INTERNATIONAL JOURNAL OF APPLIED ETHICS
Volume 10 (2024), pp. 21-37

RESEARCH ARTICLE

Self-paternalism: The Concept and Key Normative Problems

Morten Ebbe Juul Nielsen

Ph.D, Associate Professor, Section of Philosophy/Comm, Faculty of Humanities, University of Copenhagen

Submitted: January 22, 2024; returned: September 1, 2024; revised: September 17, 2024; accepted: December 10, 2024

Abstract: Paternalism is normally defined as an act where some agent, e.g., the state, imposes something on *another* agent, e.g., a citizen. That is, one party is acting paternalistically towards another party. In this paper, I argue that *self-paternalism* is not an oxymoron: one may act paternalistically towards oneself. Self-paternalism typically involves a diachronic aspect, pitting the choice or will of an agent *a* against the choice or will of *a* at some *later* time. I argue that the prior act or decision can be felt as an alien imposition for *a* at some later time even if it originates from him- or herself. We cannot always say that "the true will" or "the settled preferences" of the agent is identified by the decision to bind oneself an earlier time. Furthermore, doing away with the worry about paternalism by subsuming *a*'s original choice or decision under the harm principle seems to have some very disagreeable moral implications. Furthermore, it is argued that while the "self" in self-paternalism indeed gives us some reason to find such acts morally permissible, this may be outweighed by other, non-trivial concerns: Self-paternalist acts may express an undue mistrust in one's own willpower or moral agency. They can be judgmental in a way that conflicts with proper self-respect. Moreover, the restraint or "penalty" attached to certain acts of self-paternalism may be too burdensome. Finally, if an act of self-paternalism contributes to the atrophy of our "moral muscles", such an act may be morally dubious. I conclude that there is room for both self-paternalism as a concept, and for wrong acts of self-paternalism.

Keywords: Self-Paternalism, Paternalism; Autonomy; Self-Determination; Self-Respect

Introduction

It is only natural if some believes that "self-paternalism" is an oxymoron. Surely, paternalism

involves a relation between *different* agents – an agent acting paternalistically towards another agent. However, consider the following: A person is about to eat and pours herself a drink, with the intention only to have *one* drink, so she can resume work after eating. However, worrying that one drink will lead to another, she locks all liquor in a cupboard, puts the key in an envelope addressed to herself, and drops the envelope in a mailbox. She gets home, finishes her "self-allowed" drink, and soon regrets having put the liquor under lock and key (see Andreou 2018, p.62.)

However, on a widespread understanding of paternalism it involves some party *A*, imposing (nudging, manipulating, coercing…) *X vis-à-vis* some *other* party, *B*, against the will of *B*, for the (intended) benefit of *B*. The above case may be construed to fail to be *paternalism* on two counts: 1) the imposing party cannot be identical to the party imposed upon, and 2) it is actually the will of the agent in the case not to succumb to temptation, hence, it is not against the "true will" of *B*.

I shall first argue that this view of self-paternalism is not right. Self-paternalism is not an oxymoron; it is a species of paternalism.

Some may agree that self-paternalist acts *are* paternalist but that, *qua self*-paternalist, they are always morally permissible. After all, it is the agent herself that does the imposing. I do not accept this view: I will argue that while self-paternalism may be *prima facie* and often also all things considered morally permissible, this is not always the case. There is room for both self-paternalism as a concept, and for *wrong* acts of self-paternalism. The latter is addressed in the second half of the paper.

Sustained discussions of self-paternalism, especially of recent origin, do not abound. In fact, one of the only ones is a six-page entry in the Routledge Handbook of the Philosophy of Paternalism (Andreou 2018).

Given this meagre interest, why should we engage in serious thought about self-paternalism? I believe there are weighty practical grounds for exploring the issue. For instance, developments in technologies such as "wearables" (e.g., watches that monitor certain bio-measures), surveillance/tracking, and IT generally, greatly expands the means and opportunities for *self-disciplining* through self-paternalist acts. Combine this observation with what Rose (2006) calls a general turn towards a "somatic ethics" (a societal focus on health and the bodily as the prime scenes for self-expression and evaluation.) Insofar one may be concerned with what Crawford dubbed "Healthism",[1] there is good reason to look closer at self-paternalism in this regard. In fact, modern technologies make it possible to monitor

[1] Crawford defined healthism as "…the preoccupation with personal health as a primary—often *the* primary—focus for the definition and achievement of well-being; a goal which is to be attained primarily through the modification of life styles.", Crawford 1980, p.368.

oneself in myriad new ways; ways that have not yet been subject to much moral scrutiny.[2]

1 Paternalism

It is generally agreed that paternalism involves some form of altering the choice situation for a given agent with the purpose of protecting or promoting the welfare or interests of the agent. For instance, so-called "sin taxes" on alcohol or tobacco may have a paternalistic rationale: by altering (increasing) the price for alcohol or tobacco, the choice to buy these products becomes less attractive, and this is meant to lessen consumption and protect the health of the public.

However, there is little consensus on the exact definition of the term. Quong offers an excellent exposition of the weaknesses of various understandings of the term (see Quong 2010, section 3.1) The key points in Qoung's criticism of various influential definitions of paternalism is that they are either over- or under-inclusive given widely shared intuitions of what paternalism is. Consider, for example, the classical account provided by Gerald Dworkin: "the interference with a person's liberty of action justified by reasons referring exclusively to the welfare, good, happiness, needs, interests or values of the person being coerced" (Dworkin 1997, p.61.) This fails to catch acts that do not restrict person's liberty of action but expands their options, but which intuitively still count as paternalist, such as incentivizing a course of action because the paternalizer mistrusts the paternalizee's willpower (Quong 2010, p.75).

