

MAKING A CLEAN BREAK: ADDICTION AND ULYSSES CONTRACTS

CHRISOULA ANDREOU

Keywords

addiction,
advance directive,
compulsion,
mental illness,
self-control,
temptation,
Ulysses contract

ABSTRACT

I examine current models of self-destructive addictive behaviour, and argue that there is an important place for Ulysses contracts in coping with addictive behaviour that stems from certain problematic preference structures. Given the relevant preference structures, interference based on a Ulysses contract need not involve questionably favouring an agent's past preferences over her current preferences, but can actually be justified in terms of the agent's current concerns and commitments.

I

Addicts often contemplate the future with anxiety and search desperately for effective measures to help them avoid tempting but destructive conduct. In an effort to control her appetite for the destructive conduct to which she is prone, an anxious addict may pursue strategies that either facilitate voluntary restraint or incorporate constraint. To facilitate voluntary restraint, an addict may arrange her life so that giving in to temptation will have high immediate costs, rather than just high delayed costs. For example, an addict may make side bets (where the stakes need not be financial) that increase the costs of giving in to temptation. She may, for instance, bet that she will quit drinking and arrange things so that the cost of her losing the bet is that someone from whom she has been hiding her addiction (perhaps a family member or employer) will be informed about it. In terms of strategies that incorporate constraint, an addict may recruit others to stop her from opting for the tempting but destructive conduct that she contemplates with anxiety. This was, in essence,

Ulysses' strategy when, fearing that he would be lured into danger by the song of the seductive Sirens, he instructed his companions to tie him to the mast of his ship and to secure him with even more bonds if, when hearing the song of the Sirens, he sought to be set free. At the core of this strategy is what has since been called a *Ulysses contract*, which features a formal or informal advance directive that instructs certain others to force one to do or refrain from doing something regardless of one's anticipated resistance or protestations.¹

An addict whose addiction is causing trouble for others might be pressured or enticed into forming a Ulysses contract. Here my concern is with Ulysses contracts that are initiated by the addict from a desire for self-control rather than prompted by pressure or enticements from others.

Thomas De Quincey's *Confessions of an English Opium Eater* describes an opium addict's attempt at employing a Ulysses contract with the aim of

¹ As I will explain shortly, Ulysses contracts, or at least those I am concerned with here, differ significantly both from traditional advance directives and from standard contracts.

eliminating addictive behaviour. According to De Quincey, a fellow opium addict, namely Coleridge,

went so far as to hire men – porters, hackney-coachmen, and others – to oppose by force his entrance into any druggist’s shop. But, as the authority for stopping him was derived simply from himself, naturally these poor men found themselves in a metaphysical fix. . . . And in this excruciating dilemma would occur such scenes as the following:

‘Oh, sir,’ would plead the suppliant porter, ‘really you must not. . . . Didn’t you say no longer ago than yesterday –’

‘Pooh, pooh! Yesterday is a long time ago. Are you aware, my man, that people are known to have dropped down dead from timely want of opium?’

‘Ay, but you tell’t me not to hearken –’

‘Oh, nonsense. An emergency, a shocking emergency, has arisen – quite unlooked for. No matter what I told you in times long past. That which I *now* tell you, is – that, if you don’t remove that arm of yours from the doorway of this most respectable druggist, I shall have a good ground of action against you for assault and battery.’²

Ulysses contracts differ significantly from both traditional advance directives and standard contracts. Whereas Ulysses contracts are designed to thwart certain anticipated future wishes, traditional advance directives are designed to communicate one’s will with respect to possible future scenarios, such as permanent unconsciousness, in which one will not be able to make treatment decisions for oneself. Because Ulysses contracts are designed to thwart future wishes, they mandate that the ‘benefits’ due be forced onto the one to whom they are due, even if the ‘beneficiary’ clearly expresses wishes that the ‘benefits’ not be provided. They thereby also differ from standard contracts. With respect to a standard contract, a party who sought to procure a good or service (such as protection) by means of a

contract could, given a change of heart, release the other(s) involved from providing the good or service. If, for example, I agree to pay you to be my bodyguard tomorrow, but then change my mind about using your services, I can normally release you from serving as my bodyguard (which is not to say that I can release myself from having to provide you with the funds and any other benefits you are due according to the contract). Because Ulysses contracts mandate that the ‘benefits’ due be forced onto the one to whom they are due, the moral legitimacy of Ulysses contracts is much more hotly contested than the moral legitimacy of traditional advance directives or of standard contracts.

