Anti-paternalism and Public Health Policy

KALLE GRILL

STOCKHOLM 2009

ISSN 1650-8831
ISBN 978-91-7415-226-5

Printed by US-AB, Stockholm, Sweden, 2009.

ABSTRACT

Grill, K., 2009. Anti-paternalism and Public Health Policy. *Theses in Philosophy from the Royal Institute of Technology* 31. viii + 165 pp. Stockholm. ISBN 978-91-7415-226-5.

This thesis is an attempt to constructively interpret and critically evaluate the liberal doctrine that we may not limit a person's liberty for her own good, and to discuss its implications and alternatives in some concrete areas of public health policy. The thesis starts theoretical and goes ever more practical. The first paper is devoted to positive interpretation of anti-paternalism with special focus on the reason component – personal good. A novel generic definition of paternalism is proposed, intended to capture, in a generous fashion, the object of traditional liberal resistance to paternalism – the invocation of personal good reasons for limiting of or interfering with a person's liberty. In the second paper, the normative aspect of this resistance is given a somewhat technical interpretation in terms of invalidation of reasons – the blocking of reasons from influencing the moral status of actions according to their strength. It is then argued that normative anti-paternalism so understood is unreasonable, on three grounds: 1) Since the doctrine only applies to sufficiently voluntary action, voluntariness determines validity of reasons, which is unwarranted and leads to wrong answers to moral questions. 2) Since voluntariness comes in degrees, a threshold must be set where personal good reasons are invalidated, leading to peculiar jumps in the justifiability of actions. 3) Anti-paternalism imposes an untenable and unhelpful distinction between the value of respecting choices that are sufficiently voluntary and choices that are not. The third paper adds to this critique the fourth argument that none of the action types typically proposed to specify the action component of paternalism is such that performing an action of that type out of benevolence is essentially morally problematic. The fourth paper ignores the critique in the second and third papers and proposes, in an anti-paternalistic spirit, a series of rules for the justification of option-restricting policies aimed at groups where some members consent to the policy and some do not. Such policies present the liberal with a dilemma where the value of not restricting people's options without their consent conflicts with the value of allowing people to shape their lives according to their own wishes. The fifth paper applies the understanding of anti-paternalism developed in the earlier papers to product safety regulation, as an example of a public health policy area. The sixth paper explores in more detail a specific public health policy, namely that of mandatory alcohol interlocks in all cars, proposed by the former Swedish government and supported by the Swedish National Road Administration. The policy is evaluated for cost-effectiveness, for possible diffusion of individual responsibility, and for paternalistic treatment of drivers. The seventh paper argues for a liberal policy in the area of dissemination of information about uncertain threats to public health. The argument against paternalism is based on common sense consequentialist considerations, avoiding any appeal to the normative anti-paternalism rejected earlier in the thesis.

Keywords: Alcohol Interlocks; Altruism; Anti-paternalism; Epistemic paternalism; Group consent; Harm principle; Interference; Invalidation of reasons; Liberalism; Limiting liberty; Private sphere; Product safety regulation; Public health policy; Reason-actions; Self-regarding; Social responsibility; Uncertain information; Withholding of information.

LIST OF PAPERS

This doctoral thesis consists of an introduction and seven papers:

1. Grill, K. 2007. 'The Normative Core of Paternalism'. *Res Publica* 13(4): 441-458.
2. Grill, K. 'Anti-paternalism and Invalidation of Reasons'. Submitted manuscript.
3. Grill, K. 'Paternalistic Interference'. Submitted manuscript.
4. Grill, K. 2009. 'Liberalism, Altruism and Group Consent'. Forthcoming in *Public Health Ethics* 2(2).
5. Grill, K. 2009. 'Anti-paternalism and Public Health Policy: The Case of Product Safety Regulation'. Forthcoming in Dawson, A (ed.) *The Philosophy of Public Health*. Ashgate.
6. Grill, K. & Fahlquist, J.N. 'Responsibility, Paternalism and Alcohol Interlocks'. Responsibility, Paternalism and Alcohol Interlocks. Forthcoming in Dawson, A., Donckers, H. & Maes, L. (eds.) *Ethics of Health Promotion*. Springer.
7. Grill, K. & Hansson, S.O. 2005. 'Epistemic paternalism in public health'. *Journal of Medical Ethics* 31(11): 648-653.

NOTE: In this digital version, paper 4 has been updated to closer resemble the published version, though there are still a few very minor differences.

Front cover picture: Odysseus and the Sirens. Detail from an Attic red-figured stamnos (pottery), ca. 480-470 BC. From Vulci.

Author: Kalle Grill, Department of Philosophy and History of Technology, Royal Institute of Technology, Stockholm, Sweden. Email: kgrill@kth.se.

PREFACE

Almost six years ago, I set out to understand what paternalism was all about and why some people were so fiercely against it. I considered myself a liberal of sorts, and still do, but it seemed to me that the value of liberty was invoked too readily and with too much confidence, in this area as in many others. After all, there seemed to be some agreement that paternalism involved on the one hand a limiting of liberty, but on the other hand a promotion of good. Preserving liberty and promoting good seemed to me both important. Why then should paternalism always be (*prima facie*) morally wrong?

This early impression has stayed with me. I still consider myself a liberal and I am still uncertain what I mean by that. However, I am ever more convinced that for all the greatness of liberty, it is not the moral trump card it is too often made out to be. Individual health and well-being are very important, and can sometimes be secured at the expense of liberty. As tyrannical as it may seem to restrain a person for her own good, as cruel can it be to stand by and let someone perish from her own mistakes or confusion. It may yet be that some liberal self-realization or Millian individuality is the highest form of life, though a life of community and shared joys and sorrows is at least a close second. However, it cannot be that mere everyday restriction, coercion, or intrusion should never be suffered for preserved health or survival.

I sometimes wish with the liberal tradition that there were rational true selves with unshakable preferences inside all of us, who could be asked to direct our lives when we need steadying. However, for better and worse we are just the fallible human beings that we are, prone to bias and misjudgement and heavily influenced by our surroundings. Such are the selves that should be at liberty. Equally fallible, of course, are governments and other authorities. Nowhere can we turn for enlightened direction, neither to ourselves nor to some external director. We must simply make do with what little ability we have, helping ourselves and each other enjoy the most liberty and the most well-being that we can attain, making difficult choices in the process.

ACKNOWLEDGEMENTS

Writing this thesis has mostly been lonesome work, as academic work often is. That said I am very grateful for the many opportunities I have had to discuss my ideas with friends and colleagues at the Division of Philosophy, at quite a few international conferences, and during my visit to the University of California in San Diego (UCSD). I am no less grateful for the love and support of my close friends and family over the years, especially my mother Lisa Grill for her constant interest in and encouragement of my sometimes esoteric work, my friend and former girlfriend Hanna Ögren for her support and her refreshing 'naive philosophy', and my present girlfriend Camilla Claesson for her embrace of my various philosophy projects, as well as for her analytic mind and open heart.