Quong's own definition is close to Shiffrin's much discussed motive-based account (see Shiffrin 2000, pp. 211ff; Quong 2010, pp. 80-83). This so-called *judgemental definition* is as follows:

> 1. Agent A attempts to improve the welfare, good, happiness, needs, interests, or values of agent B with regard to a particular decision or situation that B faces.
>
> 2. A's act is motivated by a *negative judgement* about B's ability (assuming B has the relevant information) to make the right decision or manage the particular situation in a way that will effectively advance B's welfare, good, happiness, needs, interests, or values.

The first clause reflects the standard view that paternalist acts must be aimed at improving the welfare or something related to the welfare of the paternalizee. The second that paternalist acts reflects that the paternalist acts out of a belief in his or her superior insight or agency *vis-à-vis* some aspect of the judgement, will, or rationality of the paternalizee. Quong claims that the

[2] To get an overview of the opportunities for health related self-monitoring, see Ruckenstein and Schull 2017. Note that other fields that are not directly related to health promotion, such as banning oneself from gambling sites, excluding oneself from internet access for appointed hours, etc. are also made possible with new technologies.

judgmental definition is superior to the alternatives because it catches the core of what we mean when we use the phrase "paternalism".[3]

I will here take this definition as an outset. In any event, I believe that the arguments offered here are robust across a range of plausible definitions of paternalism, at least *mutatis mutandi.*[4] Some short clarificatory remarks.

First, the relevant (lack of) ability on B's behalf pertains to "practical reasoning, willpower, and emotion management." (Quong 2010 p.81). The paternalizer can feel justified in interfering because he or she feels superior in any of the three areas. This is particularly relevant in the present context because, arguably, many key examples of self-paternalism will involve the agent's doubts about his or her *own* (future) *willpower* in certain situations. The gambler doubts whether the willpower to withstand the temptation to gamble is available at the right time, for instance.

Second, I think that this definition of paternalism expresses what may be morally troubling specifically about *paternalism* that distinguishes it from other morally troubling acts. For instance, restrictions on persons' liberty may often be morally wrong, but need not to have anything to do with paternalism. Furthermore, one may sometimes interfere in what is otherwise other persons' legitimate areas of control to protect *other* parties, and again, that would not be termed paternalism under normal circumstances. What seems to distinguish paternalist acts is precisely the element that Quong expresses when he says that paternalism involves treating another as, in a certain sense, a child. In the case of self-paternalism, it involves treating oneself, in a certain sense, as a child.

Third, the argument here is *not* that if a given act may plausibly be dubbed "self-paternalist", it is therefore morally wrong. Serious debate about paternalism concerns the limits of paternalism, not whether it is never, or always, justified.[5]

---

[3] According to Quong, his definition "…captures our sense that to treat someone paternalistically is to treat that person like a child in the specific sense of acting in that person's best interests because you believe, in this situation, the person lacks the ability to do so himself or herself." (Quong 2010, p. 81)

[4] Quong's definition (and by extension Shiffrin's) has been the subject of some criticism, e.g., Mills 2013, Birks 2014. Birks argues that the definition may fail to capture some particularly problematic forms of strong paternalism, and that it may not be able to distinguish between less and more troubling forms of paternalism, and especially the latter is a point of inquiry it would be interesting to pursue here. However, for reasons of space, I cannot do so. In any event, I believe Quong's (and Shiffrin's) judgmental or status-based account(s) capture something particularly interesting about self-paternalism: the mistrust in one's own future ability to perform adequate self-control.

[5] Regularly we accept acts that on most analyses would be dubbed "paternalistic" (see Dworkin 2020; Kleinig 1983, p.39).

On the other hand, we should not accept that all things considered promotion of the good is *always* a sufficient and overriding reason that make paternalistic acts morally right. Shiffrin makes the following methodological assumption:

> It is difficult to know how to adjudicate between competing characterizations of paternalism. One plausible approach…involves not merely testing formulas against intuitions, but also testing formulas with an eye to arriving at a conception of paternalism that fits and makes sense of our conviction that paternalism matters. (Shiffrin 2000, p.212)

For instance, it may be the case that for the dyed-in-the-wool act consequentialist, acts of paternalism carries nothing but instrumental significance. However, I guess that for even quite strongly pro-paternalists, paternalist acts *do* have a special normative status – at least when they themselves are the objects of paternalist acts! In any event, in this paper I agree with Shiffrin's observation, and hence also that "paternalism matters", at least sometimes.

2 Is self-paternalism paternalism at all?

One may query whether self-paternalism is a misnomer: if a necessary condition for an act to be paternalist is that the act goes against the will of the subject imposed upon, and self-paternalism is as act willed by the very subject imposed upon, how could self-paternalism be paternalism? (see, e.g., Grill 2012, p.361, but compare Andreou 2018, pp. 59f).

However, self-paternalism either necessarily[6] or at least typically involves a diachronic aspect: Agent *a* wills at some time to interfere with *a´*s will, or choice set, at some later time, if situations of type X arise. E.g., *a,* who sincerely wants to limit smoking, sets up financial limits to how much *a* can spend on tobacco at a later time, even if *a* at a later time really wants to spend more on tobacco than is allowed by the limit. Essentially, many self-paternalist acts relate to handling *temptation*. One anticipates that one will be tempted to do something that one wants to avoid, and so one engages in self-paternalism.

As spelled out by Andreou (for the following, see Andreou 2018, pp.59ff), apart from temptation, self-paternalism also involves admitting the possibility that one may, at later time, become the victim of *akrasia*. Akrasia is the term for situations where one acts voluntarily against one's better judgment. If one anticipates this, one may decide to alter future choice sets, for example by putting the liquor under lock and key, as in the example in the beginning of this paper.

