

Voluntary Rehabilitation? On Neurotechnological Behavioural Treatment, Valid Consent and (In)appropriate Offers

Lene Bomann-Larsen

Received: 5 January 2011 / Accepted: 3 March 2011 / Published online: 18 March 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract Criminal offenders may be offered to participate in voluntary rehabilitation programs aiming at correcting undesirable behaviour, as a condition of early release. Behavioural treatment may include direct intervention into the central nervous system (CNS). This article discusses under which circumstances voluntary rehabilitation by CNS intervention is justified. It is argued that although the context of voluntary rehabilitation is a coercive circumstance, consent may still be effective, in the sense that it can meet formal criteria for informed consent. Further, for a consent to be normatively valid (“take the wronging out of the act”) under a coercive circumstance, the subject to be treated must (1) have the sovereign authority to consent, and (2) the offer-giver must be in the right normative position to make the offer. While I argue that subjects do have the sovereign authority to consent to treatment, I also argue that inappropriate offers yield invalid consents. Considerations on inappropriate offers should therefore inform which kinds of CNS intervention-based rehabilitation schemes the state may propose as part of the criminal justice system. Yet as I conclude in this paper, while there are some intrinsic constraints on voluntary rehabilitation programs, the main constraints on

voluntary rehabilitation are likely to be contingent overrides. However, CNS intervention is not ruled out as such in the context of voluntary rehabilitation.

Keywords Voluntary rehabilitation · CNS intervention · Autonomy · Valid consent · (In)appropriate offers

Preliminaries

Within the criminal justice system, offenders may be offered rehabilitation as an alternative to imprisonment. For instance, they may be offered to undergo a rehabilitation program to change their undesirable behavioural pattern, as a condition of early release. It seems humane, and pragmatically wise, to sometimes offer rehabilitation as a voluntary alternative to incarceration,¹ in particular for recidivists who will predictably commit the same type of crime over and over again. In addition to traditional methods, such as cognitive behavioural treatment (CBT), neurotechnological options are on the table. And presumably, increasingly so as we learn more about how the brain functions and how it affects behaviour, and how it can

L. Bomann-Larsen (✉)
Department of Philosophy, Classics, History of Art
and Ideas/Centre for the Study of Mind in Nature,
University of Oslo,
PO Box 1020 Blindern, 0315 Oslo, Norway
e-mail: lene.bomann-larsen@ifikk.uio.no

¹ I am presupposing that treatment versus prison is a voluntary option. Involuntary sentencing to treatment carries with it a different set of questions than those addressed here. Further, I will only be concerned with autonomous, legally competent agents; not psychotics or others who are unfit to be punished and therefore sentenced directly to treatment on paternalistic grounds.

be altered and modified by psychopharmacological and other methods—e.g. Transcranial Magnetic Stimulation—that directly affect the central nervous system (CNS) [1, 2].

At present, CNS interventions are mainly used to correct abnormal or dysfunctional conditions like depression and Parkinson’s disease. However, in the context of voluntary rehabilitation we are discussing treatment of conditions that are, to quote Martha Farah, “medically unremarkable but socially undesirable” [2:26].² Areas where CNS intervention holds a promise to be put to use include anger treatment, weak impulse control, sexual misconduct and bad parental behaviour [2]. Hence the question arises of whether, and to which extent, such interventions ought to be included in rehabilitation programs within the criminal justice system.

The main question to be addressed in this article is under which circumstances voluntary rehabilitation is justified. This is typically a question for consent theory; to be justified, rehabilitation requires consent. Firstly, it is possible that consent is not *effective* under the constraining circumstance represented by a threat of incarceration. By ‘effective’ consent I mean consent given under conditions that satisfy formal criteria necessary for consent to be *real*.³ Most importantly, the subject must freely accept the proposal; e.g., if coerced to “consent”, she is not actually consenting at all.

Secondly, it is possible that there are constraints on what one may consent to, which render consent *normatively invalid* even if formal conditions for consent obtain. By ‘normatively invalid’ consent I mean a consent that fails to take the wronging out of the act done to the consenting person, despite of her un-coerced acceptance of the proposal.

Note that a consent which succeeds in taking the wronging out of an act X only ensures that the consenting person is not wronged by X. There may be

other overriding moral reasons why X should not be done. E.g., the legitimate claims of others may count in and render X all-things-considered impermissible. A valid consent then, only gives a *pro tanto* permission to X. I will largely set the issue of all-things-considered permissibility aside in this article, and concentrate on the normative validity of the consent itself.

There are several ways in which a consent may be normatively invalid; i.e. fail to take the wronging out of X. First, formal consent-conditions may not obtain. I assume that effective consent is a necessary condition for valid consent, since without it, there is no real consent at all. However, it is not sufficient: it seems that some effective consents will fail—in the coercive circumstance of voluntary rehabilitation—to take the wronging out of the act. One way in which consent may fail to take the wronging out of the act is on the theory that there are moral limits to personal sovereignty; that the individual is not in a position to consent because there are limits to her authority over herself, e.g., that there are certain rights she cannot waive. But the idea that one can wrong oneself is a substantive moral idea and difficult to defend within a political-liberal framework. I therefore suggest we focus our attention on the *offer*. Rather than argue that a consent may be invalid for the reason that A does not have the authority to consent to X, I argue that consent to X is invalid if, under the circumstances, the consent has been inappropriately asked. However, not all CNS interventions are inappropriately offered.

Effective Consent in the Context of Voluntary Rehabilitation

Distinguish between Cognitive Behavioural Treatment (CBT) on the one hand and CNS interventions on the other [2]. While CBT appeals to persons as reasons-responsive by making it possible to reject or accept strategies for changing one’s behaviour as one goes along, CNS interventions bypass reasons-responsiveness by interfering directly with the sub-conscious mechanism itself; thus making resistance and acceptance not just obsolete, but impossible.

