accepted as substantively self-regarding, independently of its foreseeably detrimental impact on others.

2. Tests of decisional capacity are meant to fulfil a function, relevantly similar to that described in the premise above.

2.1. The choices to which these tests apply, e.g. treatment refusals are presumed substantively self-regarding. What is at stake is whether a person is capable to make such a choice in a particular context so that this choice is worthy of protection from interference independently of further values or principles.

3. A reliable way of distinguishing both autonomous from non-autonomous and capacious from incapacious decisions is crucial to fulfilling the central function as outlined in premise 1.

¹³ This intuition is reflected in the distinction between separate principles of bioethics, where respect for autonomy is contrasted with considerations based on nonmaleficence, beneficence, and justice, see Beauchamp and Childress (1979).

3.1. Although treating relevant autonomous choices as non-autonomous often attracts criticisms for being paternalistic, the opposite mistake is just as pernicious because it condones various forms of unfreedom, including coercion, manipulation, and compulsion.

4. The two divides above should overlap so that decisions, acknowledged as autonomous in the sense specified in subsection 1.1., are also deemed capacious as stipulated in 2.1., and vice versa.

4.1. If the preceding condition is not fulfilled, some substantively self-regarding initiatives will have to be both protected and overruled.

4.2. Hierarchical accounts of autonomy offer a prima facie plausible way to draw the distinction at issue.

4.2.1. For they take as relevant only the question of whether the agent endorses the choice under scrutiny at the moment of choice; the suggested procedural constraints are minimal and aim to eliminate only patent mistakes about facts and practical contradictions (Frankfurt 1998: 159–176; 1999: 129–141).

4.2.2. This seems to be an intuitive means to avoid the imposition of evaluative commitments from outside, which is the central aim of both autonomy and capacity attribution in the context of substantively self-regarding initiatives.

4.3. The conditions set out in the Mental Capacity Act (2005) are sometimes interpreted as aligned to a hierarchical understanding of autonomy. These conditions include: the ability to understand the information relevant to the decision; to retain that information; to use or weigh that information as part of the process of making the decision and; to communicate a final decision.

4.3.1. The Woollorton case is an example of such an interpretation, especially as presented in McLean (2009), on which the following description is based.

Kerrie Woollorton, a 26-year old woman has ingested antifreeze on nine previous occasions but had accepted lifesaving treatment afterwards. She was deemed to have an untreatable emotionally unstable personality disorder and, possibly, to be depressed. In 2007, days before her death, she

had drafted an advance statement indicating that she did not wish to be treated should the same circumstances arise in the future, even if she called for an ambulance. Rather than being treated, she wanted to die in a situation where she was not alone and where comfort care was provided. A document containing a rejection of treatment was presented by her on admission to hospital, after ingesting antifreeze for a tenth time. This document was accepted as valid. In addition, she made a contemporaneous refusal of treatment and was considered to satisfy the criteria for decisional capacity. The medical professionals involved did not give lifesaving treatment. A subsequent coroner's ruling upheld their decision as lawful.

5. The claim that this decision was incorrect could be based on the following grounds:

5.1. Treatment refusal after a suicide attempt is ineligible for protection under the principle of respect for autonomy because it is not self-regarding in the required sense (see subsection 1.2. above);

5.2. Although both substantively self-regarding and capacious, Wooltorton's treatment refusal should have been overridden on grounds of beneficence.

5.3. Wooltorton's treatment refusal was not capacious. Hence, either the underlying procedural interpretation was not adequately applied or this interpretation is itself inadequate.

6. Leaving the first two hypotheses aside, both options identified in subsection 5.3. above could be supported by the idea that depression involves a significant impairment and even a breakdown of intentional agency. According to ICD-10, a considerable difficulty in continuing with ordinary activities is typical of moderate depression, whereas severe depression leads to inability to continue with any, except very limited activities. Related symptoms include:

6.1. Indecisiveness and diminished ability to concentrate and think; and

6.2. General loss of interest and pleasure, sense of worthlessness, hopelessness and suicidality.

7. The symptoms listed in 6.1. could be easily accounted for by a hierarchical interpretation of capacity (see 4.2 – 4.3 above); however, the symptoms listed in 6.2. remain unaddressed.