Another typical, but importantly different component in self-paternalism is the prospect of

[6] Whether conflicting, synchronic desires may make it reasonable to speak of synchronic acts of self-paternalism – say, when an agent both want and don't want to smoke a cigarette - is an issue I bypass here

procrastination. Andreou defines this as "…cases where there is some prospect of procrastination in relation to an important goal that can only be achieved via a series of individually trivial contributions each of which requires some sacrifice." (Andreou 2018, p.63) For example, if one has the goal to achieve better physical health, this implies that one must be involved in a continued effort to exercise. Each session of exercise may be a trivial contribution, meaning that any *single* session may be rather inconsequential. What matters is that one engages continuously in a training regime. But since each individual exercise session can be dropped, one may be tempted to postpone exercising – to procrastinate. Anticipating this, one may self-paternalistically set up incentives or disincentives to lessen the risk of procrastination.

I need here to make a slight detour. How do we distinguish self-*paternalist* acts from mere promises to one-self? It would seem weird or over-reaching if all promises to oneself about future behaviour were acts of self-paternalism. I propose that self-paternalist acts include either a) some sort of manifest interference in order to produce the desired result (as in giving one's car keys to a friend before a wet night out) and/or b) some sort of self-imposed *penalty* attached to non-compliance with the promise (e.g., "if I don't run at least 4x5 km this week, I can't go to the concert Saturday"). The penalty may be draconic but may also be as mild as vague self-criticism and a fleeting feeling of guilt. This may imply that some acts that we would normally dub "promises to oneself" blends into self-paternalism *post factum*. For instance, if one's non-compliance to some promise to oneself is accompanied by feelings of guilt, it may look like self-paternalism. Yet, some promises to oneself are *not* accompanied by feelings of guilt when broken, e.g., when reasons shift sufficiently to justify non-compliance. Moreover, many promises to oneself does not concerns one's own future welfare, but may, for example, concern the welfare of others. Hence, not all promises to oneself are acts of self-paternalism.

In a brief discussion, Kleinig looks at whether we should refrain from calling self-paternalism paternalism at all (see Kleinig 1983, p.56.) It may be claimed that when a person consents to future impositions, then (normally?) we can say that the consented to imposition reflects one's *settled* wants as against one's *episodic* wants. However, we can dismiss this because, even if prior consent has been obtained, *when* the imposition against the consenting party kicks in, it is still an *imposition*.

The argument from settled vs. episodic wants has a certain plausible ring to it: if I am quite sure what my settled wants are, and that they may be jeopardized by some episodic or fleeting want, I may indeed have good reasons to "bind myself" in various ways. By accepting an at-the-source deduction of my salary that goes directly to my pension, I do not have to worry that I will be tempted to burn off too much of my salary to set aside enough money for when I retire

(for this and the following, see Kleinig 1983, pp.56f).[7]

However, the distinction between settled and episodic wants is not watertight. First of all, what is at some earlier time is a settled preference may at a later time be replaced by another: Olga, having agreed to undergo a special weight-loss programme which, if she does not lose the required weight, entails a financial penalty, finds herself at a later time more comfortable with her weight and wants to opt out of the programme. Definitely, her act at an earlier time would be paternalistic if imposed when she wants to opt out. Secondly, it may be harder to distinguish between settled and episodic preferences than the argument supposes, and one may even have conflicting settled preferences. A person may have the following preferences at some given time: a) to limit consumption of alcohol, and b) to be a spontaneous person. What if b) at some point later in time "wins" over a), and that person wants to drink more than would be the case if a) had "won out"? Can we always be sure that a) reflects a more genuinely settled preference than b) without invoking rather dubious forms of moralism? If not, then treating that person at a later time in accordance with what only hold partially as that person's preferences at an earlier time would indeed seem paternalistic (or dubiously moralistic).

Davis argues that when we have two conflicting preferences, then we should respect the "resolution preference" – that is, a third preference – the agent has about how to resolve the conflict. Presumably, for many self-paternalism cases, this tilts the right course of action in favour of the "original" preference/choice. When *a* decides to limit her gambling at some time, it goes hand in hand with the formation of a resolution preference along the lines of the following: "I prefer to resolve eventual preference conflicts in favour of my preference here and now, and not in favour of any conflicting preference in the future". That is, the formation of this third preference works as a sort of tiebreaker (see Davis 2004).

This seems inadequate: The agent can be assumed to form *another* resolution preference along with his or her new preference at some later time favouring the satisfaction of the new preference, unless the conflicting preference (the one arising later) is so weak or fleeting that it is easily "vetoed" by the first (and in *this* case, surely stronger) preference. This does not imply

[7] Probably, the *locus classicus* of this discussion is found in Berlin's influential essay "Two Concepts of Liberty" (Berlin 1969.) Here, Berlin distinguishes between our "higher self", equating it with (at least an important part of) our rationality, or our sense of self-mastery, and our "lower self" – our "nature", as it were, filled with passions and unreflected desires. Berlin equated to some degree our higher selves with positive liberty. Even if Berlin warned us against various ways in which emphasising positive liberty, he surely also thought that self-mastery and rationality is important: without our higher selves, we would not be rational human beings. The higher self may reflect our "settled wants" (and the lower self then reflects our "episodic wants") An interesting question here, then, is whether we can act self-paternalistically out of a higher-order desire to act in accordance with our higher self. I believe this is a very plausible thought. However, what reason do we have we assume that we are always *right* in our introspective deductions from our "higher selves"? For instance, can't what we believe to be "messages from our higher selves" not be expressions of neurosis, that, if used as a background for self-paternalist acts, can indeed turn out to be felt as alien impositions on our empirical, lower self?