Two of my close friends are also my best critics – my fellow graduate students Lars Lindblom and Niklas Möller. When it comes to the intricacies of liberal political philosophy as well as my particular takes on it, Lars has been my most dependable and

thorough critic. Niklas has consistently provided excellent comments on argument and presentation more generally (he has also vigorously tried to slow down my thesis work by involving me in courses, readings and discussions on meta-ethics and philosophy of mind and language). My main supervisor Sven Ove Hansson has provided useful critique on terminology and argument. Without Sven Ove this thesis would not have been written, as he authored the original research plan and secured the funds. My assistant supervisor Martin Peterson has provided reliable and speedily feed-back on presentation and sometimes argument, ever urging me to get clearer. This has been good advice more often than I realized at the time. Other colleagues at the Division of Philosophy have provided useful critique on parts of the thesis, especially Sara Belfrage and Dan Munter.

During my six months at UCSD, Richard Arneson was my host and very generous discussion partner. Our common probings into the subtleties of anti-paternalism were crucial in giving me a sense that I was on the right track, or at least not on an unreasonable one. The UCSD philosophy department in general was very welcoming and I am thankful.

Financial support from the Swedish Council for Working Life and Social Research (FAS) and, for the UCSD visit, from the Swedish Foundation for International Cooperation in Research and Higher Education (STINT), is very gratefully acknowledged.

Stockholm Kalle Grill
February 2009

TABLE OF CONTENTS

Introduction

> [I]t is possible, and at times justifiable, to coerce men in the name of some goal (let us say, justice or public health) which they would, if they were more enlightened, themselves pursue, but do not, because they are blind or ignorant or corrupt.
> –Isaiah Berlin[1]

1. Overview

This is a thesis on political morality and practical ethics. To paraphrase Joseph Raz (1986, p. 4), it is a thesis on ethics, which concentrates on certain moral issues because of their political implications. The moral issues are the nature of paternalism and anti-paternalism, and the reasonableness of the latter. The political implications are quite general but are in the thesis explicitly exemplified in the area of public health policy, with special attention to the tendency of such policy to limit the liberty of groups of people. Product safety regulation, mandatory alcohol interlocks, and the dissemination of information about uncertain threats to public health are investigated in some detail. The most novel contribution is perhaps the attention paid to reasons and their role in anti-paternalist doctrines, both as formulated in theory and as applied in practice.

Throughout, the focus is more on normative issues than on conceptual analysis. Paternalism is (in the first paper) defined generically, to allow more detailed conceptions to capture variations on the opposition to benevolent limiting of liberty that is part of the liberal tradition. This opposition is then analyzed and criticised (in paper two and three). Setting aside commonsense consequentialist or empirical arguments, the purely normative principle that benevolence can never justify limiting liberty is found unreasonable on several grounds. Nonetheless, an attempt is made (in paper four) to contribute to the anti-paternalist cause by suggesting some rules by which the liberal may possibly justify limiting the liberty of groups of individuals without invoking their good, but only their (partial) consent. Both the more constructive interpretation of anti-paternalism and the more negative rejection of this doctrine in the more theoretical first part of the thesis inform the discussion of concrete policy areas in the more practical later part (papers five, six and seven).

In this introduction, I will first briefly describe the history of anti-paternalism as I see it, with a focus on John Stuart Mill and Joel Feinberg as the main proponents of this doctrine. After so setting the stage, I will move on to consider seven definitions of paternalism and discuss their strengths and weaknesses. This will further demonstrate the context and background against which this thesis has developed and will naturally lead me to explain my own view of how paternalism should be understood. In the fourth chapter I discuss what it means to act for the good of someone else and argue that the inherent morally problematic aspect of paternalism is not overriding a person's own

[1] In 'Two Concepts of Liberty' (2002 [1969], p. 179).

conception of her good, but rather limiting her liberty. In connection, the distinction between paternalism and moralism is defended. Not until the fifth chapter do I give a summary of the arguments against anti-paternalism that are developed in the first three papers. I place these arguments in the wider context of practical reasoning and in particular in the context of what I call influence-regulating principles – principles that block reasons from influencing the moral status of actions according to their strength. In this chapter I also explain how liberalism can do without paternalism and how my argument against anti-paternalism relates to more general theories of (political) morality. In chapter six I try to say something more positive about the value of health and of liberty and consider so called libertarian paternalism as one way to justify public health policy. The very short seventh chapter contains a brief survey of the six articles that follow upon this introduction.

2. A BRIEF HISTORY OF ANTI-PATERNALISM

Much of contemporary debate on paternalism takes as its starting point John Stuart Mill's *On Liberty* (1991 [1859]). Mill does not himself use the term paternalism, but famously formulates this liberty principle:

> That principle is, that the sole end for which mankind are warranted, individually or collectively, in interfering with the liberty of action of any of their number, is self-protection. That the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant. He cannot rightfully be compelled to do or forebear because it will be better for him to do so, because it will make him happier, because, in the opinion of others, to do so would be wise, or even right. (p. 14)

This moral principle has a positive and a negative part. The positive part is the *harm principle*, saying that preventing harm to others is a valid purpose for limiting liberty (or interfering with liberty – I make no distinction).[2] The negative part is that no other purpose is valid. This part can be further divided into resistance to particular purposes – to promote physical good, to promote moral good, to promote doing the right thing (being moral). In his debate with Patrick Devlin over the relationship between law and morality and the proposal to decriminalize private homosexuality, H.L.H. Hart (1963) made a point of distinguishing *legal moralism* – 'to enforce positive morality' – from paternalism – 'to protect individuals against themselves' (p. 31).

Hart's characterization of paternalism is suggestive but imprecise. Neither Mill nor Hart were especially concerned with conceptual analysis. Mill was above all concerned to specify the principle he advocated, Hart the principle he rejected. Hart

[2] Following Feinberg (1986, p. ix). Sometimes 'the harm principle' refers to what I (following e.g. Arneson 1989, p. 409) just called 'the liberty principle'.

rejected moralism, but accepted paternalism. His common sense arguments for this acceptance are instructive:

> Choices may be made or consent given without adequate reflection or appreciation of the consequences; or in pursuit of merely transitory desires; or in various predicaments when the judgment is likely to be clouded; or under psychological compulsion; or under pressure by others of a kind too subtle to be susceptible of proof in a court of law. (p. 33)

As a result, people make choices that harm them, and when they do we should sometimes stop them.

Our understanding of the many ways in which we are products of our environments have only deepened since Hart's times. It is therefore not surprising that Mill's most distinguished follower in the 20th century made voluntariness a central concept in his resistance to paternalism. Joel Feinberg (1971; 1986) restricted his liberalism to the subject of criminal law and defined *legal paternalism* so:

> It is always a good and relevant (though not necessarily decisive) reason in support of a criminal prohibition that it will prevent harm (physical, psychological, or economic) to the actor himself.' (1986, p. 4)

Feinberg rejected this principle in favour of what he would have preferred to name *soft anti-paternalism* but, yielding to convention, called *soft paternalism* – the principle that the state may limit a person's liberty for her own good when and only when her conduct is not voluntary enough, or if intervention is needed to establish how voluntary it is (1986, p. 12).[3] *Hard paternalism*, in contrast, allows limiting a person's liberty even when her conduct is fully voluntary. Feinberg spends a large part of his article and later book discussing how voluntary is voluntary enough. The criteria include basic competence (not an infant or insane or comatose); absence of manipulation, coercion and duress; informedness; and absence of distorting circumstances (fatigue, agitation, passion, drugs, pain, neurosis, time pressure) (p. 115).