8. Yet, with respect to establishing whether a possible breakdown or severe impairment of intentional agency has taken place, the distinction between the two groups of symptoms is arbitrary. Consider the following excerpts from a memoir of depression:

“I can remember lying frozen in bed, crying because I was too frightened to take a shower and at the same time knowing that showers are not scary. I ran through the individual steps in my mind: You sit up, turn and put your feet on the floor, stand, walk to the bathroom, open the bathroom door, go to the edge of the tub...I divided it into fourteen steps as onerous as the Stations of the Cross. I knew that for years I had taken a shower every day. Hoping that someone else could open the bathroom door, I would, with all the force of my body, sit up; turn and put my feet on the floor; and then feel so incapacitated and frightened that I would roll over and lie face down. I would cry again, weeping because the fact that I could not do it seemed so idiotic to me.” (Solomon 1998: 48–49).

“During the worst of my depression, when I could hardly eat, I could not have done myself real harm. In this emerging period, I was feeling well enough for suicide. I could push myself to do pretty much all of what I had always been able to do, but I was unable to experience pleasure. Now I had the energy to wonder *why* I was pushing myself and could find no good reasons.” (*ibid*: 52).

8.1. The first quote offers a plausible description of a breakdown of intentional agency. Although it involves both groups of symptoms as listed in 6.1 and 6.2., the defining feature amounts to a double rift between evaluation and motivation: on the one hand, the agent is not motivated by his evaluative judgments, e.g. ‘showers are not scary’; on the other, he finds himself acting upon motives that he cannot comprehend, let alone value, e.g. ‘it seemed so idiotic to me’.

8.2. The second quote offers a plausible description of a severe impairment of intentional agency. Similarly to the previous one, it involves both groups of symptoms as listed in 6.1. and 6.2., and the underlying mental state points to a mismatch between evaluation and motivation. The agent

does not find his evaluative judgments worthy of acting upon and, with respect to motivation, aggression toward the self takes over.¹⁴

9. In light of premises 1 and 4 above, both decisional capacity and the related autonomy conception are incompatible with either a breakdown or a severe impairment of intentional agency as described in subdivisions 8.1. and 8.2. The reason is that respect for choices, made under either of these conditions, would promote a particular kind of unfreedom, whereby the agent lacks either effective control or authority over her motives. This kind of unfreedom is compatible with *autocracy*, whereby I mean a state, in which the agent is unconstrained by undue influence from outside; yet, although left on her own, she does not manage to attain successful self-government.¹⁵

9.1. Autocracy is one way in which personal autonomy may become unstuck.

10. Together with premise 3 and in particular 3.1., the preceding conclusion supports the idea that with respect to depression, the underlying procedural approach may lead to paradoxical decisions to uphold non-autonomous but autocratic choices in the name of autonomy, as in the Wooltorton case.

11. In order to avoid this unwelcome upshot, it may seem natural to opt for a value-laden account of both decisional capacity and the related autonomy conception. However, in light of subsections

¹⁴ This may not be fully explicit in the excerpt I refer to, however, the proposed account is clearly confirmed by the follow-up episodes in the memoir, where the agent stops short of committing suicide, then engages in various kinds of reckless behaviour, which he abandons only because he realises that this could harm not only him, but also third parties. Similarly, Solomon's change of heart about suicide is essentially motivated by other-regarding concerns: 'it would be sad for my father to have worked so hard at saving me and not to have succeeded' (*ibid*: 52).

¹⁵ On the idea of autonomy as an actual state of self-government, see Feinberg 1986, Ch. 18, esp. 31–44.

1.1 and 4.2.2, this move is apparently blocked. For this kind of solution undermines the crucial intuition, according to which personal autonomy is an independent source of justification for substantively self-regarding initiatives. This becomes clear if we consider in some detail a promising value-laden alternative.

11.1. For instance, if we take it that substantively self-regarding initiatives, set out in a state of mind relevantly similar to 8.1 and 8.2, are non-autonomous by virtue of being patently irrational, this would mean that, whenever we respect this kind of initiatives as autonomously chosen, we effectively respect their putative rationality. In so doing, we implicitly endorse *orthonomy* (Pettit and Smith 1996) as an ideal of intentional agency. In contrast to personal autonomy as inner consistency or self-integration that we achieve by acting only upon motives that we are able to identify with, orthonomy essentially requires that we adjust our motives to independent norms. The core intuition is that we are unfree whenever we choose to engage in pursuits that we would disavow as unworthy, if we were to adopt an ideal observer's perspective. This kind of unfreedom is relevantly similar to constraints that false beliefs impose on our thinking. Arguably, our freedom of thought is undermined, if we are allowed to believe anything we like, regardless of logic and evidence. On this ground, it is plausible to infer that our freedom of action is equally undermined, if we are allowed to pursue any project we like, independently of its worth. Following this line of reasoning, the issue whether a substantively self-regarding initiative is worthy of protection from interference boils down to whether the initiative accords with some independent norm, such as rationality. Hence, the focus of protection clearly shifts from *self-determination* viz. *autonomy* to *correct* determination viz. *orthonomy* of the will.