that we must embrace the "Current Preference-thesis" – which implies we should always choose to honour or support the current preferences of an agent. But it does imply that "counting" preferences in the way suggested by Davis does not necessarily help us decide which preference to favour in our evaluation.[8]

I noted that we should use a conception of paternalism that rings true to our intuitions about why and when paternalism may be morally wrong. I think a common emotion arising when an agent is being treated paternalistically is as follows: The act is felt to be *an imposition by an outside force*; as something alien to one's wants, preferences, or sense of agency. (It is not an emotion that arises necessarily because there may be paternalist acts that work "behind the scenes" and are never perceived by the agent acted upon. But in such cases, the same moral emotion would typically arise if the agent was informed about the act, or so I would guess.) For self-paternalism, this raises the question whether we can feel alienated *vis-à-vis our own* (prior) preferences, dispositions, decisions etc. I believe this to be obviously the case.

To see why, consider the phenomenon of *regret*. During a normal course of life, we will come to regret earlier decisions and commitments. One may regret entering a marriage, a career plan, a commitment to a political cause, and so on. To take the perhaps most obvious case: even if one enters a marriage with the clear ambition to honour the "Till Death Do Us part"-commitment, it seems not unreasonable that one may regret this earlier commitment. Note also that even if one regrets the prior commitment, a change of mind may also come with harsh self-criticism. Changes in one's beliefs or values may obviously render one's earlier commitments a burden. Submission to this earlier commitment may, I venture, make that commitment feel like an imposition by an outside force, *even if* the commitment made came from oneself. In short, we may feel *alienated* by our prior commitments. (Arguably, "imposition by an outside force" and "alienation" are different concepts, and I do not by the latter refer to the classic discussion of alienation as used in critical theory (see, e.g., Forst 2017). I use "alien" and "alienation" in the colloquial sense only.)

If it can be admitted that 1) we may (reasonably, at least sometimes) come to regret earlier commitments, and that 2) if we for various reasons[9] cannot easily disentangle ourselves from the commitment, then I believe that it is rather obvious that we can feel earlier self-paternalistic plans and commitments as outside impositions. We can, in that sense, end up being alienated

---

[8] Davis discusses cases where competence diminishes over time. A key case here is that of advance directives. Here, a conundrum arises (see Radden 1994.) This concerns changes of heart that may, or may not be, the result of dwindling competence. I believe Davis is mostly on the right track when discussing cases where the relevant competences is undermined due to, e.g., dementia. In such cases, I believe there should be a strong presumption in favour of the preference expressed at t1, but compare Andreou 2018, p. 61. Davis acknowledges that changes in competencies are necessary for his argument to gain traction, see Davis 2004, p. 275.

[9] There may be many different sources of this. For instance, if you are a person that feel strongly about your own earlier commitments, it may seem as a particularly heavy defeat if you even consider changing paths.

towards our own earlier plans and wows, but still feel that abandoning them comes with a high price.

A further point about denying that self-paternalism could ever be *bona fide* paternalism: If we accept as a principle or as a solid rule of thumb that prior consent always justifies later impositions in ways rendering it incoherent to talk of *paternalism* (due to the imposition being self-imposed), I have the hunch that the argument overreach. A person at some earlier time has a strong desire to live an ascetic lifestyle, refraining from most worldly goods. She swears a vow never to eat meat. 5 years later, she changes her mind. Suppose eating meat actually is bad for her; hence, an imposition of her earlier vow not to eat meat would indeed be for her good. Given her change of mood *vis-à-vis* eating eat is of a reasonable stable kind, I fail to see why an imposition upholding her prior vow in such a case should not be termed paternalistic.

Kleinig discusses a specific defence of hard paternalism revolving around the notion of "future selves." Since self-paternalism typically involves a diachronic aspect, this discussion is eminently relevant for the present purposes. Recall Parfit's (Parfit 1987) view about identity: Albert at some later point is not entirely identical to Albert at some earlier time, and at some point (Parfit argues) it makes sense to speak of another Albert altogether. On this view, identity is a matter of degree; at some point, we move beyond any meaningful identity between a person at earlier and later times. But if Albert in 2025 is not Albert at in 2010, then a given harm that Albert in 2010 inflicts at Albert in 2025 (say, by Albert in 2010 failing to take reasonably good care of his health so Albert in 2025 suffers unnecessarily), then paternalist-like measures imposed on Albert in 2010 are not necessarily paternalist at all. They may be subsumed under the harm principle. In essence, the harm principle allows, under certain conditions, restrictions against harm to *others*. So, if Albert in 2010 isn't really Albert in 2025, then we may justifiably impose restrictions on Albert in 2010 and some time after that, in order to protect *another* person, that is, Albert in 2025. But this would not be paternalism, precisely because Albert in 2010 isn't really Albert in 2025.

Kleinig notes some worrying flaws with the argument (Kleinig 1983, pp.46ff). If part of its justification lies in the fact that some persons who experience harm because of earlier decisions change their minds ("I really should have worn that crash helmet") then it does not apply in cases where the harm is fatal. But naturally, the proponent would like cases with a deadly outcome to be covered by the argument as well. The argument seems only to have force in cases where a person *actually* experiences harm. If a policy of mandatory crash helmets is meant to be justified by the argument, and only, say, 2% (which is probably an overestimation) of motorcycle riders experience a dangerous crash, how do we justify imposing the helmet policy on the remaining 98%? Furthermore, Kleinig is sceptical about Parfit's view of identity, especially when applied to the case of paternalism: "What is perhaps most puzzling about the Argument from Future Selves is why we should be attracted to the view of personal identity that underlies it…It suffers from a certain ad hocery, especially in view of the counter intuitiveness of some of its extensions." (Kleinig 1983, p. 46.) Kleinig is perhaps a bit brusque here. But there is another problem. Unless one adopts some version of strict consequentialism, there is something deeply morally troubling about the view we would have to have about ourselves if we should take the argument from future selves in the presented version seriously.