When we fall under the threshold of sufficiently voluntary, we are not in tune with what Feinberg sometimes calls our 'true self', and so restraining our non-voluntary actions is not really limiting *our* liberty. While the harm principle protects a person from others, soft paternalism protects him from his "'non-voluntary choices", which, being the genuine choices of no one at all, are no less foreign to him.' (1986, p. 12) When we stay over the threshold, on the other hand, there are no good reasons to intervene since no wrong is done: What harm we voluntarily bring on ourselves is not wrongful, in line with the Roman Law maxim of *volenti non fit injuria* – 'To one who freely consents to a thing no wrong is done, no matter how harmful to him the consequences may be.' (1971, p. 107)

[3] Feinberg thereby rejects not only legal paternalism as specified, but also the stronger version where 'sometimes' is substituted for 'always', given that the actor acts voluntarily.

Soon after Feinberg's first rejection of hard legal paternalism in 1971, the debate on paternalism turned more conceptual, as will be explored in the following section. On the more normative side, most of the attention was on soft paternalism (e.g. Hodson 1977; Van de Veer 1986) and on further restricting anti-paternalism by making exceptions for the promotion of liberty or autonomy (e.g. Dworkin 1972; Kleinig 1983; Sneddon 2001; Husak 2003; De Marneffe 2006). Gerald Dworkin (1972) argues that liberty-preserving hard paternalism should be accepted and that Mill paved the way for such acceptance with his rejection of voluntary slavery. Others followed Dworkin in proposing other accounts of liberal values that similarly justify hard paternalism.

The most noteworthy recent development on the normative side of the debate is arguably the embrace of paternalism by Richard Thaler and Cass Sunstein (2003a; 2003b) under the provocative label *libertarian paternalism*. Libertarian paternalism is paternalism in the sense that 'it attempts to influence the choices of affected parties in a way that will make choosers better off.' (2003b, p. 1162) It is libertarian in the sense that 'people should be free to opt out of specified arrangements if they choose to do so.' (p. 1161) Thaler & Sunstein justify benevolent influence on choice by appeal to a sort of expanded soft paternalism. To briefly return to Mill, Hart found his picture of the human being naive:

> [A] middle-aged man whose desires are relatively fixed, not liable to be artificially stimulated by external influences; who knows what he wants and what gives him satisfaction or happiness; and who pursues these things when he can.

We might in turn find Hart's picture of the middle-aged man naive (and sexist), but the point is that Mill assumed that, with a few exceptions he did not spend much ink discussing, people act on their set preferences. Feinberg recognised that we often do not so act, but apparently assumed that we all have a middle-aged man inside of us, our true self, with equally set preferences. The point that Thaler & Sunstein has driven home is that because of the great impact of default rules, starting points and framing effects, most of the time we simply have no preferences independently of the choice situations we face. Therefore, adjusting those choice situations to promote our good does not limit our liberty.

Much has been written on paternalism and anti-paternalism that is not captured in this brief history of course. I will turn to some more conceptual points in the following section. Parallel to the more principled arguments surveyed, there has always been straightforward consequentialist arguments for and against paternalism, mainly of two kinds: 1) People, and especially people in the government, cannot be trusted to limit other people's liberty for their own good, because people are selfish, power corrupts, etc. 2) Everything we do and are affects others one way or other and so there is always a non-paternalistic rationale for limiting liberty. Both of these arguments are valid, but they do not address what I take to be the central normative dimension of paternalism – whether or not the promotion (not just as an attempt or an excuse but actually) of a person's good can contribute to justifications for limiting her liberty.

3. DEFINITIONS

As noted, I argue in the first paper for defining paternalism on normative grounds. More specifically, detailed conceptions of paternalism should be based on substantial normative views on what reasons are morally problematic when invoked for what actions, or on attempts to describe such substantial normative views. In this chapter, I will defend this approach in general terms. I will also explain more in detail what I think this normative approach entails concerning the defining and understanding of paternalism. First, however, I will survey seven definitions from the literature on paternalism over the last decades and discuss their differences, their strengths and weaknesses. This is not a comprehensive account (more thorough overviews, though less up to date, can be found in Van de Veer 1986 and Nikku 1997). I merely aim to contrast the approach to anti-paternalism defended and employed in this thesis to other influential approaches.

3.1 PATERNALISM 1972-2008

In his argument for liberty-preserving paternalism, Gerald Dworkin (1972) adopted this brief characterization:

> By paternalism I shall understand roughly 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. (p. 65)

The characterization comes very close to the anti-paternalist part of Mill's liberty principle. Paternalism is basically *benevolent interference*. Later in the article, it becomes clear that Dworkin finds paternalism morally problematic only when it is not consented to (p. 77).

Dworkin is mainly concerned with the normative question of whether benevolent limiting of liberty is ever justified. He supposedly neither intends a full conceptual definition nor a precise description of a normative position, but merely to set the stage for his argument (as indicated by the 'roughly'). The debate on paternalism took a more conceptual turn with Bernard Gert and Charles Culver (1976) objecting to defining paternalism in terms of interference with liberty of action. Gert & Culver point out that paternalism need not involve an attempt to control behaviour, but can be a case of 'killing, causing pain, disabling, depriving of pleasure, deception, or breaking a promise.' (p. 51) These things cannot, they claim, be understood as interferences with liberty of action. Dworkin (1983) later concedes this point but defends his earlier definition as appropriate given his interest in 'the proper limits of state coercion' (p. 105). Gert and Culver explicitly do not restrict themselves to legal paternalism. They offer this general definition:

> A is acting paternalistically toward S if and only if A's behaviour (correctly) indicates that A believes that
>
> (1) his action is for S's good

(2) he is qualified to act on S's behalf
(3) his action involves violating a moral rule (or doing that which will require him to do so) with regard to S
(4) he is justified in acting on S's good independently of S's past, present, or immediately forthcoming (free, informed) consent
(5) S believes (perhaps falsely) that he (S) generally knows what is for his own good. (pp. 49-50)

What is most noteworthy about this definition is that it is put entirely in the head of the paternalist. No interference with liberty is required, but neither is an actual violation of a moral rule. As noted by Donald Van de Veer (1986, p. 37), I may, according to this definition, believe that I am exorcising your demon through some obscure but harmless ritual and be acting paternalistically, even though there is no benefit to you nor any other effect on you. The focus is very much on the inner life and character of the agent, and not at all on the action and its consequences. This is particularly inapt for discussing paternalism in politics, where effects are, arguably, more important than motives.[4]

Apart from the heavy focus on beliefs, Gert & Culver's definition is traditionally Millian in its focus on benevolence (condition 1) and lack of consent (condition 4). The interference condition (3) is formulated in terms of violating a moral rule, which, if it were not for the belief qualification, would ensure that paternalism raises moral issues. Condition 5 contains the traditional exclusion of incompetents (infants, animals), a very weak voluntariness condition. The purpose of condition 2 seems to be to exclude, for some reason, cases where the agent is obviously less competent than the person she acts towards, such as when the agent is 'a small child' (pp. 50-51).