According to Martha Farah, public acceptance of CBT and ditto scepticism against CNS interventions

² One may wonder if such treatments are not more properly labelled enhancements, since they are not strictly speaking medical. Still, I will refer to them as treatments since they are arguably not improvements beyond normal functioning, but corrections of behaviour that is deemed socially unacceptable. For discussion of the distinction between enhancement and treatment, see [2, 3].

³ I use ‘effective’ rather than ‘actual’ or ‘real’ to avoid alluding to the contrast with ‘hypothetical’ consent, which is not an issue here.

may be explained by an intuition about individual freedom⁴ as

(...) the freedom to think one's own thoughts and have one's personality (...). In the anger management class, a person is free to think: 'This is stupid! No way I am going to use these methods'. In contrast the mechanism by which Prozac curbs impulsive violence cannot be accepted or resisted in the same way [2: 28].

Moreover, Farah suggests that while it is mitigating that the treatment is not forced but offered in a context of choice—with conventional therapies and incarceration as alternatives—the implicit coercion in the situation still represents a problem.

Farah raises two separate but related questions: (1) does the implicit coercion in the situation undermine consent? And (2) is CNS treatment compatible with respect for autonomy?

To start with the second question; we must determine what is the relevance for autonomy of the distinction between treatment that can be resisted/accepted and treatment which intervenes directly into the mechanism. On the one hand, CBT does not seem to be autonomy-undermining at all, insofar as the way it changes the person—if successful—is by means of persuasion, that is, by reason-giving which appeals to the person's autonomous acceptance.

Persuasion is the intentional and successful attempt to induce a person, through appeals to reasons, to freely accept—as his or her own—the beliefs, attitudes, values, intentions or actions advocated by the persuader. (...) All choices made and acts performed on the basis of persuasion are non-controlled (...) and autonomous [4: 262].

Arguably CBT is a type of persuasion, thus understood. It changes the person in a way that is not threatening to autonomy, but rather presupposes and sustains it. No matter the result for one's identity, we may say, autonomy is not undermined by CBT because the autonomous process guarantees an autonomous result. CNS interventions on the other hand, are not persuasions, and if we believe they are

compatible with respect for autonomy, this point must be argued; in particular if the result is *less* autonomy rather than more autonomy. I will return to this latter point in [Are There Constraints on what One may Consent to in Voluntary Rehabilitation?](#) Here I will only argue that the fact that CNS treatment skips the autonomous process is not in itself a reason why the subject cannot validly consent to CNS treatment. An act of consent is an act of "autonomous authorizing" [4:235]. It is true that CNS intervention bars the opportunity to say: Hey, stop, I don't want to continue this process. Thus CNS interventions differ from CBT in that there comes a point—when consent has been given and acted upon—beyond which it cannot be withdrawn; one cannot change one's mind, as it were. Arguably, however, this is true of many autonomous choices upon which we act, both in general and in the form of consents, not least in the medical context. If I consent to a hysterectomy, I cannot withdraw my consent when I wake up from the narcosis. It could be objected that CNS interventions are different because they affect the brain, which is "the seat of our selves"; hence it is all the more important that the procedure takes small and, ideally, revocable steps.⁵ I admit the importance of this. While it does not follow that CNS intervention is principally ruled out for the reason that it affects our identity and that it does so in a particular way by bypassing our reasons-responsiveness,⁶ there is certainly a call for careful consideration both of the constraints on such interventions, and in the development of the procedures themselves. I will return to this towards the end of the article.

Granted that consent to CNS treatment is an autonomous act of authorizing and hence as such not incompatible with autonomy, let us return to Farah's first question: can consent be effective (real) under the threat of incarceration? For consent to be effective, it must be informed and freely given. I

⁴ The kind of freedom she refers to might well be labelled individual *autonomy*.

⁵ I owe this point, and the phrase, to an anonymous referee.

⁶ The same is true of medical procedures that affect the brain, say when the removal of a tumour causes permanent changes in the brain. Of course, there is a difference between accepting surgery to save one's life and accepting surgery to address criminal behaviour, but all I am arguing is that this difference does not bar the permissibility of consenting to the latter.

stipulate that the relevant consent is informed. But what does it mean that consent is *freely* given, and how do we make sense of this notion in the present context?

Consent is effective only if it is the result of a free exercise of choice. Beauchamp and Faden suggest *non-control*—i.e. absence of controlling influences—as the condition of free choice. Non-control is defined by the ability of the subject to *resist* the influence of others, and the degree of resistibility demarcates the degree of non-control. Influence may be choice-restricting or actually choice-facilitating. Only choice-restricting influences like coercions and manipulations are controlling factors in a sense that may undermine consent.

Beauchamp and Faden define *coercion* as occurring when “one party intentionally and successfully influences another by presenting a credible threat of unwanted and avoidable harm *so severe* that the person is *unable to resist*⁷ acting to avoid it.” [4: 339]. In the present context, the coercive element is the prospect of incarceration. However, it may be false to construe imprisonment in this context as a threat of harm. It is surely a coercive circumstance, having to choose between staying in prison or undergoing CNS treatment. The subject is not free to choose any option she wants. But while coercive circumstances do restrain freedom of choice, they do not thereby undermine autonomy. Consider the cancer patient facing the option of undergoing a risky experimental surgery or being doomed to certain death from her disease. She is clearly not free to choose whichever option she wants (presumably, what she wants is get well without risk), but she can still autonomously consent to surgery. What would undermine her consent as autonomous authorizing—short of tying her down and forcing her to undergo surgery—would be to threaten her with repercussions other than natural death resulting from her rejecting the surgery, or manipulating her by way of lies or distorted information about the risks of undergoing or not undergoing surgery.