We would, in a sense, have to make ourselves slaves of our future selves. This raises the question: which self, exactly? The one that *may* be harmed in ten years, or the one that, at the deathbed *may* regret that the earlier version of him- or herself never took any risks, or? We should, according to this line of argument, internalise a subservient attitude to a master who may indeed be inordinately fickle, or very harsh. An alien imposition indeed! Surely, this is a perverse moral view and cannot be an attitude that a reasonable morality forces us to internalize.

In essence, what has been argued here is that self-paternalism is not an oxymoron. Self-paternalism typically involves a diachronic aspect, pitting the choice or will of an agent *a* at some earlier time against the choice or will of *a* at some earlier time. The prior choice, or limitation, or self-imposed penalty, can be felt as an alien imposition for *a* at some later time even if it originates from *a*. We cannot always say that "the true will" or "the settled preferences" of the agent is identified by the decision to bind oneself at some earlier time, even if there may be a weak presumption in favour of such earlier choices. Furthermore, doing away with the worry about paternalism by subsuming *a*'s choice at an earlier time under the harm principle seems to have some very disagreeable moral implications.

3 Normative concerns over self-paternalism

As noted by Radden (Radden 1994, pp. 787f), there is a strong current in modern moral thinking linking self-determination (or autonomy) to our capacity for "binding ourselves" to future paths or choices. This includes, of course, to binding ourselves to a promise *not* to pursue a specific path. Radden mentions notabilities in the debate such as Feinberg and Arneson.

Radden concurs that the ability to bind oneself is one important aspect of self-determination, but only as one of two aspects: "A different expression of autonomy which is equally central, I would argue, is found in our ability to change our minds in the light of new ideas, beliefs, and desires." (Radden 1994, p.788). Hence, self-paternalism involves two autonomy-related issues: Self-paternalist acts is or can be an important expression of our autonomous will, but so can changing our mind. Re-doing our earlier commitment can also be an autonomous act. This reflects the basic conflict of intuitions that cases of self-paternalism may elicit. On the one hand, a wish to respect the choices, including choices to "bind oneself", of persons. On the other hand, a similar wish to respect our opportunity to revise our beliefs and wants. It does not suffice to lump these two together. If our actual beliefs or preferences should always take precedence, then respect for "binding choices" becomes undermined – and *vice versa* (but keep the earlier discussion of "resolution-preferences" in mind here!)

A central way in which self-paternalism may be morally worrying concerns self-respect. Following Quong's definition of paternalism, paternalist acts involve a negative judgment about the object of paternalism's ability to reach the right decision. Paternalism is a way of treating a person as a child. In the case of self-paternalism, the "child" is oneself. Should we

respect *a*'s verdict at some earlier time that future *a*'s decision at a later time may be so childlike that (self-)paternalist measures are justified? Or should we rather respect *a*'s decision at some later time to "defect" from the earlier decision to bind him- or herself? A way of approaching this question is to look closer at the concept of self-respect, for it seems natural to think that, if there is something normatively troubling about self-paternalism, and the wrongness of paternalism lies in a form of lack of respect for persons, then self-paternalist acts (may) express a worrying lack of concern for self-respect.

Following Dillon 1992, self-respect is standardly divided into recognition self-respect and evaluative self-respect (see Dillon 1992; Stark 2019). The first concerns the respect we owe to ourselves simply in virtue of being a person, when viewed as an autonomous chooser.[10] "Recognition respect for oneself as a rational autonomous agent involves acknowledging and appreciating one's autonomy, one's rationality, the value one has in virtue of having these capacities, and the moral status of equality that is grounded in this shared value…"[11] Evaluative (or "appraisal") self-respect concerns one's self-respect (or lack thereof) *vis-à-vis* specific and concrete achievements, such as conforming to certain ideals of worthiness. "One merits evaluative self-respect, in other words, not in virtue of *being* a person, but because of the virtue *of* one's person."[12]

According to Stark, those that argue that persons ought to care about their recognition self-respect imply, among other things, that persons ought to "abstain from self-destructive behaviour" as well as to "avoid subordinating themselves to others unnecessarily" and to "take responsibility for their conduct."[13] This reflects the duality noted in the above about self-determination. Self-paternalist acts may reflect a conscientious effort to avoid self-destructive behaviour. But it may also be seen as an unnecessary or exaggerated act of subordination to one's earlier plans. In the same vein, such acts can be viewed as prudent ways of exercising responsibility, or as neurotic expressions of self-deprecation. The implication here is that if one thinks self-respect *is* something we should care about (think, for instance, of the high value Rawls accords to the social bases of self-respect (see Rawls 2001, p.59)). We ought to be concerned about when self-paternalism constitutes an attack on proper recognition self-respect. For it seems implausible that *all* acts of self-paternalism are by that very fact prudent and appropriate acts of self-preservation or exercises of due responsibility.

Evaluative self-respect is also relevant to consider, and the duality we have just seen plays out once more. A given self-paternalistic act may be viewed as conforming to some plausible

[10] Stark (Stark 2019) argues that this notion is Kantian insofar as it relies on a moralized or at least highly specific notion of autonomy. I believe the notion makes sense also on a much leaner concept such as being an autarchic agent, where the only necessary condition for agency is that it makes sense to speak of a person as a choosing agent. See Gaus 2011, pp.346ff.