Van de Veer (1986) has another issue with Dworkin's rough characterization. The term 'justified' implies, he argues, that Dworkin's 'definition' is normative in that paternalism is supposed to be justified, and that Dworkin fails to consider the motives of the agent. An action that is justified by personal good reasons may be "'paternalistic in result'", Van de Veer admits, but the action as such is not paternalistic (p. 27). Van de Veer sometimes (p. 26) takes Gert & Culver, too, to assume that paternalism is justified (by condition 4) and sometimes not (because of the belief qualification) (p. 37). In contrast to their (overly) rich definition, Van de Veer's (1986) own is minimalist in its core content, though excessive in its detail:

A's doing or omitting some act X to, or toward, S is paternalistic behavior if and only if
1. A deliberately does (or omits) X
and 2. A believes that his (her) doing (or omitting) X is contrary to S's operative preference, intention or disposition at the time A does (or omits) X [or when X affects S–or would have affected S if X had been done (or omitted)]

[4] In a later article, Gert and Culver (1979) repeat their definition but treats condition 3 as if it was not phrased in terms of beliefs, but in terms of actual violation (pp. 199-200 and esp. note 4).

and 3. A does (or omits) X with the primary or sole aim of promoting a benefit for S [a benefit which, A believes, would not accrue to S in the absence of A's doing (or omitting) X] or preventing a harm to S [a harm which, A believes, would accrue to S in the absence of A's doing (or omitting) X]. (p. 22)

Interestingly, Van de Veer's definition is as subjectivist as Gert & Culver's. On this definition, too, it may be paternalism if I believe that I am exorcising your demon through some obscure but harmless ritual.

Van de Veer takes it to be an advantage that on his definition, paternalism need not raise any moral issues. Very reasonably, however, Van de Veer is mainly interested in specifically those cases of paternalism that do raise moral issues. As he notes, 'many paternalistic practices happen to involve acts which are presumptively wrong' (p. 21). We might then infer a Van de Veerian definition of morally interesting paternalism by adding this interference condition:

and 4. A's doing (or omitting) X presumptively wrongs S.[5]

This condition would also offer a connection with actual effects, independently of beliefs (it could of course also come in a belief variation).

Van de Veer's condition 1 only requires that the doing is an intentional action, something that is taken for granted in other definitions. Condition 2 is foremost a consent condition, though without the added condition 4 it also serves the role of a weak interference condition. While in Gert & Culver's definition the agent believes she acts 'independently of' consent, in Van de Veer's she believes she acts 'contrary to' consent. The first formulation includes acting where there is no (belief that there is a) preference either way. This makes sense, as there is an important distinction to be made between expressed consent and lack of expressed consent. Important because an interference that is explicitly consented to is arguably not morally problematic, it does not limit liberty in a morally relevant sense.

Van de Veer's condition 3 is the benevolence condition, which in his rendering explicitly includes both harm-prevention and good-promotion. It is, like Gert & Culver's, subjective to the agent. No actual benefit is required. While Gert & Culver requires only that the agent believes the action will be beneficial, Van de Veer settles for the stronger requirement that the action must be primary or solely intended to benefit the person acted towards. However, this difference is softened by Gert & Culver explaining that 'insofar as A's behaviour toward S is paternalistic, it is only S's good, not the good of some third party, which is involved.' (p. 50)

Like Van de Veer, David Archard (1990) tries to improve on Dworkin's and Gert & Culver's accounts (without position himself in relation to Van de Veer). Like Van de Veer's, his definition is explicitly designed to be non-normative or morally neutral.

[5] Van de Veer would not appreciate this addendum. In fact, he laments his own earlier 'fixation on the *morally interesting* subset of cases of paternalism, interesting because controversial' (p. 35).

However, in contrast to Van de Veer, Archard means to incorporate the morally interesting aspect of paternalism in his definition:

> P behaves paternalistically towards Q iff:
>
> (1) P aims to bring it about that with respect to some state(s) of affairs which concerns Q's good Q's choice or opportunity to choose is denied or diminished;
> (2) P's belief that this behaviour promotes Q's good is the main reason for P's behaviour;
> (3) P discounts Q's belief that P's behaviour does not promote Q's good.

Part of the background to Archard's interference condition (1) is that Dworkin (1983), going conceptual, argues that Gert & Culver's definition is too narrow, since paternalism need not involve the violation of a moral rule. It is paternalism, Dworkin argues, for a husband to hide his sleeping pills from his suicidal wife (p. 106). Archard's definition is meant to accommodate this case. Though the husband is not obligated to share his pills, his wife's opportunity to choose is diminished because he would have done so '[i]n the normal course of events' (p. 37). On the other hand, *increasing* a person's choices cannot, Archard argues, be paternalistic. The definition is morally neutral because we have duties to promote other's well-being, but also, supposedly, duties to respect their choice, and the definition does not say which is stronger (pp. 41-42).

Archard's benefit condition (2) is all but identical to Van de Veer's (assuming a sole reason is a main reason). Condition 3 can be understood as a rich consent condition. Archard finds it essential to paternalism that it involves 'the usurpation of one person's choice of their own good by another person.' (p. 36) Therefore, it is not paternalism if P's belief about Q's belief is incorrect, as there is then no actual usurpation (p. 39). Further, Archard prefers an 'independently of' understanding of consent, and accepts that the consent may be given previously but not that it be merely anticipated (p. 40).

Seana Shiffrin (2000), like Archard, aims to provide 'a conception of paternalism that fits and makes sense of our conviction that paternalism matters' (p. 212):

> [P]aternalism by A toward B may be characterized as behavior (whether through action or through omission)
>
> (a) aimed to have (or to avoid) an effect on B or her sphere of legitimate agency
> (b) that involves the substitution of A's judgment or agency for B's
> (c) directed at B's own interests or matters that legitimately lie within B's control
> (d) undertaken on the grounds that compared to B's judgment or agency with respect to those interests or other matters, A regards her judgment or agency to be (or as likely to be), in some respect, superior to B's. (p. 218)

What is most noteworthy with this definition (characterization) is that there is no benevolence condition. Shiffrin argues that it may be paternalism to interfere with a person for the sake of a third party, or without concern for anyone's welfare, as long as the interference condition is satisfied (pp. 215-7). Condition d is supposed to specify a motive; Shiffrin seems to hold that the sheer belief that one can do something better (in some unspecified sense) than another can move one to act, giving the example of articulating a point at a talk. Indeed, this motive is the central normative component of paternalism on Shiffrin's account, since it 'delivers a special sort of insult to competent, autonomous agents.' (p. 220). The three remaining conditions all concern the specification of interference (conditions a and c are all but identical). However, with its focus on legitimate control, this specification may be taken to incorporate a consent condition. Shiffrin earlier in the article explicitly prefers a 'independently of' to a 'contrary to' understanding of consent (and supposedly of legitimate control) (p. 214).

In contrast to Archard, Shiffrin thinks that increasing a person's range of options (her choices) can be paternalistic, if the person prefers not to have those extra choices. This is so because in such cases, where A increases B's options against her will, 'A forcibly substitutes her judgment about the right way for B to exercise and develop her agency.' (p. 214) In other cases, the substitution is not so much of judgment as of agency – A and B may agree on what should be done but B may think that she can make it happen more effectively, even though it falls under A's 'sphere of legitimate agency'. It is this focus on substitution of judgment or agency that leads Shiffrin to reject the benevolence condition. She argues that 'we should have the same sort of normative reaction' regardless of who, if anyone, benefits from the interference.

Peter De Marneffe (2006) objects to Shiffrin's anti-paternalism and argues that substitution of judgment is common and not necessarily insulting. The motive for imposing speeding limits may be the good of drivers or the good of third parties – in either case the government substitutes its judgment for that of the driver (pp. 77-79). We are imperfectly rational and so may make mistakes both in our considerations of the interest of others and of our own interests (p. 80). Paternalism would be morally problematic if paternalistic reasons were silenced somehow, or if there were moral rights against paternalism, but De Marneffe finds no good reason to accept any of these ideas (at least not on a Scanlonian account of rights) (pp. 83-86).