11.2. Orthonomy is compatible with a loss of personal autonomy, as in cases where the correct determination of the will is extraneous to the will so determined.¹⁶

¹⁶ Berlin's critique of positive freedom (2002: 178–200) could be seen as an expression of this concern.

11.3. The trouble does not go away, if we conceive correct determination as a prerequisite for genuine self-determination since this conceptualisation implies that self-determination should always be kept in check by further principles or values. It cannot provide an independent rationale for any initiatives, substantively self-regarding or otherwise.

11.4. The difficulty at issue – finding a reliable way to distinguish autonomous from non-autonomous choices vis-à-vis substantively self-regarding initiatives – is effectively dismissed rather than offered a solution.

12. Hence, decisional capacity and the related autonomy conception cannot be consistently defined in either value-neutral or value-laden terms.

IV. A possible solution: loss of self-esteem as paradoxical identification

The outcome of the preceding reconstruction is not entirely aporetic, for it helps outline the contours of a satisfactory solution. In essence, this should be a formal rather than a substantive approach to both decisional capacity and the related autonomy conception, the exclusive focus of which is a person's relationship to her motives, as in higher-order accounts. This conclusion follows directly from the earlier observation that the problem posed by depression comes down to a kind of deficit of authority or control over one's motives rather than a conflict with particular evaluative judgements. By focusing on the person's relationship to her motives, we are able to integrate the compelling intuition, according to which there is a close link between autonomy and self-respect. However, unlike substantive conceptions, which take the accord with the value of self-respect as an extra test applicable to the content of initiatives that have already satisfied the procedural conditions for being autonomously chosen, self-respect would be inbuilt in the formal constraints on the kind of relationship to one's motives that is compatible with personal autonomy.

This difference in strategy is significant. If this revised formal account is successful, it will offer a way of distinguishing substantively self-regarding initiatives that should be protected from

interference from those that should not, which is both principled and reliable. By being principled, I mean that the criteria will be solely autonomy-based instead of falling back onto further considerations on an ad hoc basis, whenever an apparently autonomous self-regarding initiative seems too striking or counterintuitive to be granted protection. This is a clear advantage vis-à-vis value-laden accounts. In contrast, the improvement with respect to standard higher-order accounts lies in the ability to reliably distinguish autonomous from non-autonomous initiatives and avoid the kind of paradoxical decision to uphold non-autonomous but autocratic choices in the name of autonomy.

Key to the proposed solution is the claim that identification, the kind of relationship to one's motives that typically conveys autonomy to ensuing initiatives, becomes paradoxical in depression. Crucially, depressive identification is self-alienating: it leads to ambivalence, inner conflict and, ultimately, loss of self-esteem, which I take to involve both enmity toward the self and a sense of worthlessness based on a perception of the self as an inadequate source of actions.

This central claim draws on Freud's account of melancholia (1915) and, in particular, the idea that melancholia stems from a person's identification with someone or something that she should not or can no longer care about. According to Freud, the underlying process becomes intelligible by comparison with mourning, for in both cases the objective is coming to terms with a significant loss. However, the loss associated with melancholia is rarely as clear-cut as that conducive to mourning. One reason is that what is experienced as lost in melancholia may not only be a beloved person, but also an abstract idea, such as friendship, inspiration, one's very *raison d'être*. Another reason is that, unlike most cases of bereavement, the affliction suffered by the melancholic person can effectively be attributed to the object of her loss.

This becomes clear if we look at cases, which Freud considers as central. In such cases, a person gets hurt, disappointed, or abandoned by an intimate relation. Ambivalence, Freud considers, is key to the experience conducive to melancholia. For one and the same person becomes an

appropriate target for contradictory attitudes. She is still a friend or a lover that one trusts and is fond of; yet, she also appears as someone that one should be wary of and resent.