In the context of voluntary rehabilitation, the subject is already in prison. Thus the incarceration to which the subject has already been sentenced for

her crime may be construed as a factual circumstance; to be aligned with the death from cancer in the medical example. The prison sentence is not a threat *made in order to* make the subject accept CNS treatment, but something to which the convict has made herself liable by—*ex hypothesis*—knowingly and willingly breaking the law. In contrast, were the authorities to say “unless you consent to treatment we will *increase* your sentence”, the convict would be victim to manipulative coercion.

However, this assessment of formal consent-conditions seems to circumvent a serious problem. Consider, as a possible counter-example, the following case described by Joel Feinberg [5], about a proposal from the governor to the prisoner on death row that his sentence will be commuted if and only if he agrees to be part of a medical experiment. “This translates as ‘If you do not agree to be an experimental subject, I will have you executed, and if you do agree, I will commute your sentence.’ [5:218]. This situation is different from that of the cancer-patient because while it is not within anybody’s power to prevent the death of the cancer patient unless she undergoes experimental medical treatment, it is within somebody’s power to hinder the death of the prisoner, independently of whether or not he takes part in a medical experiment. The case as described therefore makes the prospect of death seem like a threat, rather than akin to a fact, and the example appears detrimental to the argument that effective consent may obtain under such circumstances. However, we should maintain that if the prisoner is in fact *liable* to execution, he by stipulation deserves to be killed, and would be killed, were it not for the governor. Hence the governor’s proposal is not a threat after all, but an offer.⁸ In the absence of a threat, it seems the prisoner does *effectively* consent. What he does not necessarily do, is give a *valid* consent; i.e. a consent which is sufficient to take the wronging out of the act (of experimenting on him). Arguably, the offer here made is inappropriate for various reasons, and as I will argue shortly, inappropriate offers fail to take the wronging out of the act despite effective consent.

⁷ My emphasis.

⁸ Feinberg concludes that despite appearances, this is an offer, not a threat [5].

Suffice it to preliminarily conclude that as long as the sentence is fixed and transparent and not subject to change conditional on which choice the subject makes, he is not more coerced than the cancer patient who consents to experimental treatment as an alternative to certain death. Granted a distinction between manipulative coercion (which restricts autonomy) and coercive *circumstances* (which restrict freedom), *effective* consent thus appears obtainable even in the coercive circumstance of voluntary rehabilitation.

Are There Constraints on What One may Consent to in Voluntary Rehabilitation?

Presumably, the aim of CNS-treatment in the context of criminal justice will normally be increased autonomy, by way of strengthening self-control, inhibiting overwhelming urges, and so on. Yet it is not implausible that the end-state of CNS intervention could on occasions be *decreased* autonomy. One reason is that the side-effects of a targeted intervention may be such that they negatively affect the autonomy of the agent (I return to this in [How CNS Interventions may Affect \(De Facto\) Autonomy](#)). The agent may be informed about such possible side-effects and still agree to the treatment. Further, autonomy arguably involves the option of doing wrong, and the capacity to deliberately set aside this option—or not—in the face of competing reasons. Though perhaps far-fetched, it is conceivable that it is in the state's interest to have a population of non-autonomous Moral Saints rather than a population of autonomous wrong-doers, and that the state could therefore push for interventions that decrease autonomy on this score. Admittedly the first possibility is most realistic, but both possibilities make it worthwhile to address not just the question of consent to autonomy-increasing treatment, but the harder case of consent to autonomy-decreasing treatment.

Granted that effective consent may be (formally) obtained in the context of voluntary rehabilitation, let us now move on to the question of whether there are normative constraints on what one may consent to. As suggested above, such constraints are sometimes framed as limitations on the subject's authority; i.e., of duties towards self. E.g., on

a Kantian view an agent has certain duties towards Humanity in her own person, which are morally binding on her choices. She is no more permitted to treat herself as a mere means than to treat others as a mere means, say, by killing herself out of “self-love” [6:1,013–1,014]. Similarly, J S Mill believed that the individual cannot have a right to self-enslavement [7].⁹

It may be morally the case that there are duties towards self. But within the legal-political context of the criminal justice system, where the issue of voluntary rehabilitation must be addressed, Kantian, Millian or similar comprehensive and disputed doctrines must be relegated to the private domain.¹⁰ They should not be made the basis of public institutions [8]. Feinberg's discussion of autonomy is helpful in this context:

Should persons be permitted to consent to treatment that is likely to diminish their freedom by restricting their future options irrevocably, for the sake of a good they have come to value more than their freedom? On the assumption that consent can be fully voluntary in such circumstances, it would be to respect an individual's choice to permit him his (...) manipulative treatment, and it would violate his autonomy to deny it for what we take to be his “own good”. An autonomous being has the right to make even unreasonable decisions determining his own lot in life, provided only that his decisions are truly voluntary (hence truly his own) and do not injure or limit the freedom of others [5:67].

Feinberg helpfully distinguishes between *de facto* autonomy (internal and external conditions for exercising autonomy) and *de jure* autonomy. *De jure* autonomy equals the sovereign right to self-government, and cannot be constrained at all. For that reason, he claims, *de jure* autonomy includes the

⁹ “By selling himself for a slave, he abdicates his liberty; he foregoes any future use of it beyond that single act. He therefore defeats, in his own case, the very purpose which is the justification of allowing him to dispose of himself” [7:1,107]

¹⁰ The Kantian or Millian is free to decline the offer of treatment.

right to give up one's capacity for autonomy (aka the *actual* condition of self-government), for whatever reasons one has. If we take *de jure* autonomy seriously, there seems to be no reason why a sovereign agent may not consent to autonomy-undermining treatment.