De Marneffe observes that policies may be paternalistic in their effects without a paternalistic motive and is therefore drawn to an account in terms of justification. However, he finds no such account that allows that unjustified policies can be paternalistic. Therefore, and perhaps to accommodate linguistic intuitions, he proposes this hybrid definition:

> [A] government policy is paternalistic toward A if and only if (a) it limits A's choices by deterring A from choosing to perform an action or by making it more difficult for A to perform it; (b) A prefers A's own situation when A's choices are not limited in this way; (c) the government has this policy only because those in the relevant political process believe or once believed that this policy will benefit A in

> some way; and (d) this policy cannot be fully justified without counting its benefits to A in its favor. (pp. 73-74)

The definition is rather traditional, with a the interference condition in terms of choice, reminiscent of Archard's; b a consent condition (where if A has no explicit preference we consider what A would prefer if she thought about it [p. 73, note 15]); c a motivational benevolence condition; and d a justificationary benevolence condition. What is novel is the hybrid between a motivational and a justificationary definition.

In a definition of paternalism for *The Stanford Encyclopedia of Philosophy*, first formulated in 2002, Dworkin (2008) drops the 'of action' bit from his early characterization of interference with liberty and adds interference with autonomy as an alternative, proposing this very Millian definition:

> I suggest the following conditions as an analysis of X *acts paternalistically towards* Y *by doing (omitting)* Z:
>
> 1. Z (or its omission) interferes with the liberty or autonomy of Y.
> 2. X does so without the consent of Y
> 3. X does so just because Z will improve the welfare of Y (where this includes preventing his welfare from diminishing), or in some way promote the interests, values, or good of Y.

There is very clearly an interference condition (1), a consent condition (2) and a benevolence condition (3). In an earlier article, Dworkin (1983) equates violation of autonomy with substitution of judgment (p. 107), moving him close to Shiffrin. The consent condition is of the 'independently of' kind. As pointed out by De Marneffe, Dworkin's benevolence condition is ambiguous between motivational and justificationary interpretations.

Dworkin's recent definition completes the survey. I have returned repeatedly to the three conditions that specify the form of paternalism that Mill rejected in his liberty principle. In the following section, I will summarize the conceptual debate in relation to the anti-paternalism of Mill and Feinberg.

3.2 DISCUSSION OF DEFINITIONS AND MY OWN VIEW

The survey of recent definitions shows a development in our understanding of the interference condition – from interference with liberty of action to presumptive wrongs and on to diminishing of choice or substitution of judgment or agency. However, Dworkin keeps or returns to the more traditional 'interference with liberty', adding 'or autonomy', perhaps with the thought that these two liberal notions may be specified to capture such things as diminishing of choice and substitution of judgment. Since, arguably, it is the interference condition that carries the moral weight of anti-paternalism, it is very important that it be adequately specified (in paper three I argue that it has not been).

Concerning consent, authors in general, very reasonably, agree that the important distinction is between acting with a person's (expressed or possibly inferred) consent and without it (I have passed over a lot of more detailed issues on the nature of consent). Implicitly, Shiffrin incorporates consent into her account of interference, which is very reasonable on the assumption that acting towards a person with her consent is not really to limit her liberty. This assumption is supported by e.g. Husak (1981, p. 31): 'It would seem that a necessary condition for describing an act as an interference with the freedom of the agent is that the agent did not consent to it.' In fact, Mill could be read to make this assumption since he seems to equate 'interfering with the liberty of action' with 'the exercise of power over someone against his will', with an explicit consent condition only in the second formulation. In most of the papers in this thesis, I make the same assumption.

Merging interference and consent, we may talk of the action component of paternalism as that part which concerns the specification of the action type, in all its intricacies, including whether or not and in what sense the action has been consented to. This component can be further delimited in terms of other properties of the target of the action. Gert & Culver and Van de Veer restrict paternalism to targets over some degree of competence. Others explicitly do not, or leave the matter aside. When moving to normative considerations, however, most all find that voluntariness or some such property is essential to distinguishing morally problematic interference. As noted, one of Feinberg's contributions is to take this condition very seriously. On Feinberg's account, it is not interference, in the morally relevant sense, to restrain a person who acts insufficiently voluntarily.

The benevolence condition is rather stable over the definitions surveyed. With Shiffrin the notable exception, the joint assumption is that paternalism is acting for someone's good or benefit, where no distinction is made between harm-prevention and benefit-provision, or between act and omission. This is consistent with Mill's and Feinberg's rejection of the act/omission distinction (Mill 1991 [1859], p. 15; Feinberg 1984, chapter 4). While Mill's and Feinberg's focus is on harm-prevention, this is only natural given that it is less morally problematic to limit a person's liberty in order to prevent harm to her than to provide her with a benefit, if there is a distinction to be made.

On Shiffrin's account, the essential aspect of paternalism is the unwelcome involvement with a person's sphere of legitimate control in combination with the paternalist regarding herself as superior in some sense. Therefore, motive in a strict sense does not matter. De Marneffe criticises Shiffrin on the ground that substitution of judgment may have different motives, but this seems to be part of Shiffrin's very point. Shiffrin explicitly admits that her account is dependent on an account of legitimate control that she does not provide while at the same time insisting that the 'motive' of disrespect for agency is not restricted to rights-violations (pp. 218-9). This implies that there is a sphere of authority where unwelcome involvement is prima facie wrong regardless of motive. This is an embryo to a defence of traditional anti-paternalism, but also more generally of anti-(unwelcome involvement). A shared rationale for both anti-

paternalism and for a broader liberalism of non-interference would to some extent muddle the distinction between benevolence and other reasons for interference. However, since the rationale for anti-paternalism is still very much under discussion, it seems reasonable to keep the possible distinction and confine paternalism to *benevolent* unwelcome involvement.

Pace Shiffrin then, there is widespread agreement that paternalism involves benevolence. Benevolence, however, can be either a motive or a justification, as noted by Van de Veer and brought out more clearly by De Marneffe. With the early Dworkin a possible exception, all definitions surveyed agree that paternalism requires a paternalistic motive. These motivational accounts have the consequence that if I, as a third party, observe you imposing a benefit on a person against her will, for some obscure reason (such as to make good on a bet), and I judge your action to be justified by this benefit, there is no paternalism involved. This is so even if I have the power to stop you and would have done so was it not for the benefit I want imposed on the person (as long as my omission does not count as an interference with the person). If I actively manipulate or force you into imposing this benefit, I may perhaps be said to be acting through you, though this is not at all obvious. Even such cases, therefore, may involve no paternalism. There is also the opposite or corresponding case where I, for some higher purpose, let you impose a benefit on a person, even though I could have stopped you, and even though you are blind to the higher purpose and act only for the sake of the benefit. On motivational accounts, the person is a victim of paternalism even if, were it not for the higher purpose, she would not have been imposed upon.