A further layer of ambivalence adds up with the fact that the person wronged does not face this interpersonal problem so that it can be resolved by means of either reconciliation or revision of the prior relationship as valueless, deceitful or, perhaps, just over. Instead, the prospective melancholic redirects the resulting frustration toward the self. A possible explanation is that she avoids articulating the problem in such terms out of fear that neither of these options is really available to her. Thus, she takes herself to be unable to either mend the relationship at stake or let it go for good. This contributes to the paradoxical identification with the object of one's frustrations and disappointments, which is at the heart of the phenomenon. The person wronged is now able to shift the blame for what was inflicted upon her onto herself and, consequently, to perceive herself as a wrongdoer rather than a victim. In so doing, she gets the chance to maintain her former attachment and resist the need for change. This, however, comes at the price of internalising the interpersonal conflict at the root of the problem. To give an example, a woman who discovers that her husband has been cheating on her may be reluctant to confront him because she fears that this could precipitate the end of their marriage. Yet, having chosen to pretend that she is unaware of his infidelity, the cheated wife needs to work out a plausible story about why she would want to continue living with someone who obviously disrespects her. In this story, she is likely to portray herself as the ultimate culprit who has drawn away the kind and loving husband of hers by becoming, say, increasingly plain and boring. The identification with the unpleasant character that she has invented for herself enables the unfortunate wife to maintain her esteem and affection for the husband, whom, according to her story, she does not deserve.

We can now see why the process of paradoxical identification eventually ends up with a loss of self-esteem, understood as a phenomenon whose complementary sides are sense of worthlessness and hostility toward the self. Both follow directly from the fact that the melancholic takes inappropriate responsibility for her misfortune. To return to the preceding example, the cheated

wife reinvents herself as undeserving so that she may exculpate her husband. In her eyes, he now appears right to disrespect her, for, allegedly, she is unworthy of respect. Moreover, since the cheated wife casts herself as the one to blame for the failing marriage, her sense of worthlessness ties directly in with hostility toward herself. Having paradoxically identified with the husband who has wronged her, she becomes the target of her original desire to make him pay for this.¹⁷

Building on this model, it is possible to integrate various kinds of abstract ideas as eligible objects, whose perceived loss may be similarly compensated for by paradoxical identification. For instance, an ambitious professional may be unwilling to acknowledge his disappointment with a career move that he carefully planned and worked hard for. Like the cheated wife from the example above, he may wish to keep up the pretence that his current position is truly rewarding. In order to do so, he would have to assume that only his own inadequacy prevents him from flourishing at the new workplace. The ensuing paradoxical identification with an organisation that he finds frustrating would enable him to resist bringing up the discrepancies between the conditions offered and the opportunities in place. In turn, this could save the ambitious professional both a potentially costly interpersonal conflict and a radical revision of his career plan. However, as in the unfortunate marriage scenario, the avoidance of a would-be professional disaster comes at the price of greatly diminished self-esteem.

To recap the argument of the current section, the notion of paradoxical identification helps explain the lack of authority over substantively self-regarding initiatives that, as shown in the

¹⁷ Freud (1915: 251): "... the patients usually still succeed, by the circuitous path of self-punishment, in taking revenge on the original object and in tormenting their loved one through their illness, having resorted to it in order to avoid the need to express hostility to him openly. After all, the person who has occasioned the patient's emotional disorder, and on whom his illness is centred, is usually to be found in his immediate environment."

previous section, is typical of depression.¹⁸ These initiatives cannot be considered as autonomous in so far as they substantiate a conflicting relationship to one's motives. This is because the paradoxical identification with something or someone that the agent implicitly loathes, naturally leads to internalised ambivalence. In essence, the agent remains irresolute and fails to take control over her motives. Instead of making up her mind, she absorbs a practical contradiction that she faces. Initiatives that could be traced back to this process are non-autonomous, for they are not actively determined by the agent: she merely goes along with them. Moreover, she cannot be recognised as the ultimate source of such initiatives, whether she is willing to endorse them *ex post factum* or not.¹⁹ This analysis points to a necessary formal condition that any process of identification should fulfil in order to effectively convey autonomy to related initiatives. In order to be authoritative in this respect, identification with one's motives should be undertaken under the guise of the good.