However, I cannot consent to renouncing my duties, because my duties are not within the jurisdiction of my self-government, as it were. "(...)even an [de jure] autonomous person cannot alienate certain duties, and the responsibility for what she does" [5:78]. Presumably, what Feinberg means is that *as long as I am* a moral agent I cannot renounce my moral obligations. But the claim could be interpreted in a different way. If I incapacitate myself as an agent, I no longer have duties, and I can no longer be held responsible for what I (fail to) do. Only *de facto* autonomous persons can be bearers of duties and appropriate targets of responsibility attribution. Perhaps I do not have the authority to undermine my moral agency, because that entails renouncing all moral obligations?

To make sense of the notion that I might have an obligation to *stay* a moral agent, we need to determine to whom I have such an obligation. First, I may have *special* obligations to stay a moral agent; e.g., as a parent I have an obligation to make sure I can fulfil my parental duties in the future [9]. Some people then, may not permissibly incapacitate themselves, if by doing so they violate the legitimate claims of others. As said at the outset, valid consent only gives a *pro tanto* permission; it takes the *wronging* out of the act, but not necessarily the *wrongness* out of the act. The legitimate claims of others could trump a valid consent.

Second, citizens may have duties towards the state (i.e. the democratic collective), which lie outside the domain of their sovereign self-government. Citizens may arguably have an obligation to retain a least some *de facto* autonomy in order to fulfil their democratic duties. If this is so, the citizen cannot validly consent to *de facto* autonomy-undermining treatment by reference to her *de jure* autonomy because she is bound by an obligation which is not a direct duty to herself, but derived from her duties as citizen, and those duties lay beyond the scope of her authority.

The problem with this argument is that if the obligation to remain autonomous is ultimately a duty towards the state, the state apparently has the authority to release the subject from this duty. While the duty to stay autonomous cannot be *waived* because it is not within the authority of the single subject to do so, the subject can be *released* from the duty by the authority of the state; and, presumably, the state also has the authority to set the conditions for its forfeiture.¹¹ Thus the state can release the subject from her duty to stay *de facto* autonomous, without violating her *de jure* autonomy (consent is still asked).¹² Whether it lies within the authority of the citizen to renounce her *de facto* autonomy, or she has a duty towards the state to stay *de facto* autonomous—a duty from which the state has the authority to release her—the result is the

¹¹ It could further be objected that it does not lie within the authority of the state to offer its citizens to give up their *de facto* autonomy, because the state draws its legitimacy from expressing the will of the people. And a polity without *de facto* autonomous citizens has no will. The state is therefore not in a position to offer its citizens to undermine the very autonomy in virtue of which it acquires its legitimacy. It is true that the *de facto* autonomy of citizens is *constitutive* of the democratic state. Can the democratic state legitimately offer treatments that ultimately undermine its own existence? It would certainly not be prudential to do so, but the prudential concern only arises if all, or very many, citizens were at risk of being turned into non-autonomous "vegetables". But we do not conduct that kind of generalization when assessing the justifiability of incarceration or capital punishment (arguably, the democratic state would cease to exist if more or less the entire population was in prison, or dead), so why should it make a difference in the context of rehabilitation? The democratic state is not threatened by the fact that quite a few inhabitants are not *de facto* autonomous. There may be a threshold somewhere, but arguments derived from democratic legitimacy do not seem to work against the permissibility of the state offering autonomy-undermining treatment to some of its citizens.

¹² We should grant that the State's legitimacy as a whole is derived from the degree to which it respects its citizens as *de jure* autonomous, but this is of no help to us here. For instance, while respect for *de jure* autonomy may render it illegitimate of the state to deny suffrage to its citizens (even those imprisoned), since the right to self-government arguably implies the right to vote, it does not imply that the state cannot offer citizens to renounce their *de facto* autonomy. All it implies is that the state cannot take away citizens' *de facto* autonomy without their consent.

same. There is no inalienable obligation to *de facto* autonomy, and consent suffices to renounce it.¹³

However, it could be that moral agents have a *general* moral obligation to ensure that they can fulfil their future moral obligations, on the same basis as we have special obligations of this kind [9]. If so, *de facto* autonomy, as a prerequisite for moral agency, may not be renounceable, either by the subject or the state.¹⁴ This is a plausible response, but I doubt it works in the context. First, it is disputable whether we have a duty to remain able to fulfil unspecified positive obligations. It may well be that I have no right to turn myself into a future killing machine or a child molester. That follows from the harm principle. But renouncing *de facto* autonomy is not just about renouncing the capacity to fulfil positive obligations, it is also renouncing one's capacity for doing harm. If we believe negative obligations not to harm beat positive obligations to benefit, then the trade-off indicates that it is better not to be an autonomous agent at all, than to be an autonomous agent with a propensity to harm people. Or to put it otherwise; if the price for losing an uncontrollable urge to assault and rape people is loss of motivation to support starving children in Sudan, then it seems worth paying. That there should be a general *obligation* to stay an autonomous agent, against one's will and disregarding of how one executes this agency thus seems to be a much harder case to argue than the duty to remain able to fulfil special obligations.

¹³ The obligation of the state to respect citizens as free and equal—understood as political persons capable of having and revising their own conception of the good; “to revise their wants and desires” [10: 261]—could of course entail more than respect for *de jure* autonomy alone. It could be argued that the state cannot respect one as free and at the same time offer one to give up one's capacity for freedom; i.e., for changing on one's own accord one's conception of the good. I grant that this might be a counter-argument against offering globally autonomy-undermining treatment, and I address this problem to some extent in (In)appropriate Offers by the State. Despite this, it seems that respect for citizens' entitlement to their own conception of the good generally entails allowing them to give up their autonomy for some (to them) greater good (say, peace of mind). Hence it seems that voluntary CNS treatment outside of the criminal justice context is not permissibly prohibited. If people want to change their behaviour by various CNS means, the state should not make it illegal to do so. This does not, of course, imply anything by way of morally obliging the state—or anyone else—to offer such treatments.