While there is widespread agreement on the importance of a benevolent motive, the surveyed accounts diverge on whether this motive should be the only reason or the main reason, or whether actions are paternalistic to the extent that actions have this motive (suggested by Gert & Culver and more explicitly proposed by Kleinig, 1983, p. 12), or whether the motive should be necessary, as specified by De Marneffe:

> In my view, what is relevant to the paternalism of a policy is the truth of the counterfactual that it would not be the government's policy had some government official not counted a paternalistic reason in its favor. (p. 74, note 16)

As long as motivational paternalism does not affect the justification of actions, these are all plausible accounts. Motivational paternalism may for example be relevant to the moral evaluation of a person's character. If, however, paternalism is relevant for which actions are justified, as argued by Mill and Feinberg, then only De Marneffe's account is adequate, since on all other accounts a personal good reason can tip the balance in favour of paternalistic action without making the action as such paternalistic. Such tipping would be rejected as unacceptable by anti-paternalists such as Mill and Feinberg. Both Van de Veer and Archard consider the possibility that benevolence may be a contributory reason for interference without being the main reason. They both conclude that such contributory reasons cannot make an interference paternalistic (Archard p. 38; Van de Veer pp. 27-28). Remember, however, Feinberg's definition of legal paternalism:

A personal good reason is a good and relevant, *though not necessarily decisive*, reason in support of interference.

On my own view, the normative core of paternalism is the invocation, in favour of limiting a person's liberty, of a reason that concern her good, regardless of whether this reason is the only reason, the main reason, a sufficient but secondary reason, a weaker contributory and possibly redundant reason, or, along de lines of De Marneffe, a necessary contributory reason in any set of sufficient reasons (this is the thesis of the first paper). If we are concerned with justification, the last form is the most salient. I agree with De Marneffe's caption of justificationary anti-paternalism as the principle that if a policy (or an action) limits the liberty of a person A (without her consent), then 'it is wrong for the government to adopt this policy unless it can be fully justified without counting any benefit to A in its favor.' (p. 76) I take it that this is Mill's and Feinberg's position.[6]

All the definitions surveyed aim to define paternalistic policies, actions, or behaviour. I think this is a mistake. On what I call the *reason account*, what is paternalistic is always the combination of a reason and an action, or in other words a reason for an action. These two components can be specified in various ways to capture various normative positions. As noted, reasons can have different forms of impact and be on different levels of intentionality (motives, justifications). The reason account prepares the ground for principles that operate on reasons, as will be discussed below and in paper two. In contrast, the definitions surveyed are all committed to the *action account*, on which what is paternalistic is an action or policy, which is partly defined in terms of what reasons are (or could be) invoked to support it. The action account is perhaps more true to our linguistic habits, but it is unable to appropriately capture various forms of anti-paternalism.

3.3 DESCRIPTIVE AND NORMATIVE DEFINITIONS

The authors of the definitions surveyed above disagree on to what extent definitions should be normative. Some (Van de Veer) prefer definitions that do not raise moral issues at all, while others (Archard) aim to capture the negative side of paternalism, regardless of whether or not this aspect is balanced out by a positive side. Only Shiffrin explicitly aims to capture something that is prima facie wrong. I propose that this is exactly what definitions of paternalism should do – to capture something that is prima facie wrong, or is regarded to be so.

Dworkin (2008) states on defining paternalism that '[a]s a matter of methodology it is preferable to see if some concept can be defined in non-normative terms and only if that fails to capture the relevant phenomena to accept a normative definition.' I disagree. Not only because, as Dworkin admits, '"paternalism" as used in ordinary contexts may be too amorphous for thinking about particular normative issues', but because a concept that is mainly used to describe a normative position may as well be

[6] On De Marneffe's hybrid definition of paternalism a policy can only be paternalistic if benevolence is both justificationary and motivationally necessary. I take anti-paternalists to oppose beneficial limiting of liberty, regardless of motive.

normative. As far as I can tell, 'paternalism' has no distinct descriptive meaning in everyday conversation, nor in some specialized field or profession. The etymology of the word (Latin *pater* for father) supports characterization along the lines of treating someone as a (good) father would treat a child. In discussions of paternalism in moral and political philosophy, however, we are mainly interested in the benevolent limiting of liberty, not in other areas of parenthood. The main issue, furthermore, is the moral status of benevolent liberty-limiting, and especially the traditional liberal resistance to such liberty-limiting. Best then to define what exactly is resisted, instead of taking a detour around some non-normative definition.

Why prefer non-normative definitions? Donald Van De Veer (1986) argues: 'If we wish to avoid begging the moral question (by simply assuming or supposing an act is wrong in labelling it "paternalistic") we need to identify a morally neutral definition' (pp. 16-17). This is not true. We can argue about whether or not that which is captured by the normative definition is indeed wrong, rather than just suppose that the labelling does the job. It seems that Dworkin and Van De Veer are interested in the concept of paternalism because of its alleged normative properties – its wrongness. They want their definitions to capture this moral controversy. However, for some reasons they do not want the moral controversy to be a defining criterion. I do not see why not. Van De Veer notes that there are many value-laden terms in our language, exemplifying with a long line of pejorative terms such as 'bastard' and 'redneck' (p. 16). However, there are also more theoretically interesting value-laden terms, such as 'justice' and 'betrayal'. I do not know whether Dworkin would prefer to define those terms in non-normative terms or whether Van De Veer would find it question-begging not to do so. I propose, however, that such terms can only be fully understood by giving proper attention to their alleged normative properties.

What might make sense is to try to capture, somewhat more precisely and in less normative terms, what people who have taken an interest in paternalism have been interested in. A majority of the diverse treatments of paternalism over the past decades share a concern with actions that have all of the following four qualities (including those surveyed and also of importance Hodson 1977; Arneson 1980; Husak 1981; Dworkin 1983; Kleinig 1983; Arneson 1989; Archard 1994; Husak 2003; Arneson 2005, as well as many more specialized contributions in medical ethics and other areas of practical ethics):

1) It amounts to (or is believed or intended to amount to) involvement in some person's life that can reasonably be interpreted as limiting or interfering with that person's rights, liberty or self-determination.
2) It is not (or is not believed or intended to be) an immediate response to an informed, authentic and rational request by the person.
3) It is (or is believed or intended to be) to the benefit of the person.
4) Its moral status is not (or is not believed or intended to be) in any obvious way determined by other factors than the interests of the person.

This is the closest I will get to proposing my own non-normative characterization of paternalism. The first and second criteria ensure that the action is in some way problematic in relation to liberal values. The references to reasonable interpretation and obvious justification in criteria one and four, respectively, avoid dependence on normative elements while allowing for commonly accepted relevance boundaries. Criterion one excludes actions (such as asking for the time) that no one finds problematic. Criterion four excludes actions (such as stopping someone from blowing himself up in public; or spending a substantial part of the health budget on a program of extremely costly, forced plastic surgery) that most everyone agrees can be evaluated without considering the interests of the targeted person ('no one' and 'everyone' allows for interpretation). The element of subjectivity, as well as the disjunctive components and the inherent vagueness of many of the terms, make the criteria quite inclusive, while still providing some guidance as to in what way the problem (or family of problems) of paternalism is an ethical problem in its own right, distinguishable from more general questions of rightness.

4. REASONS AND THE GOOD

There is an action and a reason component to paternalism. Paper three is focused on scrutinizing accounts of the action component and paper two on the relationship between the two components. In this chapter I will discuss the reason component in somewhat greater detail than is done in any of the papers.