Although the image of Ulysses tied to the mast can be found in philosophical work concerning addiction, sustained philosophical discussion of Ulysses contracts has figured primarily in work concerning mental illness.³ Debate has focused on the issue of whether a Ulysses contract drawn up by a patient with the cooperation of his psychiatrist can legitimately be enforced in the contexts in which it is most needed, namely, when the patient is in the early stages of relapsing into mental illness but before the relapse has progressed to the point where the patient has lost his status as a competent agent. By the time a patient has lost his status as a competent agent, (limited) interference with the patient’s conduct aimed at protecting his vital interests is generally recognized as legitimate, whether or not the patient formed a Ulysses contract, or even just a plan or settled preference, in anticipation of his current situation. But in the early stages of relapse, when the symptoms of relapse may include little more than resistance to continuing a course of treatment, interference with the patient’s conduct, though tempting, is questionable, since, if the patient’s preferences concerning treatment have changed – if, for example, the patient now sees the side effects of the treatment as worse than the risk of relapse – there

² T. De Quincey. 1994 [1821]. *Confessions of an English Opium Eater*. Hertfordshire: Wordsworth Classics: 22–23.

³ See, for example, R. Dresser. Ulysses and the Psychiatrists. *Harv Civ Rights-Civil Lib Law Rev* 1982; 16: 777–854; R. Dresser. Bound to Treatment. *Hastings Cent Rep* 1984; 14: 13–16; M. Quante. Precedent Autonomy and Personal Identity. *Kennedy Inst Ethics J* 1999; 9: 365–381; J. Savulescu & D. Dickenson, The Time Frame of Preferences, Dispositions, and the Validity of Advance Directives for the Mentally Ill. *Philos Psychiatr Psychol* 1998; 5: 225–246; R. Spellecy. *Kennedy Inst Ethics J* 2003; 13: 373–392.

seems to be no basis for respecting the preferences of his past self, as expressed in the Ulysses contract, over the preferences of his current self.⁴ Hence there is the worry that interference based on a Ulysses contract involves an unwarranted ‘departure from respect for the individual’s present free choice’.⁵ But the appeal to freedom introduces complications: on the one hand, it has been suggested that respecting freedom involves respecting an agent’s freedom to bind himself; on the other hand, it has been suggested that being free involves *both* having the strength of will to carry out one’s prior decisions if one does not change one’s mind *and* having the freedom to change one’s mind.⁶

This paper will shift attention to the usability and morality of Ulysses contracts in the context of addiction. I focus on addiction because some interesting theoretical research in this area can (and here will) serve as the foundation for novel theoretical and applied results concerning Ulysses contracts. The broad philosophical question at the heart of this piece is ‘Are there preference structures that warrant interference based on Ulysses contracts?’ Based on an examination of current models of self-destructive addictive behaviour, I will argue that the answer is ‘yes’. I will consider only informal – rather than legally binding – Ulysses contracts between an addict and her friends or family. As will become apparent, while some situations involving addiction are well-suited to the use of Ulysses contracts, others are not. In situations of the former sort, which involve problematic preference structures, enforcing a Ulysses contract need not involve arbitrarily respecting the preferences of a past self over the preferences of a current self, or assuming that the agent’s past preferences are more authentic or sound than her current preferences.

⁴ Note that the language of current and past selves is not meant to imply that a human being with significantly changing preferences cannot be unified enough to persist through time as a single person. Throughout this paper, I will proceed without attempting to settle the question of when something counts as a succession of persons inhabiting the same body as opposed to a single but changing embodied person.