¹⁴ I owe this point to an anonymous referee.

None of these considerations explain why a subject does not have the authority to renounce her *de facto* autonomy by consenting to autonomy-undermining treatment. The first leg of valid consent therefore seems securely fastened; personal sovereignty is not constrained by a general duty to stay autonomous, all else equal. Now, before proceeding to the discussion of the second leg of valid consent—constraints on offers—let us look at how CNS interventions may affect autonomy in more detail.

How CNS Interventions may Affect (De Facto) Autonomy

The result of CNS treatment may be both more and less autonomy for the subject.¹⁵ Define autonomy as the *capacity* to act according to one's own decisions, without the controlling influence of others, and to form these decisions on the basis of one's own beliefs, desires and values. Enhancing the capacity to form decisions and act on them, say, by strengthening self-control, presumably increases autonomy. But as mentioned above, interventions could also decrease autonomy, by making it harder, or impossible, for an agent to form or act on certain types of decisions.¹⁶

The capacity to form (one's own) decisions and act on them does not only track rationality and the absence of controlling influences; it arguably also tracks the *will* to act. ‘Capacity’ in this sense is not an inert faculty, but a power; a motivational force. An agent wholly lacking in motivation cannot be autonomous. Thus CNS interventions which impair the motivational capacity of an agent decrease her autonomy.

Assume a rough distinction between lower and higher order pro-attitudes. Distinguish between two ways in which an agent may be motivationally incapacitated:

(1) She has a higher order pro-attitude to do X but cannot motivate herself to actually do X. (2) She lacks any higher-order pro-attitude to be motivated to do X.

¹⁵ In this section ‘autonomy’ refers exclusively to *de facto* autonomy.

¹⁶ Again, imagine the Moral Saint, who can only do good. The Moral Saint is not autonomous since she does not in any real sense form decisions on the basis of weighing reasons. Her actions are automatic rather than autonomous.

In (2), there appears not to be a problem. It is not a problem (autonomy-wise) that I cannot bring myself to be motivated to do X if I have no reason to be motivated to do X.¹⁷ Is there a problem in (1)?

Suppose a sex offender undergoes chemical castration so he is no longer motivated to abuse strangers. As a side-effect of the treatment he is also no longer motivated to have sex with his wife. The drug is not selective; it is all or nothing. However, he wants to be motivated to have sex with his wife. It is very important to him. Is the man now less autonomous, since he cannot be motivated to do what he wants to be motivated to do with his wife?

The answer is, I think, that it depends. Say our chemical castrate decides to have sex with his wife, but cannot perform. If he had been forcefully castrated, he would clearly be less autonomous; his lack of capacity to act would reflect someone else's decision, not his. But given consent, the reason for his performance failure is his previous decision to incapacitate himself; a decision entailing acceptance of the cost of not being able to do what he wants to do with his wife.¹⁸ He has tied his own hands. Thus despite the conflict internal to his motivational structure, the castrate is still autonomous. To claim otherwise would commit us to the implausible view that if A binds herself by an informed choice of some sort at t1, she is not autonomous at t2, if she at t2 has a desire to do something which is ruled out by her previous choice. But that is absurd. If A decides at t1 to have a child, A cannot complain at t2 that not being able to go to Bali on her own undermines her autonomy. Consenting to treatment which inhibits motivation locally therefore seems not to make an agent less autonomous, even if it rules out certain desirable actions.¹⁹

The agent would be clearly less autonomous if his motivational capacity was inhibited globally. A non-

targeted treatment which unselectively inhibited all affective states could incapacitate the agent *tout court*. An agent left with no desires strong enough to act on is hardly an autonomous agent at all. The difference between autonomy-undermining and non-autonomy undermining treatment could possibly be explained by global effects on motivational capacity, as contrasted to local effects on the motivational structure. Yet we could imagine targeted and local inhibiting of someone's motivational structure having global impairment as a side-effect. The person who takes sex to be incredibly important may, upon losing his sex-drive, also lose the ability to see the point of anything, and hence lose motivational capacity overall.

However, as argued above, it is not given that it does not lie within the authority of an agent to consent to autonomy-undermining treatment for whatever reasons she has. For some, being less autonomous—perhaps lacking in motivation to do much of anything—could still be considered the better option.

This conclusion may seem uncomfortably permissive and in principle allow for unconstrained modification of socially undesirable behaviour. But it need not be. In the next two sections I will discuss the second leg of valid consent, and argue that consent is invalid if the offer is inappropriate. On the basis of an appropriateness-constraint on valid consent, we can maintain that it is not within the state's jurisdiction to offer global autonomy-undermining treatment, because behaviour that is not strictly related to the crime lies beyond the scope of state authority.²⁰

(In)appropriate Offers

Acceptance made under a threat is not an effective consent: "Your money or your life? Your choice!" I will set coercive pressures [5] created in order to force compliance aside. But also non-created coercive circumstances can be exploited: "I will help you out of the water if you give me all your money." Here, consent is formally speaking *effective* because there is

¹⁷ Unless one believes that the fewer life-options a person has, the less autonomous this person is. I will set that issue aside here, as it would make for a paper of its own.

¹⁸ Provided, of course, that he was informed about side-effects.

¹⁹ It could be objected that even if the castrate is still autonomous, he is less autonomous than before. But is that true? Pre-intervention, his inclination to abuse strangers conflicted with his higher-order pro-attitudes, so it seems there was already a deficit in his capacity for autonomy. In fact, he may therefore be more autonomous post-intervention, because his behaviour now at least is the result of his own choice.