4.1 KNOWING BEST

It is very common to think that paternalism necessarily involves a presumption on the part of the paternalist to know better than the person targeted what is best for her. This is a mistake. For example, Richard Arneson states:

> The essence of paternalism is overriding the individual's own evaluation of where her own good lies (along with her decision as to the degree to which she will pursue that good by her choices rather than seek alternative goals). Restriction of people's liberty intended to give effect to their current evaluations of where their own good lies and their own present will as to how far it should be pursued are not rightly deemed paternalistic, as many commentators have noted. (2005, p. 266)

I agree with the second sentence. Specifying the liberty value that anti-paternalists are interested in protecting is difficult (this is one of the main points of paper three). However, it seems that this value is not diminished if a (competent) person chooses to have her future options restricted. In other words, restricting a person's options (her 'liberty') is immediately justified by her consent, and so consented-to option-restricting does not limit liberty in the relevant sense. We should notice, however, that it is certainly possible to help someone restrict her future options in accordance with her present will, without agreeing with her on what is her good (and to that extent 'override' her 'own

evaluation of where her own good lies'). I may help Odysseus tie himself to the mast even though I think it harmless to jump ship when the Sirens sing, if tying him up promotes his good in some other way – such as by keeping him on a strict diet, something he himself in no way considers a good thing.

My objection is to the first sentence. I agree with the bit in parenthesis – paternalism must be unwelcome in some sense, it must conflict with a person's decision or will. It is normally important that a person can decide freely how to pursue what she thinks is her good and paternalism is most typically a value conflict involving liberty in this sense. However, overriding a person's decision for her good need not involve disagreeing with her on what is her good, since she need not act on her own view of her good (which is in fact rather straightforwardly entailed by the 'to the degree' and 'how far' bits in the quote). Consider the nicotine addict who believes that smoking is bad for her and yet wants to smoke. She may even believe that in the long run smoking is worse for her than the effects of temporary coercion, so that her all things considered judgment of what is conducive to her good favours coercively preventing her from smoking. Still, as long as she does not consent to being coerced, coercing her in conflict with her present will and for her good would arguably involve paternalism (this case is discussed in paper two and named 'The Smoker'). One might avoid this conclusion by claiming that whenever we can contribute to what a person considers her good by limiting her liberty she is irrational and therefore doing so involves no paternalism. This strong claim seems to depend on the unreasonable assumption that paternalism is only relevant in relation to people who always maximise their own good (or more precisely who promote their good so efficiently that no one else can ever help promote it by the slightest limiting of their liberty).

Furthermore, it is not problematic to disagree with a person on what is her good, as long as one does not limit her liberty. In fact, it is not even problematic to actively promote, for her good, what she does not consider her good, as long as doing so is in accordance with her present will. For example, it involves no paternalism if I help a (competent) person, in response to her request, to destroy what she but not I think is her own good, in order to promote what I but not she thinks is her good. Her motive might be to promote what she thinks is some greater cause. Assume, more concretely, that Anne thinks that her good is best promoted by her having as much money as possible. She asks Betty for help in giving away money to charity in order to promote the greater good of alleviating suffering. Betty does not believe in alleviating suffering but believes that it is better for Anne to give away some of her money (since psychological research shows that practicing altruism makes us happier) and helps her for this reason. Betty is helping to give effect to Anne's decision, but disagrees with her on what is her good.

All this shows that what is morally problematic, and therefore what involves paternalism, is the overriding of decisions, not the overriding of evaluations of good. From the perspective of a presumptive good-promoter, there are three important factors in Arneson's quote – whether we agree with the person on what is her good, whether we restrict her options, and whether we act in accordance with her present will. These three factors can be combined in eight ways and all combinations are feasible. I have agreed

with Arneson that it involves no paternalism to restrict a person's options in accordance with her will. I have pointed out that this is true regardless of whether or not we agree with her on what is her good (Odysseus on a diet). I have argued against Arneson that it does involve paternalism to restrict a person's options against her will even though we agree with her on what is her good (The Smoker). If we do not agree on what is her good, restricting her options against her will is of course a typical instance of paternalism. I have also argued that it does not involve paternalism to disagree with a person on what is her good as long as we do not restrict her options (Anne and Betty). This is true regardless of whether or not the person wants us to restrict her options, as long as we have no duty to do so. If we do have such a duty, not restricting her options, for her good, may involve paternalism. Two non-problematic possibilities remain – we may agree with a person on what is her good and not restrict her options. The person may want us to restrict her options, but if there is no disagreement on what is her good, choosing not to will not involve paternalism (we may just prefer to do something else).

Arneson is an influential interpreter of anti-paternalism but not (any longer) an advocate. In an earlier article (1989), he makes the adequate and interesting observation that the insistence on respecting people's own views of their good can actually undermine anti-paternalism: 'From a single-party welfarist consequentialist standpoint, the insistence that "autonomy trumps" is just another species of perfectionist imposition of values on the agent in defiance of the agent's own considered evaluation.' (p. 435) This is true. However, I prefer, more generously, to assume that the anti-paternalist tendency to insist that people be allowed to define their own good is a confusion.

4.2 ACTING ON JUDGEMENTS

In order for anti-paternalism to be of any interest, the doctrine must operate on reasons that would be relevant and valid was it not for the doctrine. A claim that we have no reason to ever help others does not support anti-paternalism, but rather makes it redundant. To reject anti-paternalism is not to make a claim about what are in general good reasons for action, nor about what is in fact good for a person. It is to claim that reasons that concern a person's good are as appropriate for actions that limit her liberty as they are for any other actions.

Good reasons are most obviously such that refer to the actual effects of an action. Paternalism is perhaps most appealing when limiting a person's liberty does in fact promote her good. Indeed, it is a common position that only facts provide reasons (e.g. Parfit manuscript, section 1.1; Raz 1990. p. 18). However, since what is in fact good is a matter of philosophical and often empirical dispute, and in any case very hard to know, we may want to distinguish between acting on such beliefs as are well supported and such that are not (this point is also made in footnote 11 of the second paper). I therefore propose that we accept as reasons for action also *judgements* to the effect that something of value will be affected.

We may distinguish a number of different judgers or 'authorities', who can make judgements on what is good for a person. Admittedly, a theory of the good can identify the good with what some authority judges to be the good. In such a case, the fact

of the matter and the judgment of the authority coincide (if the theory is correct). It may still be worthwhile to distinguish the two, since people may agree on the normative status of the judgment of the authority while disagreeing on what is the correct theory of the good. An authority may make judgements both with regard to what is the correct theory, and with regard to what the correct theory entails in a certain case (e.g. what makes a certain person happy in certain circumstances). The two most salient authorities in the context of paternalism are the person acting and the person(s) affected by the action, though possible authorities also include third parties, such as experts, people who know the affected person well, and possibly even general opinion. More complex kinds of judgment are also possible and important, especially the acting person's judgment on the affected person's judgment on what is for her good. We may also introduce idealized or qualified versions of these authorities such as the person acting when reasonable and informed, or the person affected when expressing her settled values.

From a liberal perspective, acting toward someone based on one's own judgment on what is for her good may seem dubious, while acting toward someone based on her judgment may seem unproblematic. Things are, however, not that simple. As argued in the previous section, people may well resist an action even if this action promotes their good according to their own judgment. Also, acting on one's judgment on what is for the good of others may be very reasonable and even trivially non-problematic. We may certainly act on our own judgment of what is good for others if this is within our rights, as for example when we make a decision concerning what to give a friend for her birthday based on what we think would make her most happy (rather than what we think that she thinks would make her happy). Often, we do not know (and cannot in any reasonably efficient way find out) what is truly for the good of others, nor what they judge to be for their good. Equally often we cannot help but to significantly affect others one way or other (by our action or inaction). It would then seem preferable to act on one's own judgment, rather than to simply refrain from reasoning, or to disregard the good of others. We may of course try to guess what others would judge to be in their interest, but sometimes we have very little information on which to base such guesses.