⁵ Dresser, Ulysses and the Psychiatrists, *op. cit.* note 3, p. 792.

⁶ For a clear discussion of both positions, see J. Radden. Second Thoughts. *Phil Phenomenol Res* 1994; LIV: 787–801. Radden defends the latter position.

II

There are a variety of tempting activities that people get hooked on, often with very negative results. Among the most familiar are drinking, taking drugs, smoking, snacking, and spending money. Following the norm in the contemporary theoretical literature on addiction, I will not draw a sharp distinction between harmful addictions and bad habits. Relatedly, I will assume that addiction has many faces, and that the different faces of addiction may be captured in a variety of valuable models of self-destructive addictive conduct. I will now review the main models of self-destructive addictive conduct and consider their implications with respect to the usability and morality of Ulysses contracts in the context of addiction.⁷ The first two models considered make it clear that some situations involving addiction do not fit neatly with the use of Ulysses contracts. The third model captures cases in which appealing to previously expressed plans or preferences is not directly relevant to justifying interference. Of particular interest with respect to the broad philosophical question at the heart of this piece are the fourth and fifth models considered, which suggest that in some situations involving addiction, problematic preference structures make the use and enforcement of Ulysses contracts quite appropriate.

The Hidden Costs Model

According to Richard Herrnstein and Drazen Prelec, self-destructive addictive behaviour often stems from a failure to focus on some of the indirect and diffuse costs of the behaviour. One might, for example, gradually go from drinking moderately to drinking frequently and heavily without realizing

⁷ My review of the main models of addiction is adapted from my review in C. Andreou. Going from Bad (or Not So Bad) to Worse: On Harmful Addictions and Habits. *Amer Phil Quart* 2005; 42: 323–331, which surveys some of the theoretical literature on addiction and proposes the *intransitive preferences model*. For a more extensive survey and sample of the theoretical literature on addiction see J. Elster & O.-J. Skog, eds. 1999. *Getting Hooked*. New York, NY: Cambridge University Press, 1999. See also G. Yaffe. Recent Work on Addiction and Responsible Agency. *Phil Pub Affairs* 2002; 30: 178–221. My focus will be on models concerning intrapersonal addictive dynamics. For insightful discussion concerning interpersonal addictive dynamics, see for example K.O. Moene. Addiction and Social Interaction. In *Getting Hooked*, *op. cit.* note 7, pp. 30–46.

that this consumption pattern is gradually changing one's sensibilities and circumstances for the worse. By the time one is drinking heavily, having a drink, though appealing, is no longer the special treat it once was; furthermore, one may now lack the motivation to engage in other previously valued and rewarding activities, such as spending time with one's children. Though it may be clear 'from the outside' that 'addictive behaviour is having profound effects on the rest of the addict's life', this may not be apparent to the addict, or at least may not be a fact the addict is willing to face, particularly if the benefits of his behaviour are immediate while its costs are subtle and spread over time.⁸ So, though it may be responsible for his malaise, the addict may associate his addictive behaviour with relief and blame his troubles on everything but this source of temporary solace.

If one's self-destructive addictive behaviour is due primarily to a failure to focus on some of the indirect and diffuse costs of the behaviour, then one will not contemplate future addictive behaviour with anxiety and so one will not seek to control one's conduct with the help of a Ulysses contract. As such, the availability of external enforcement for self-imposed Ulysses contracts will not solve one's problem.

The Consistent Discounting of Future Utility Model

According to Gary Becker and Kevin Murphy, harmful addictions and habits can flourish even in the absence of hidden costs. So long as one discounts future utility – so long, in other words, as one gives more weight to present satisfaction than to future satisfaction – one can be led by the attraction of an immediate reward (or by the aversiveness of an immediate punishment) to follow a course of action that is less rewarding than another available alternative. Becker and Murphy focus on the case of the addict who has a stable preference to indulge rather than to abstain even though the delayed costs of indulging are high. They see such cases as cases of