²⁰ I present this as the main reason, and focus my discussion on that dimension. It may also be that it is impermissible for anyone to offer such treatment to others because the offer fails to respect the other as a moral person.

no threat. In fact, it is an offer.²¹ Being in the water is something you would be even if I did not make the offer. Whether you decide to give me your money or decide to stay in the water is your autonomous choice (though of course, it is not a free choice, since your freedom is constrained by being in the water). But is your consenting to give me all your money a *valid* consent? If it is, it follows that I am not *wronging* you by taking all your money, and it seems that I am. There is something wrong with the proposal; I try to gain unfair advantage by your misfortune. It is a *coercive offer* [5].

If the option offered is modified, it appears more acceptable: “I will help you out of the water if you do me a favour in return.” Here too I try to gain advantage from your misfortune—I could just offer to help you out—but I do not seem to be wronging you simply by stating this condition. On balance, I do not seem to be gaining an unfair advantage, since you are also gaining an advantage; hence the deal looks more like a transaction between equals than an exploitation of someone’s misfortune. Though I am not acting in a morally recommendable way, your consent to do me a favour in return appears valid; and when you later help me, say, fix my fence, you do not have a complaint that I am somehow wronging you by this. “You owe me one”.²² This suggests that some kind of symmetry between the offer and the circumstance is normatively significant. The offer should not be “excessively profitable” [5:252]. Still, I doubt that asymmetry between the circumstantial option and the offered option can fully explain what makes an offer inappropriate.

Consider a case where A says to B that: “I will pay medical help for your (otherwise dying) child if I get to have sex with you” [5:238].²³ On a commonsensical view, the offered option is less harsh than the circumstantial option, but the offer still seems unacceptable, and not only because A exploits B’s bad circumstances for his own advantage. If A says “I

will pay medical help for your (otherwise dying) child if C gets to have sex with you”, it does not do much of a difference, even if A does not gain any advantage. This example suggests that certain offers are wrongfully made in coercive circumstances, proportionality-considerations aside.²⁴

A further complicating matter arises when we ask *why* certain offers are inappropriate. *Ex hypothesis*, if A consents to X, A is not wronged by X. By consenting to X A changes the moral value of X. Thus there is (*ceteris paribus*) nothing wrong with X’ing A; A has released B from the duty not to X, and presumably B acts permissibly by X’ing A. If A’s consent renders X morally okay, what is it then that makes it inappropriate to offer X?

The answer, I think, is that the offer itself is a wronging. It fails to recognize the other as a moral equal. Exploiting someone’s vulnerability or excessively profiting from a situation is generally inappropriate because the very speech-act is a violation of the other’s fundamental claim to moral respect. If A is already wronged by the offer to X, her consent to X cannot take the wronging out of X’ing her.

The examples suggest that there are certain offers that no one is in a normative position to ask of anyone because the offer itself amounts to a wronging. But there is another dimension to appropriateness as well. Some options which B would be wrong to offer A, are not wrong of C to offer D—even if the offer is coercive. Special relations create special obligations, but also special permissions. Consider an example: A wife asks her husband to choose between his gambling and his marriage. This can be construed both as a threat and an offer: “If you stop gambling, I will stay with you, but if you do not quit, I will leave you.” The offer is certainly coercive. But it does not seem that she is wronging him. Similarly, the state can say: “If you do not serve the military, we will send

²¹ Feinberg has a very sophisticated discussion of the relation and distinction between threats and offers, and their implications for consent. For present purposes, I will just pick out those elements that are clearly relevant to the question of voluntary rehabilitation.

²² In fact, this may be true even if I do not explicitly state the conditions.

²³ Paraphrased.

²⁴ In this example, A’s offer displays that A is a bad person. We could change the example to something like: A offers to save B’s child if B accepts to have sex with C, in order to distract C from setting off a bomb that would kill the entire city’s population. In this case, A makes the offer for a moral reason, and great good will come of it. However, while it might be true that B should have accepted to have sex with C in order to save the city, it is still a question of whether it is a fair offer to B, given the coercive situation she is in, and hence whether B can be said to have consensual sex. Moreover, it is questionable whether A is in a position to make this offer to B at all. I owe this example to Kasper Lippert-Rasmussen.

you to prison.” What is special with these (coercive) offers is that they cannot be appropriately made by anyone else. The wife and state are, *qua* wife and state, in a privileged position to make a certain kind of coercive offers. But substitute the wife for the state, and the inappropriateness becomes evident.

This suggests that for an offer to be appropriate it must be made by someone who is in the right normative position vis-à-vis the other, and some offers will be inappropriate *tout court* because no one is in a right normative position to make them. In the dying-child example, B is not in a normative position vis-à-vis A to coercively offer her to have sex with him (no one is). This is what makes the offer inappropriate, and hence what renders A’s consent null and void.²⁵ In the gambling-example, the wife, and only the wife, is in the right normative position to make a coercive offer.

(In)appropriate Offers by the State

In the following I will specify how the “appropriateness-constraint” operates in the context of voluntary rehabilitation. But first a caveat: as explained in the previous section, there are many factors that can make an offer inappropriate, or wrongful, and hence fail to yield valid consent. For instance, it may well be that an offer to give up one’s *de facto* autonomy is incompatible with respect for someone’s moral standing, and thus inappropriately made in all contexts.²⁶ But for the purpose of this paper, it suffices to consider how offers are inappropriately made by the state, also for distinguishing between treatment that aims at undermining autonomy and treatment aiming at correcting

specific problems that may have autonomy-undermining side-effects. As we shall see, it suffices for ruling out the former kind of offer in voluntary rehabilitation that the offer is inappropriately made by criminal justice authorities, for the reason that it lies outside the scope of criminal justice authority.