A related implication is that some initiatives at issue, such as treatment refusals with foreseeably fatal consequences may turn out not to be self-regarding in the required sense. For they could be pinned down as misplaced reactions of either hostility toward identifiable others or generalised resentment. In both instances, apparently self-regarding initiatives become open to reinterpretation as fundamentally other-regarding and, for this reason, ineligible for protection under the principle of respect for autonomy that was specified earlier.

¹⁸ See esp. steps 8 to 10 of the reconstruction above.

¹⁹ See Radden (2008) for a comparative analysis of two strategies whereby personal identity is conceived in depression memoirs: 'symptom alienating' and 'symptom integrating'. An important conclusion is that both strategies could be equally self-alienating: either by refusing to acknowledge one's mental states as one's own or by internalising a ready-made, extraneous perspective toward one's experiences.

V. Personal Autonomy and the Guise of the Good

It may be tempting to challenge the proposed solution as leading to unacceptable consequences. The challenge is as follows. If, as I argued, paradoxical identification is incompatible with personal autonomy in the relevant sense, protection from interference should be removed from many substantively self-regarding initiatives that some depression sufferers effectively identify with. This move would substantially shrink, if not eliminate the scope for self-determination they are left with. It would equally be at odds with respect for depression sufferers as persons.

This challenge builds on two related assumptions. According to the first, respect for persons is inseparable from respect for their autonomous agency. According to the second, respect for personal autonomy is inseparable from respect for substantively self-regarding initiatives. Jointly, the two assumptions support the idea that the only way to respect a person in the grips of depression is to condone her depression-induced initiatives and, for instance, let her die, if she attempts suicide, as was implied by the Wooltorton case.²⁰

To understand the underlying dialectic and successfully take on the challenge, let us consider the concept of practical identity introduced by Christine Korsgaard. According to Korsgaard, practical identity is “a description under which you value yourself, a description under which you find your life worth living and your actions to be worth undertaking” (1996: 101). For instance, the fact that I am committed to a particular profession gives me a cluster of reasons for action that I would otherwise not have. In a similar vein, defining myself as a theatre enthusiast, a good friend, or a keen cyclist would each speak in favour of certain projects to the exclusion of others. Following this line of thought, Korsgaard concludes:

“It is necessary to have *some* conception of your practical identity, for without it you cannot have reasons to act. We endorse or reject our impulses by determining whether they are consistent with the ways in which we identify ourselves...For unless you are committed to some conception of

²⁰ See step 4.3.1 of the argument scheme in section iii.

your practical identity, you will lose your grip on yourself as having any reason to do one thing rather than another – and with it, your grip on yourself as having any reason to live and act at all.” (*ibid*: 120-21).

In light of these observations, depression appears as a distinct impediment to any practical identity that a depressed person could possibly undertake rather than a separate practical identity. This is not to say that depression may lead to an unreasonable or objectionable practical identity, providing the depressed person with reasons for action that do not make sense from an observer’s perspective. Instead, the thought is that reasons for actions, recognised as valid from the agent’s perspective, get overridden or suspended because of depression and this seems idiotic to the agent herself (Solomon: 1998). We are finally able to resolve the initial paradox posed by depression: the perplexing gap between evaluation and motivation can now be explained as a result of the agent’s paradoxical identification with something that she implicitly loathes. Related initiatives could be compared with the so-called mixed actions, whereby the original link between evaluation and motivation is thwarted by various kinds of undue influence, such as credible threats. Paradoxical identification has a similar effect. Like coercion, the behaviour it affects cannot be considered as fully intentional.

This analogy could remain unnoticed because, in cases of paradoxical identification, the voluntariness of the resulting behaviour is impeded by the agent herself rather than another person. This aspect, however, is neither paradoxical, nor of consequence to the analogy between depression and coercion, which was alluded to earlier in terms of autocracy. In fact, there is nothing astonishing in the idea that an agent can incapacitate herself. It suffices that I take myself to be an inadequate source of actions in order to make myself so. This follows directly from the intuitive understanding of intentional action as trying to achieve certain objective. Clearly, I cannot undertake such an action, unless I implicitly believe that the objective is not unattainable

for me. Hence, it is rather the mechanism underlying the impediment or breakdown of agency brought about by a depressed self that stands in need of explanation.