⁸ R. Herrnstein and D. Prelec. 1992. A Theory of Addiction. In *Choice Over Time*. G. Loewenstein and J. Elster, eds. New York, NY: Russell Sage Foundation, 337.

rational addiction, since according to standard decision theory, it is permissible to discount future utility, so long as one does so in a way that allows for stable preferences that are consistent over time.⁹

Like the hidden costs model, the consistent discounting model suggests that, even if self-imposed Ulysses contracts were reliably enforced, this would not undermine self-destructive behaviour in all contexts of addiction. If one's self-destructive addictive behaviour is due primarily to consistent discounting of future utility, then though one may contemplate the future costs of one's addictive behaviour with anxiety, one will also contemplate future indulgences with relish and see the gratification associated with indulgence as worth the delayed costs. So, again, one will not seek to control one's conduct with the help of a Ulysses contract.

The Irresistible Cravings Model

As George Loewenstein stresses, harmful addictions are often supported by powerful visceral factors. Paradigmatic visceral factors include drives, such as thirst, and sensations, such as pain. With these factors come urges and cravings that are not directly responsive to one's judgments about what conduct would be most sensible given one's concerns. Such urges and cravings limit 'the scope for volitional control of behaviour' and can be irresistible, 'overwhelm[ing] decision making altogether'.¹⁰ Visceral factors can thus compel an agent to act against his considered preferences – a fact that can help explain why people sometimes behave in self-destructive ways. Consider T. C. Schelling's example of a man who, because of a stomach wound, must refrain from drinking water for 36 hours. Though drinking water prematurely can prove fatal, as the man's thirst increases, 'his mind's capacity to think about the consequences of drinking the water steadily diminish[es], and his mind's preoccupation with the need to quench thirst increase[es]'.¹¹ As Schelling

⁹ See G. Becker and K. Murphy. A Theory of Rational Addiction. *J Polit Econ* 1988; 96: 675–700.

¹⁰ G. Loewenstein. A Visceral Account of Addiction. In *Getting Hooked*, *op. cit.* note 7, pp. 235–236.

¹¹ T.C. Schelling. Self-Command: A New Discipline. In *Choice Over Time*, *op. cit.* note 8, pp. 172–173.

suggests, this situation resembles that of the cocaine addict 'on the verge of relapse in the throes of a craving'.

Unlike the addict whose addictive conduct stems primarily from a failure to recognize hidden costs or from consistent discounting of future utility, the addict who is subject to irresistible cravings that compel him to act against his considered preferences is likely to search desperately for effective measures to help him avoid the tempting but destructive conduct to which he is prone. A Ulysses contract may, in this sort of case, seem quite useful. Notice, however, that in terms of justifying interference, it is the addict's state of compulsion that is crucial here. Given that he is incapable of acting on his considered preferences, the addict in the throes of irresistible craving is not functioning as a free and competent agent. In this sort of case, interference is not supposed to involve a departure – not even a justified departure – from respect for the individual's present free choice. Rather, it is supposed to enable the addict to act in accordance with his preferences, to which his behaviour would naturally conform if his control over it were not compromised by irresistible craving. So if there is any reason to believe that his current preferences differ from those expressed in his prior plan, interference aimed simply at freeing the addict from the grip of irresistible craving will favour his current preferences.

The Preference Reversals Model

According to George Ainslie, at least some addictive behaviour can be traced to temporary preference reversals. Like Becker and Murphy, Ainslie focuses on the fact that we discount future utility. But unlike Becker and Murphy, who assume consistent discounting of future utility, Ainslie suggests that there is growing evidence that our natural discount curves support temporary preference reversals.¹² As such, if an agent does not take anticipatory measures to ensure the coordination of her behaviour through time, but simply acts on her 'natural spontaneous preferences', we can expect that, on a regular basis, she will *voluntarily* do things that she planned

against and will soon regret. She may, for example, spend or consume more than she prospectively and retrospectively endorses.