As Antony Duff has pointed out, criminal responsibility is a matter of “who is answerable for what, to whom” [11:19]. There are some acts for which we are responsible to the state and hence must answer for to the state, but many acts for which we are responsible only to non-state agents; be if friends, spouses or employers.²⁷ I suggest that what citizens are answerable for to the state determines the scope of behavioural conditions for which the state can appropriately offer convicts treatment; otherwise, the state illegitimately trespasses the borders of its jurisdiction vis-à-vis citizens. Not all wrongs are “public wrongs” [11], and not all socially undesirable behaviours are the state’s concern.²⁸

Therefore, only behaviour that is criminal—more specifically, the behaviour for which the agent is convicted in the particular case—falls within the scope of what the state is in a position to offer to rehabilitate. If, in addition to being a sexual offender, a man is also in general an unpleasant person and a notorious liar, neither of these are crimes; and the state has no authority to offer him treatment for those personality traits.

The appropriateness-constraint helps us address the challenge to consent under coercive circumstances raised in [Effective Consent in the Context of Voluntary Rehabilitation](#). The governor who offers the man on death row to commute his sentence if he agrees to a medical experiment makes an offer beyond the legitimate scope of the governor’s authority. The conditions attached to the commuting of the sentence are therefore inappropriate (even if formal consent-conditions obtain), and the reason is not that participating in a medical experiment is risky

²⁵ Consent is only relevant when the offer is made in coercive circumstances; otherwise one is free to ignore the offer. Of course, an offer made in un-coercive circumstances can still be demeaning, and thus inappropriate, in the sense that it fails to respect the other as a moral equal.

²⁶ I was reminded of this important point by an anonymous referee. Note that if this is so, the offer is also inappropriate if given e.g. in a psychiatric ward. One could also think that treatment that had autonomy-undermining side-effects was generally ruled out in virtue of being a demeaning offer. But imagine the paedophile who, in despair about what he might do, asks his doctor for a chemical castration, and who accepts the foreseeable side-effect that this might have severe impact on his motivational capacity. To me it seems more disrespectful to deny him treatment than to offer it.

²⁷ And for some acts, we are presumably answerable to the universal moral community.

²⁸ A further consideration is related to treatments that create permanent changes in the person. There may be limits to what one can validly consent to on behalf of one’s future self. I will not explore this thesis here, but it could be the case that today’s citizen is simply not in a position to consent on behalf of her future self, and making her the offer would be making the offer to the wrong person.

or dangerous. The matter would be the same if the governor offered to commute the sentence on the condition that the convict agreed to serve him coffee every day for the rest of his life. It is not the severity of the offer, but the inappropriateness of it, which renders the consent invalid. Our intuitions on invalid consent may weaken when the option is serving coffee rather than taking part in a medical experiment, but all the same; the offer is inappropriate because the governor, *qua* governor, has no right to ask the convict to be his servant, any more than he has the right to ask him to take part in medical experiments.

Of course, it is presupposed here that the medical experiment in question is not a form of treatment for the benefit of the convict, but rather an experiment to which he will be used as a means for the possible benefit of others. It could still be a concern that the sheer irresistibility of the prospect of execution makes consent to treatment (which, *ex hypothesis*, the governor is in a position to offer) invalid as well, even if the situation of voluntary rehabilitation is not described as one of threat, but as an offer under coercive circumstances. What difference does it make to the validity of the consent if the governor were to say “I will commute your sentence if you agree to undergo an invasive operation into your ventromedial prefrontal cortex”? Even if the convict has made himself liable to execution, there still seems to be a difference relevant to consent between the governor’s offer and the one made by the doctor to the cancer patient: “This tumour will kill you unless you consent to having it removed by surgery (but the surgery might cause some changes in your ventromedial prefrontal cortex)”.

One way of solving this problem could be to deny the option of voluntary rehabilitation to convicts on death row, for the reason that their consent cannot be valid under the circumstance. This solution is a bit odd, though, like saying: “Since we respect your sovereign right to self-determination we will not offer you treatment for the behaviour for which we will kill you, since we are afraid that you might feel pressured to accept (since we will otherwise kill you), and thus that your consent to the treatment is invalid. If we could offer it, you could walk freely, but that would amount to wronging you, so we will kill you instead”. One could suspect that this oddity indicates a deeper problem with capital punishment. Without further pursuing that line of thought, suffice it to say that

obtaining valid consent to voluntary rehabilitation is at least possible within jurisdictions not practising capital punishment (which encompass most jurisdictions anyway). After all, voluntary rehabilitation is to be offered as a condition of early release, implying that eventual return to society is the final goal.

The appropriateness-constraint implies that the solution (treatment) should aim narrowly at the problem (the criminal behaviour for which the convict is answerable to the state). A global motivational incapacitation of an agent would of course solve the local problem of, say, criminal sexual behaviour, but it would “over-solve” it. A ‘problem/fix’-requirement implies a necessity-constraint (rather than a proportionality-constraint); ideally, the treatment should not go beyond what is necessary in order to correct the behaviour for which the criminal is imprisoned. This does not necessarily rule out the permissibility of offering treatment that may, as a foreseeable side-effect, have global impact on autonomy, such as in the chemical castration-case discussed in section 4. But it does give strong reason to aim to develop targeted forms of treatment. Moreover, it rules out treatment that directly aims at undermining autonomy, that is, interventions that leave the subject literally a non-agent. In other words, though I argued in section 3 that one has the authority to consent to autonomy-undermining treatment, the offer to undergo such treatment may be inappropriate. It does not lie within the authority of the criminal justice system to offer to reduce convicts to wards, even if that in some extreme cases of dangerous recidivists could be socially desirable.