Drawing on the preceding analysis, this mechanism becomes intelligible in terms of suspended commitment to many, if not all eligible practical identities. As a result, judgements about projects worth undertaking could be seen as merely entertained rather than endorsed by a person in the grips of severe depression. The lack of motivation to act one way or the other that is so striking from both agent's and observer's perspectives, turns out to be better accounted for as a hindrance to the depressed person's capacity to value rather than as a severed link between evaluation proper and motivation.²¹

This development is fully consistent with the idea that an agent can authoritatively identify only with what she cares about. Whenever this formal constraint is not met, the agent's relationship to her motives becomes paradoxical. This could lead to a volitional halt to the extent that some of her own aspirations correctly appear to her as impossible to satisfy. For the intrinsic ambivalence of a paradoxical relationship to one's motives translates into inconsistent volitions, whereby the same outcome is both wished for and abhorred. Since these kinds of volitions are bound to end in frustration, they sustain the depressed person's sense that any endeavour is futile and her life is no longer worth living (Solomon:1998).

Returning to the challenge from unacceptable consequences, we are now in a position to critically examine the two premises, on which it depends. For instance, the analogy between paradoxical identification and coercion points to a flaw in the first premise, according to which we can only respect a person in so far as we recognise her to be an autonomous agent. The thought is that

²¹ An advantage of this solution is that it does not commit us to a particular account of the nature of either evaluation or motivation. For it is compatible with both a Humean account of valuing as a kind of desire (Roberts 2001), and a scholastic view of desires as subjective conceptions of the good (Tenenbaum 2007: 283–298).

respect for persons who have been subject to coercion clearly departs from respect for the choices, they made under duress. In fact, giving credence to the latter would amount to taking sides with the coercers and disregarding the coerced even further. Following this line of thought, it becomes compelling to separate respect for persons from respect for autonomous agency so that instances of forced or improper self-determination do not undermine the possibility for genuine self-determination in the future. In this respect, firm constraints on what a person can authoritatively decide about herself appear to be a prerequisite for rather than an obstacle to her self-determination. The notion of inalienable rights or dignity expresses well this intuition. For only by making certain self-regarding choices ineligible for protection from interference could a society effectively shield individuals from particularly grave kinds of intrusion. The rejection of a right to sell oneself into slavery is a classic example (Mill 1859). Thus, a person may be at liberty to feel and behave as another person's slave, however, she is not at liberty to require that third parties respect this arrangement and treat her as another's slave (Shiffrin 2000).

The preceding example helps articulate the significance of disentangling respect for persons from respect for autonomous agency. For it suggests that self-respect is not a necessary condition for being respected as a person.²² Just as the would-be slave's subservience does not provide us with a reason to treat her as a being of lesser worth, a depressed person's sense of worthlessness does not provide us with a reason to let her die if she attempts to commit suicide.

Furthermore, this observation points to a flaw in the second premise of the challenge from unacceptable consequences, according to which it is exclusively by respecting self-regarding initiatives that we respect personal autonomy. This is consistent with the earlier argument that depression should be understood as an obstacle to having a practical identity rather than an

²² See Langton (2007) for a related argument, according to which the value of persons is best understood as unconditional and, therefore, independent of whether a person happens to value herself at a particular moment in time or not.

alternative practical identity. As indicated by the two thought experiments in the previous section, both the cheated wife and the frustrated professional end up with diminished self-esteem because of a paradoxical identification with a problem that was imposed on them rather than with a solution of their own. Respect for initiatives motivated by reduced self-esteem can only amplify the initial paradox and curtail the scope for self-determination proper even further. Drawing on this analysis, it becomes clear that respect for personal autonomy is incompatible with respect for depression-induced self-regarding initiatives. This is good reason to doubt the objection from unacceptable consequences, for neither of the assumptions on which it depends could survive critical examination.

References

- American Psychiatric Association (2001). *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. (text rev.) [DSM-IV-TR]. Washington, DC: American Psychiatric Association.
- Anscombe, G.E.M. (1963). *Intention*, 2nd ed. Oxford: Blackwell.
- Arpaly, N. (2003). *Unprincipled Virtue: An Inquiry into Moral Agency*. New York: Oxford University Press.
- Beauchamp, T.L. and Childress, J.F. (1979). *Principles of Biomedical Ethics*. New York: Oxford University Press.
- Berlin, I. (2002). *Liberty*. (ed. by H. Hardy). Oxford: Oxford University Press.
- Charland, L. (2008). Decision-Making Capacity. In E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Winter 2008 Edition), URL = < <http://plato.stanford.edu/entries/decision-capacity/>>
- Christman, J. (1988). Constructing the Inner Citadel: Recent Work on Autonomy. *Ethics* 99(1): 109–24.