The ultimate sign of a temporary preference reversal is that one abandons a long-standing resolution when the opportunity for indulgence is imminent but hangs on to resolutions to show restraint in the future. This pattern can emerge if one's natural discount curve is such that the discounted value of a reward spikes as the opportunity to enjoy the reward becomes imminent. Once the pattern becomes familiar, it will be difficult to remain completely naïve concerning the impact of one's unsupplemented plans for future restraint. For an addict anticipating a temporary preference reversal, a Ulysses contract is thus likely to seem very appealing. But is the interference it insists on justifiable? Given that the future self one seeks to control has the status of a free and competent agent, why should the preferences of the self that created the advance directive be respected over the preferences of the self that the directive is meant to control? What sort of justification can those assigned with the task of enforcing the advance directive provide to the self whose preferences are being put aside?

Perhaps, when the time to enforce the advance directive arrives, the enforcer can suggest that although the agent is resisting interference now, she will soon be thankful for the interference.¹³ But, given that there is a significant difference between the (all-things-considered) preferences of the agent's current self and the (all-things-considered) preferences of the agent's future self, the agent's current self can plausibly resist this purported justification by pointing out that it is her future self, who is, by hypothesis, importantly different than her current self, that will be thankful. And so the question remains: What about the preferences of the agent's current self?

Perhaps the enforcer can maintain that the agent's current preferences have been given some

¹² See G. Ainslie. 2001. *Breakdown of Will*. New York, NY: Cambridge University Press.

¹³ Here presumed future (or retroactive) consent is supposed to make up for the lack of current consent. In the least controversial versions of the 'thank you theory of paternalistic intervention', interference is seen as justified only if the agent is temporarily incapacitated so that she is unable to currently provide or withhold consent. (The quoted expression is from A. Stone. 1976. *Mental Health and Law*. Rockville, MD: National Institute of Mental Health, U.S. Department of Health, Education, and Welfare.)

weight, but have been outweighed by the more enduring preferences of her past and future selves. Here we have a problem analogous to mob rule. Just as it is not enough of a justification in interfering with my conduct to say 'I'm just making you do what most people want you to do', it is not enough of a justification in interfering with the conduct of my current self to say 'I'm just making you do what most of your selves (past, present, and future) want you to do'.¹⁴ Now it may be that the agent's future self has a right to be protected against the harms she will suffer if the agent's current self gets her way. Notice, however, that relative to this revised justification, whether the agent formed a Ulysses contract or even just had a prior plan or settled preference concerning her current choice situation is not relevant, or at least not directly relevant. The same, it should be noted, is true of the *you'll-thank-me-later* justification discussed above.

In my view, what the enforcer should emphasize when the time to enforce the advance directive arrives is the following: The agent's plans – in particular, her current plans to show restraint in the future – are bound to fall apart if temporary preference reversals due to inconsistent discounting of future utility are allowed to influence her conduct. Given her commitment to her current plans – including her long-term plans concerning reduced spending or consumption in the future – she should acknowledge the legitimacy of interference based on coping devices such as Ulysses contracts during discounting-induced temporary preference reversals. This justifies interference based on Ulysses contracts from the perspective of the concerns and commitments of the agent's current self. It therefore does not assume complete identification with a past self when, by hypothesis, no such identification is present.¹⁵

¹⁴ This is so, I think, even if the selves are unified enough to collectively constitute a single but changing person. (If it were not true, one would be justified in interfering with an elderly person who decides to spend the last couple of years of her life studying literature though, for many years before this decision, she had settled on spending the last couple of years of her life studying biology.)

¹⁵ Note that while a Ulysses contract made in response to a recognition, on the part of the agent, of a problematic preference structure may not be the only strong sign that an agent has a problematic preference structure, it is arguably the clearest sign and the least paternalistic basis for interference in problematic preference structure cases.