[11] Stark 2019, np.

[12] Stark 2019, np.

[13] Stark 2019, np.

standard of worthiness. A recovering alcoholic that sets up reasonable barriers for consumption may by this add to his or her sense of evaluative self-respect. On the other hand, a person who sets up unnecessary or unduly restrictive barriers may *lessen* his or her grounds for increased self-respect when achieving otherwise worthy goals. If one could have achieved some noteworthy goal without setting up self-paternalist measures to get there, it seems, at least intuitively, to detract from the value of achieving the goal to have set up these measures.

In the present context, it is worth noting that self-respect seems often to be linked to self-discipline; the successfully recovering addict, the obese person achieving and maintaining weight-loss, the former unfit person completing a long run, and similar figures are taken to show or gain self-respect. This may be an expression of what Nikolas Rose calls "the somatic ethics" that permeates our times. But in any event, it seems noteworthy that these figures are celebrated due to what may be (and probably often is) self-paternalistic efforts, and that it is assumed that their self-respect at least partially hinges on the success of their projects.

Setting aside the obvious possible beneficial effects (and also the sometimes very admirable and respect-inducing efforts) at play, are there grounds for normative concern? Recall the discussion in the above about "slavery to our future selves". Do we, by way of self-paternalist acts, sometimes unjustifiably treat ourselves as children? Do we at times *judge* ourselves too harshly – in ways that cast doubts about the legitimacy about our self-chosen bondage?

It seems to me that this may be the case. An adult with no history of alcohol abuse could begin taking Antabuse just to make sure alcohol addiction won't arise, but it could very well be an expression of a distorted and denigrating view of the person herself. Probably cases like this are rare. Less rare, I would surmise, are cases where overly anxious (or ambitious) persons sets up a battery of self-paternalist measures such as using "wearables" to monitor all bio-health data, limiting their screentime by apps like the probably not intentionally ironically titled "Freedom", counting calories for every intake and so on. At some point, a fine masked net of self-surveillance must be an expression of such a low estimation of one's own capabilities as an autarchic agent that it becomes judgmental, and in *that* sense a worrying form of self-denigrating self-paternalism. One may of course freely adopt an attitude of "trust, but verify" *vis-à-vis* one's own (future) willpower – but even this must include a modicum of *trust* in oneself.

Of course, even an extremely fine-masked net of self-surveillance and self-paternalism, upping the costs of several possible actions, or making some of them altogether impossible, may be all things considered justifiable. Perhaps the self-paternalizing agent in question really *would* succumb to dangerous temptations again and again *in lieu* of self-surveillance etc.

However, surely there can be cases where the mistrust of one's future willpower is misplaced or exaggerated, and hence be an expression of lack of true self-respect. Answering the obvious question here, namely "when", is quite complicated. A full answer would involve robust answers to a set of other difficult moral questions. To mention two central ones: 1) do we have duties to ourselves; if yes, what duties; and what are their weight as compared to other duties and commitments? What interests is it those duties should protect? For instance, should we

include an interest in allowing for spontaneity? 2) Should we reserve our moral judgments to *ex post*-evaluations, or can we make such judgments *ex ante*? *Ex ante*, it is at least somewhat speculative to say about a person that his or her self-paternalist commitments are a) exaggerated and an expression of misplaced mistrust in the self-paternalists willpower, or b) just about right to achieve the wanted goals and a justifiable balance between prudence, autonomy, care and so on, or c) not enough to reach the goals anyway. *Ex post*, we may be in a better position to assess the efficacy and dosage of the self-paternalist acts, but even here, questions about causality may blur the precision of our assessment (e.g., would a person *not* have succumbed to dangerous temptation even if no self-paternalist measures had been in place.)

However, even if we lack precision in our assessment, we should not ignore the possibility that self-paternalists measures could be morally troubling because they express a denigrating form of mistrust of one's willpower (and concomitant aspects of agency). Our lack of precision should rather be taken into consideration when we form judgments about self-paternalist acts. As second-parties, we should take serious an agent's commitments to self-paternalist acts and reserve criticism to the more obviously neurotic or misguided cases. On the other hand, we should also temper any tendency to over-zealous *support* for self-paternalist acts.

If one is concerned about paternalism, one should of course also be concerned with judgments about (and, *a fortiori*, interferences with) persons' plans about how to conduct their own lives. This creates a certain tension when dealing with self-paternalist acts that one may suspect display an unpalatable air of lack of (self-)respect. Do you show proper respect by respecting the initial choice (the "binding ourselves"-aspect of autonomy?) Or by interfering (e.g., by support, or advice, or even by convincing oneself in the case of introspective deliberation over one's own self-paternalist commitments) with the initial commitment (the "changing our minds"-aspect of autonomy)? Admitting that this tension may exist is not equivalent to taking sides on the question (*pace* Davis). It calls for a case-by-case assessment. But it seems to me obvious that one may show morally troubling disrespect to oneself in self-paternalist acts: at some earlier time, if the self-paternalist act shows an unrealistically low assessment of one's future willpower (as in the non-drinking person that begins on a regime of Antabuse), or at a later time, if one sticks to an earlier commitment that now is an alien imposition (as in Kleinig's example with Olga, who changes her perception of her weight and wants to opt out of a diet regime.) If one shares the view that such situations may morally troubling components of lack of self-respect, then one should admit that self-paternalism may be morally troubling, on grounds of lack of proper self-respect.