Now, it could be tempting to draw the conclusion that if the crime committed warrants a severe form of punishment, then the treatment-option could justifiably be proportionally severe. One could imagine that a long sentence deserved warrants a more invasive treatment than a short sentence; that the treatment offered to a serial killer could, for that reason, be more invasive than the option offered to a kleptomaniac.²⁹ But desert-

²⁹ The length of the sentence could possibly be an incentive to accept more invasive treatment that one might not have accepted if the sentence was shorter, because a long sentence is arguably a more constraining circumstance than a short one, but that is a different matter than proportionality and can be solved practically (e.g. by only offering treatment when the remaining sentence is *n*).

considerations are irrelevant to treatment. Allowing desert-considerations to play a role in determining which treatments could be justifiably offered would require that we justify CNS interventions as a form of *punishment*, say: “For murder the punishment is a full lobotomy”. That would require a total reconsideration of the penal system, and a setting aside of the fundamental principle that punishment should not be ‘cruel and unusual’. It would raise a very different kind of debate about the role of the penal system, and the justifications of punishment in general. Moreover, if CNS interventions were construed as punishments, they would not need actual consent to be justified; punishment presupposes merely tacit consent. By breaking the law, the perpetrator presumably willingly forfeits certain rights, and can be said to, in virtue of the criminal act, tacitly consenting to being imprisoned. But the court need not ask the perpetrator for her actual consent to the sentence.

Summing up, desert-considerations are irrelevant to voluntary rehabilitation.³⁰ Treatment should not be conflated with punishment. What should be considered is the fit between the problem (e.g., uncontrollable violent behaviour that has led to a criminal act) and the solution (e.g. inhibiting the propensity to violent behaviour). The treatment should aim at correcting the behaviour on the basis of which the subject is convicted, not with an eye to punishment warranted by the crime. Moreover, the treatment should be as narrow and targeted as possible. Though targeted treatment with the foreseeable side-effect of negatively affecting global motivational capacity is not inappropriately offered, it is inappropriate to offer global autonomy-undermining treatment. The subject may therefore be wronged by such treatment even if she should effectively consent.

Conclusion

I have argued that in the context of treatment versus incarceration, *effective* consent is obtainable even if given under coercive circumstances. Granted the premise that the subject has a sovereign right to self-

³⁰ Because it must circumvent desert-considerations, the justification of voluntary rehabilitation is unlikely to make retributivists happy.

government, consent to treatment is also *valid* on the one leg, even when treatment is (de facto) autonomy-undermining, because such consent is consistent with *de jure* autonomy. Consent in coercive circumstances is valid on the other leg only if the option is offered by someone who is in the right normative position to make the offer, and lies within the domain of what that agent is in a position to offer (the appropriateness-constraint). It is suggested that the state may only offer solutions to behavioural problems for which the convict is answerable to the state. This seems to rule out certain globally autonomy-undermining interventions, but not necessarily interventions that as a side-effect decrease autonomy by negatively impacting on the motivational set of the subject. Still, it gives a strong incentive to develop narrowly targeted treatments. Finally, a valid consent only gives a *pro tanto* permission to X. Valid consent may fail to give an all-things-considered permission to X if overridden by the valid claims of third parties. Presumably will global treatment that has autonomy-decreasing side-effects on occasions be inconsistent with fulfilling one’s duties towards third parties. Hence even if consent to such treatment is normatively valid in the sense that subject is not wronged by it, the treatment may still in some cases be impermissible. While this constraint is contingent, it could be significant in restraining public policy and practice. That said, voluntary rehabilitation is most likely to be guided by an aim to increase, rather than decrease, the autonomy of subjects, and as such, this article offers an argument in favour of voluntary rehabilitation by CNS intervention.

Acknowledgments Earlier versions of this paper have been presented at Nasjonal Fagkonferanse i Etik, Oslo, Nov 2010, Nordic Network in Political Theory, Lysebu, Oct 2010, and the seminar “Sentenced to Treatment – on Therapeutic Justice and Behaviour Control”, Oslo, Sept 2010. I owe thanks to participants at these events, and in particular to Jakob Elster, Silje Langvatn, Kasper Lippert-Rasmussen, Stephen Morse, Nicole Vincent, and an anonymous referee, for their thorough comments. I am especially indebted to Robin Pierce for her paper “Setting Margins for Therapeutic Justice: Circumscribing Consent”, presented at James Martin Advanced Research Seminar Series, February 3, 2010, at Oxford, UK, which inspired the research question of this paper.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Tancredi, Laurence R. 2004. Neuroscience developments and the law. In *Neuroscience and the law: Brain, mind and the scales of justice*, ed. Brent Garland, 71–113. New York: Dana Press.
2. Farah, Martha. 2007. Emerging ethical issues in neuroscience. In *Defining right and wrong in brain science. Essential readings in neuroethics*, ed. Walter Glannon, 19–36. New York: Dana Press.
3. Levy, Neil. 2007. *Neuroethics. Challenges for the 21st century*. New York: Cambridge University Press.
4. Beauchamp, Tom L., and Ruth R. Faden. 1986. *A history and theory of informed consent*. New York: Oxford University Press.
5. Feinberg, Joel. 1986. *Harm to self. The moral limits of the criminal law*. New York: Oxford University Press.
6. Kant, Immanuel. 1992. Grounding for the metaphysics of morals. In *Classics of moral and political theory*, ed. Michael L. Morgan, 1044–1115. Indianapolis: Hacknet Pub.
7. Mill, John Stuart. 1992. On liberty. In *Classics of moral and political theory*, ed. Michael L. Morgan, 1044–1115. Indianapolis: Hacknet Pub.
8. Rawls, John. 1991. The idea of an overlapping consensus. In *Rawls: Collected papers*, ed. Samuel Freeman, 421–448. Cambridge: Harvard University Press.
9. Zimmerman, Michael J. 2007. *The concept of moral obligation. Cambridge studies in philosophy*. New York: Cambridge University Press.
10. Rawls, John. 1991. A Kantian conception of equality. In *Rawls: Collected papers*, ed. Samuel Freeman, 254–266. Cambridge: Harvard University Press.
11. Duff, Antony. 2007. *Answering for crime. Responsibility and liability in the criminal law. Legal theory today*. Oxford: Hart Pub.