- Culver, C.M. and Gert, B. (2004). Competence. In J. Radden (ed.) *The Philosophy of Psychiatry: A Companion*. New York: Oxford University Press; 258–270.
- Dworkin, G. (1988). *The Theory and Practice of Autonomy*. Cambridge: Cambridge University Press.
- Feinberg, J. (1986). *Harm to Self*. New York: Oxford University Press.
- Frankfurt, H. (1998). *The Importance of What We Care About: Philosophical Essays*. Cambridge: Cambridge University Press.
- Frankfurt, H. (1999). *Necessity, Volition, and Love*. Cambridge: Cambridge University Press.
- Frankfurt, H. (2002). Reply to Gary Watson. In Buss, S. and Overton, L. (eds.) *Contours of Agency: Essays on Themes from Harry Frankfurt*. Cambridge (Mass): MIT Press; 160–163.
- Freud, S. (1915). *Mourning and Melancholia*. In J. Strachey (ed. and trans.) *The Complete Psychological Works of Sigmund Freud*, vol. 14. London: Hogarth Press and the Institute for Psychoanalysis (1957); 243–258.
- Gallie, W.B. (1956). Essentially contested concepts. *Proceedings of the Aristotelian Society* 56: 167–198.
- Hill, Jr., T. (1991). *Autonomy and Self-Respect*. Cambridge: Cambridge University Press.
- Korsgaard, C. M. et al. (1996). *Sources of Normativity*. Cambridge: Cambridge University Press.
- Langton, R. (2007). Objective and Unconditioned Value. *Philosophical Review* 116(2): 157–185.
- McLean, S. (2009). Live and Let Die. *British Medical Journal* 339:b4112.
- Mental Capacity Act (2005). Office of Public Sector Information, URL = http://www.opsi.gov.uk/acts/acts2005/ukpga_20050009_en_1
- Mele, A. (1987). *Irrationality: An Essay of Akrasia, Self-Deception, and Self-Control*. New York: Oxford University Press.
- Mill, J.S. (1859) *On Liberty*. Indianapolis IN: Bobbs Merrill, 1959.
- O’Neill, O. (2002). *Autonomy and Trust in Bioethics*. Cambridge: Cambridge: University Press

- Radden, J. (2008). My Symptoms, Myself: Reading Mental Illness Memoirs for Identity Assumptions. In H. Clark (ed.) *Depression and Narrative: Telling the Dark*. Albany: State University of New York Press; 15–28.
- Radoilska, L. (2007). *L'Actualité d'Aristote en morale*. Paris : Presses Universitaires de France.
- Radoilska, L. (2012). Autonomy and Ulysses Arrangements. In Radoilska, L. (ed.) *Autonomy and Mental Disorder*. Oxford: Oxford University Press; 252–280.
- Re MB [1997] 2 FLR 426, URL = <<http://www.bailii.org/ew/cases/EWCA/Civ/1997/3093.html>>
- Roberts, J.R. (2001). Mental Illness, Motivation, and Moral Commitment. *The Philosophical Quarterly* 51(202); 41–59.
- Scoccia, D. (2008). In defense of hard paternalism. *Law and Philosophy* 27: 351–381.
- Setiya, K. (2010). Sympathy for the Devil. In S. Tenenbaum (ed.), *Desire, Practical Reason, and the Good*. New York: Oxford University Press; 82–110.
- Shiffrin, S.V. (2000). Paternalism, unconscionability doctrine, and accommodation. *Philosophy & Public Affairs* 29(3): 205–250.
- Solomon, A. (1998). Anatomy of melancholy. *The New Yorker* 12 Jan: 46–61
- Stocker, M. (1979). Desiring the Bad: an Essay in Moral Psychology. *Journal of Philosophy* 76(12): 738–753.
- Taylor, J.S. (2005). Introduction. In J. S. Taylor (ed.) *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*. Cambridge: Cambridge University Press; 1 – 29.
- Tenenbaum, S. (2007). *Appearances of the Good: an Essay on the Nature of Practical Reason*. Cambridge: Cambridge University Press.
- Velleman, D. (1992). The Guise of the Good. *Noûs* 26(1): 3–26.
- World Health Organisation (1992). *The ICD-10 Classification of Mental and Behavioural Disorders*. Geneva: World Health Organisation.