There are further normative concerns. Self-paternalist acts can be morally problematic when the *means* (restrictions or penalties) are too severe. This is no different from paternalist acts that may be unjustifiable because they impose too severe restrictions or punishments on paternalizees. We may as philosophy teachers mildly coerce students to engage in logic courses, but we may not lock them up in cells until they master second order logic, nor may we threaten to beat them if they fail to pass. We may in some – perhaps many – cases accept that someone sets up for him- or herself a restrictive or penalizing system that is more onerous than we would condone if the system was imposed from the outside. Still, there are limits to how

stringent a system or severe a punishment that may be justified merely by being self-imposed.[14] In more general terms, we can say that self-paternalist acts where the expected benefit does not outweigh the imposed costs are generally morally unpalatable – just as similar standard paternalist acts would be. (Whether or not such acts are "truly" paternalist or not, since they in a certain sense misfire, is a terminological struggle of no real import, I believe.)

A further point here: while individual acts of self-paternalism may only inflict a modest penalty, I believe that there is a case for being morally worried when the *sum* of such acts collectively incur a harsh penalty. This is relevant to the point earlier made about the greatly expanded opportunities for self-paternalism.

A third concern has deep roots in the history of moral philosophy and concerns the threat to long-term moral agency that may lurk behind some forms (or perhaps "levels") of self-paternalism. Kleinig says: "Mill speaks of individuality analogously to bodily fitness. Regular exercise is needed if the muscles – mental and moral as well as physical – are not to atrophy…" (Kleinig 1983, p.261). In essence, the worry (in the present context) is that adopting *too heavy* (too many and/or too restrictive) self-paternalist measures may lead to a weakening of our moral fibre and hence our capacity for choosing for ourselves.

It is probably easiest to illustrate the problem by way of standard paternalism before going into self-paternalism. In the mid 2010s, Danish politicians discussed what forms of regulation should be applied to so-called "mobile loans" – small loans (with often punitive annual costs) easily obtainable by your mobile device or laptop. One spokesperson for a political party expressed the following sentiment: since Danes are so used to everything being well-regulated, many of them would think that mobile loans are well regulated too. Since they are not, they need to be regulated. Now, regulation can of course be motivated by non-paternalist concerns, but regulating things like mobile loans can easily be construed as paternalistic. That may not be a problem in itself (indeed, I believe there are good reasons to regulate this area), but that is not the point. It is rather that if you are surrounded by paternalist measures and *rely* on such measures at every corner, it seems not unlikely that at some point, one's "moral muscles" begins to dwindle. This is worrying for two reasons. One worry is consequentialist: it may lead to bad outcomes if we do not with some regularity exercise our practical reason rather than relying automatically on the idea that someone has already set up a system where our choices cannot have bad consequences. This may be harmful to ourselves and to others. The other has a more deontological ring to it: being cocooned in a net of paternalist protections displays a certain amount of mistrust or similar negative attitude *vis-à-vis* the moral and cognitive abilities of the paternalizee.

---

[14] The matter is complex. There seems to be not only two different setups or normative questions here (the other- vs the self-imposed system). A third question concerns whether or when second parties should assist (or not assist) self-paternalist acts. Sometimes, it seems intuitively right to assist (e.g., helping a friend who wants to quit smoking with searching for lost or hidden cigarettes in her home in order to remove the temptation), whereas in other cases, it may be right to allow the self-paternalists to go about his or her project but wrong to actively assist the endeavor, e.g., by inflicting physical pain on the self-paternalist when a wow is broken.

Moving from paternalism to self-paternalism, it seems at least conceivable that one can enmesh oneself in a net of self-paternalist measures that ultimately conjure up the same worry: that the self-paternalist's moral muscles begin to atrophy, which may lead to sub-optimal outcomes, and displays a certain form of unpalatable mistrust of one's own moral and cognitive abilities. Imagine a person who monitors a whole range of bio-data, limits screen time, counts calories and so on. Even if all this is self-imposed, there seems to be a point beyond which even perfectly sensible acts of self-paternalism, simply because of their number and how weighty/intrusive they are, *as an aggregate* becomes morally problematic. There is line – admittedly a blurry one – between due diligence and neuroticism. "Trust, but verify" may be a sound policy for someone who has good reasons to mistrust his or her own future willpower, but beyond a certain point, the mistrust is misplaced and unduly self-deprecating.[15]

There is one further reflection worth mentioning here, even if it for reasons of space cannot be pursued to any great length. I believe the issue of self-paternalism connects interestingly with (at least a broad reading of) the notion of the *biopolitical* as known from Foucault and perhaps especially in the version brought forward by Rose, as well as the issue of *healthism*, as introduced by Crawford (see Crawford 1980). At least this is so when a) the sorts of self-paternalism in question concerns (physical and/or mental) *health* (which probably often is the case), and b) that one allows for using the concept of the biopolitical on the individual level.[16] Rose's *The Politics of Life Itself* shows us how amendable to deliberate changes the biological –"life itself" – has become, and how deeply politicized and moralized practices of medicine and "self-care" has become. But it is not only governments that can have an interest in (and act upon) our health: a very efficient way of achieving population health is by installing a "somatic ethics" in the citizens themselves, through means of self-surveillance and self-discipline. Self-paternalism may in that regard be said to be a central form of biopolitical (self-) governance. Modern technology gives us unheard potential for self-surveillance, from wearables to apps monitoring and ultimately regulating our sleep, meditation, screentime, calories-intake, gambling proclivities and so on. This opens new and maybe dramatic avenues for self-paternalism. Even if I am only reluctantly optimistic about the fecundity of the "biopolitical lens" as an analytical approach, at least this literature reminds us to see self-paternalism in the broader context of political and societal values and norms combined with technological innovations.