The Intransitive Preferences Model

According to the intransitive preferences model, which I have developed in a recent paper on addiction, some self-destructive addictive behaviour is fostered by stable but intransitive preferences.¹⁶ Roughly speaking, one's preferences with respect to a set of options count as intransitive (at a certain point in time) if one cannot rank the options from most preferred to least preferred because there is a sort of circularity in one's preferences. Here is a simple case of the relevant sort: one prefers option 2 over option 1, option 3 over option 2, but option 1 over option 3. Intransitive preferences often support harmful addictions and habits in cases in which the costs of individual indulgences are negligible but the cumulative effects of many indulgences are great. Consider the smoker who enjoys smoking, but also values decent health, and correctly believes that if he doesn't quit smoking, he will develop serious health problems. Because no individual instance of smoking is significant enough to take him from a state of decent health to a state of poor health, this smoker may, for any number n , invariably prefer quitting after $n + 1$ cigarettes to quitting after n cigarettes. But, because he will, by hypothesis, end up in poor health if he keeps smoking cigarette after cigarette, he may also prefer quitting after a relatively low number of cigarettes, such as 5000, to quitting after a very high number of cigarettes, such as 500,000. Indeed, he may see quitting after a low number of cigarettes as uncomfortable yet tolerable, but see (suffering the damage caused by) smoking a very high number of cigarettes as quite terrible. He would thus have the following intransitive preferences: quitting at 5001 over quitting at 5000, quitting at 5002 over quitting at 5001, . . . , quitting at 499,999 over quitting at 499,998, quitting at 500,000 over quitting at 499,999, but quitting at 5000 over quitting at 500,000. Given these preferences, he is prone to smoke cigarette after cigarette, and thus, via a series of voluntary moves, to end up having followed a course of action that he finds worse than acceptable courses of action that he passed up.

For an addict with intransitive preferences, a Ulysses contract can prove quite useful, since it can force him off what he recognizes as a seductive but

¹⁶ Andreou, *op. cit.* note 7.

self-destructive course. But is the interference required ultimately justifiable? What sort of justification can those assigned with the task of enforcing the advance directive provide to the current self whose pair-wise preference between, say, quitting now and having another cigarette is being put aside? One important point that can be made is that the agent's current pair-wise preference is not being put aside in favour of a conflicting past or future pair-wise preference concerning the same options. For, by hypothesis, the agent's preferences are stable – the intransitive preferences he has today are the same as those he had yesterday and those he will have tomorrow. But, as the agent recognized in drawing up his Ulysses contract, if he invariably follows his pair-wise preferences, he will end up in a position that is worse, relative to his concerns, than his starting position. So those charged with enforcing his Ulysses contrast can justify their interference by pointing out that the agent's advance directive is all he has to protect him against his stable but intransitive preferences, which will, if followed, lead him to a position that is worse, relative to his concerns, than his starting position. The agent should thus acknowledge the legitimacy of interference based on his directive.

III

I have been arguing that in some situations involving addiction, Ulysses contracts are both particularly useful and legitimately enforceable. In the

relevant situations, interference based on Ulysses contracts can be justified in terms of the concerns and commitments of the agent's current self.

Return now to the broad philosophical question at the heart of this piece, namely 'Are there preference structures that warrant interference based on Ulysses contracts?' According to my reasoning, there are at least two types of preference structures that at least sometimes warrant interference based on Ulysses contracts: those that involve intransitivities and those that involve temporary reversals resulting from problematic discounting of future utility. Apart from establishing this result, this paper will, I hope, stimulate research in the area of Ulysses contracts by prompting two follow-up inquiries, namely 'Are there contexts other than addiction in which such preference structures are common?' and 'Are there other preference structures that can warrant interference based on Ulysses contracts?' Attention to these questions will, I think, revitalize research on Ulysses contracts, which seems to have stagnated as soon as a stalemate was reached concerning whether Ulysses contracts are legitimately enforceable in the context of mental illness.

Acknowledgements

My thanks to Peggy Battin, Leslie Francis, Donald Hubin, Elijah Millgram, Ryan Spellecy, Mariam Thalos, Mike White, the audience at my UVSC colloquium presentation, and three anonymous referees for valuable comments on earlier drafts of this paper. Research leave supported by the University of Utah provided me with time to work on the paper; I gratefully acknowledge the support.