Conclusion

I have argued that we have reasons to believe that self-paternalism is not an oxymoron. A self-imposed restraint originating at an earlier time may seem like an alien, outside act or imposition to an agent at a later time. This is so even if the agent chose to adopt the restraint

[15] From Once upon a Time in the West: "How can you trust a man who wears both a belt and suspenders? The man can't even trust his own pants!"

[16] Foucault's notion of the biopolitical concerns most often populations, but populations are after all made up by individuals.

under circumstances at the earlier time that would otherwise point in the direction that the decision to adopt it was a perfectly reasonable thing to do. Unless one follows a very strict form of consequentialism, paternalist acts stand in need of justification that go beyond appeals to maximization of welfare. Some such acts are defensible. The special subset of paternalist acts – self-paternalist acts – are, arguably, likely to be justifiable rather often. The opportunity to "bind one's will" by self-paternalist acts itself is valuable. Not arriving by your own car to a party where you know you may be tempted to drink, and you know you may then be tempted to drive under the influence, is a good practice. The opportunity to "ban" oneself from gambling sites is a valuable opportunity. That, however, does not mean that self-paternalistic acts are by definition justified, or morally innocuous. Nor does it mean that other parties must enthusiastically assist person's acts of self-paternalism, directly, or indirectly, e.g., by enabling self-paternalism via setting up social mechanisms that offer themselves to self-paternalism. One class of self-paternalist acts may be deemed morally dubious because the restraint or "penalty" attached is too burdensome. While we may say that Odysseus was justified in binding himself to the mast in order not to be mesmerized by the Siren's song, it does not seem that making himself permanently deaf would be so. Even if we – reasonably – accept that morally competent agents may put heavier burdens on their own shoulders that what is acceptable when they put it on the shoulders of *other* agents, it does not follow that *any* self-imposed burden is morally unproblematic. Two possibly overlapping cases are interesting. Centrally, self-paternalist acts may express an undue mistrust in one's own willpower or moral agency. They can be unduly judgmental, in ways compromising reasonable concerns about proper (self-)respect. And if or to the extent an act of self-paternalism contributes to the atrophy of our "moral muscles", such an act may be morally dubious. I hope this paper can contribute to further moral inquiry into the concept and practice of self-paternalism.

Statements and Declarations:

Conflict of Interests: No conflicts of interests

External funding: No external funding

References

[1] Andreou, C. (2018) "Self-paternalism", pp. 59-65 in Grill, K. and Hanna, J., *The Routledge Handbook of the Philosophy of Paternalism.*

[2] Berlin, I. (1969) "Two Concepts of Liberty", in *Four Essays On Liberty*, (Oxford, England: Oxford University Press, 1969), pp. 118-172.

[3] Birks, D. (2014) "Moral Status and the Wrongness of Paternalism"., *Soc Theory Pract*. Jul 1;40(3):483-498. doi: 10.5840/soctheorpract201440329.

[4] Crawford, R. (1980) "Healthism and the medicalization of everyday life." Int J Health Serv. 1980;10(3):365-88. doi: 10.2190/3H2H-3XJN-3KAY-G9NY. PMID: 7419309.

[5] Davis, J.K. (2004) "Precedent Autonomy and Subsequent Consent", *Ethical Theory and Moral Practice*, pp. 267-291.

[6] Dillon, R. (1992) "How to Lose Your Self-Respect", *American Philosophical Quarterly* 29: 125–139.

[7] Dworkin, G. (1997) "Paternalism" in *Mill's On Liberty: Critical Essays*, Gerald Dworkin, ed. (Oxford: Rowman & Littlefield, 1997), reprinted from *The Monist* 56 (1972), pp. 64-68.

[8] Dworkin, G. (2020) "Paternalism", *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition), https://plato.stanford.edu/archives/fall2020/entries/paternalism/.

[9] Forst, R. (2017). "Noumenal Alienation: Rousseau, Kant and Marx on the Dialectics of Self-Determination." Kantian Review 22 (4):523-551.

[10] Kleinig, J. (1983) *Paternalism*, Manchester University Press.

[11] Gaus, G. (2011) *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World*. Cambridge University Press

[12] Grill, K. (2012) "Paternalism", pp. 359-369 in Chadwick, R. (ed.) *Encyclopedia of Applied Ethics,* Elsevier.

[13] Mills, C. (2013) "The Problem of Paternal Motives", *Utilitas* 25, no. 4 (2013), pp.446-462.

[14] Quong, J. (2010) *Liberalism Without Perfection*. Oxford University Press.

[15] Parfit, D. (1987) *Reasons and Persons*, Clarendon Press.

[16] Radden, J. (1994) "Second Thoughts: Revoking Decisions Over One's Own Future" *Philosophy and Phenomenological Research*, pp. 787-801.

[17] Rawls, J. (2001) *Justice as Fairness – A Restatement*, The Belknap Press of Harvard University Press (Cambridge: Mass.)

[18] Rose, N. (2006) *The Politics of Life Itself: Biomedicine, Power, and Subjectivity in the Twenty-First Century*, Princeton University Press.

[19] Ruckenstein, M. and Schüll, N.D. (2017) "The Datafication of Health", *Annual Review of Anthropology*. 46:1, 261-278.

[20] Shiffrin, S.V. (2000) "Paternalism, Unconscionability Doctrine, and Accommodation", *Philosophy & Public Affairs* Vol. 29, No. 3 pp. 205-250.

[21] Stark, C.A (2019) "Self-Respect", *The Routledge Encyclopedia of Philosophy*, updated 2019. DOI:10.4324/9780415249126-L092-1. (Original version in the paper edition, 1998.)