I propose that anti-paternalism is equally reasonable or unreasonable for any kind of benevolence, regardless of whether the reasons are based on facts or on the judgment of various authorities. For those who disagree, the doctrine could be restricted to reasons provided by certain kinds of judgments. However, if the resistance to paternalism is based on the idea that we may not limit a person's liberty for her good, then it seems that both the fact of the matter, the affecting person's beliefs, third party beliefs, and the affected person's beliefs should be included. Restrictions of the doctrine that exclude or are limited to some one of these different sources of reasons must explain why other reasons that similarly concern personal good are (not) included. If one rejects anti-paternalism, restricted versions are more reasonable only in the sense that they prohibit a smaller part of what should not be prohibited at all. The top candidate for exclusion is probably the judgment of the affected person; that a person judges an action to be in her own interest should reasonably provide some kind of reason for that action

even if it does limit her liberty. If limiting liberty for such reasons is deemed acceptable, however, it becomes less clear why limiting liberty for other reasons, that similarly refers to a person's good, is not.

Our intuitions concerning the reasonableness of different kinds of reasons for interference admittedly differ. Interfering based on one's own judgment on what is good for another seems perhaps most problematic, while interfering with someone based on her judgment seems least problematic. There are important differences between these kinds of reasons. It is probably true that in general, a person is more likely than someone affecting her to know what is in fact good for her. It may also be true that in general, interference based on the affected person's own judgment of her good is less severe than interference based on the affecting person's judgment of her good – in the former case the person interfered with can at least identify with the motive or aim of the action. This does not mean, however, that reasons based on the affecting person's judgment are not valid. It may also be that we mistakenly assume that limiting a person's liberty in order to promote what she judges to be her good is not really limiting her liberty at all, since in a way she wants the action to take place. As I argued above, this is not so, since our decisions and our views of our good can diverge. However, the impression that there is this connection may still bias our intuitions.

4.3 HEALTH PROMOTION AND MORALISM

In arguing against anti-paternalism I in no way intend to argue against anti-moralism. However, C.L. Ten (1971) and Heta Häyry (1992) have argued that these isms cannot be properly distinguished by the values or goods they aim to promote. Even limiting liberty in order to avoid physical harm, a rather straightforward personal good, involves a moral commitment, they claim, which leaves paternalism undistinguishable from legal moralism.[7] Häyry states: 'If legislators "know" – that is, are licensed to define – for ordinary people what is best for the individuals' own physical good, then they presumably also "know" – are licensed to define – what is best for their moral good.' (p. 196) It is not clear why Häyry takes this position. She refers to arguments by Devlin (1965) to the effect that defence of physical paternalism is necessarily founded on 'rationally untenable distinctions' (Häyry, p. 196). However, Hart's (1963, pp. 30-32) distinction between (justifiable) paternalism and (unjustifiable) moralism, further developed by Feinberg (e.g. 1984, pp. 12-13), is quite clear. Legal moralism involves preventing the inherently immoral, regardless of whether it harms anyone. Paternalism involves preventing harm (or promoting good) to individuals. While it should be recognized that attributing negative value to physical harm is taking a moral stance, to attribute negative value to certain behaviour or lifestyles as such is another and distinct stance.

[7] Ten and Häyry both hold that the distinction between paternalism and moralism can be upheld by restricting paternalism to interference with acts that are insufficiently voluntary. However, paternalism need not be so restricted and the point here is that Ten and Häyry claim that no distinction can be upheld between the kinds of values that are promoted by the respective isms.

Ten argues that what is valued by the typical paternalist is not absence of physical harm as such but rather the absence of immoral physical harm: 'Intervention is only [thought] justified when the harm is caused by the commission of an immoral act.' (p. 57) While surgical operations are generally accepted, sterilization is not, though both involve physical harm: 'The difference between the two cases seems to be based on the fact that surgical operations are not generally regarded as immoral, whereas sterilization is so regarded by some religious groups.' (Ibid.) Ten's argument, then, seems to be that if the paternalist was really concerned to protect physical health, she would go after surgical operations as well. In fact, however she only favours intervention when physical harm is caused by something she finds immoral, such as sterilization.

In reproach, why should we think that surgical operations are accepted because they are morally impeccable, and not because they are conducive to physical health? It is true that an operation can be dived into steps which taken in isolation can be described as harmful (cutting someone's stomach open etc.), but such arbitrary division can prove all sorts of acts to contain harmful or in other ways counter-intuitive parts. Surgical operations are performed with the aim of contributing to physical health and they most often do. To the extent that operations are questioned on normative grounds (such as with sterilization), this is exactly because they are atypical in that they might be seen to destroy rather than contribute to physical health. Immoral actions of course tend to be more susceptible to paternalistic intervention because they tend to be harmful. This does not show that it is their immoral nature that is the basis for intervention, rather than their harmfulness.

5. VALUES, REASONS AND PRINCIPLES

To my mind the two most important notions in moral philosophy are *value* and *reason*. What has value? What do we have reason to do? These seem to me the fundamental questions of morality and more generally of practical reasoning.[8] Complex empirical relationships between actions and outcomes ensure that we cannot move easily from the first to the second. However, when something of value (i.e. something important) is affected by an action, this gives us a reason of some strength to perform or avoid the action. In other words, if we can effectively protect or promote (or diminish) some value, we always have reason (not) to do so. To oppose this common sense assumption is to muddle the distinction between considerations that are in a basic sense relevant, such as that a person will be harmed, and considerations that are utterly irrelevant, such as that a person will sneeze. Granted this assumption, many doctrines in moral philosophy can only be understood as influence-regulating principles that prevent reasons from having influence on the moral status of actions according to their strength. I propose that moral inquiry would benefit from rejecting all such principles and focusing on the two fundamental questions.

[8] Perhaps these questions should be complemented with one or other agent-focused question, such as: What kind of person do I want to be? Nowhere in this thesis are such virtue ethical issues considered, unfortunately.

5.1 REASONS WITHOUT INFLUENCE

John Broome (2004) defines reasons in terms of their role in explanations of ought facts. Modifying his former view to allow for deontic principles, Broome states that these explanations may be merely 'potential', not 'actual' (pp. 40-41). In other words, reasons are entities that would explain ought facts if it were not for deontic principles. I prefer to say that what reasons do is that they *influence the moral status of actions*. That is, they do so unless their influence is blocked, typically by deontic principles.[9]

It is a rather common idea in moral philosophy that certain considerations should be disregarded on normative grounds and that certain other considerations have a special status that goes beyond their relative strength as reasons. Indeed, one could argue that this is the essence of a deontological approach to ethics. For example, Jonathan Dancy (1993), building on and adapting from John McDowell, argues that reasons can be 'silenced' by the presence of other reasons, or by other circumstances (pp. 47-58).[10] Interestingly, for Dancy silencing does not occur in accordance with principles, but is particular to each situation. Apparently at the opposite end of the particularism-principalism debate, Thomas Scanlon (1998) argues concerning reasons in general that certain 'considerations' may make it the case that a reason is not 'relevant', which is distinguished from reasons being 'outweighed without loosing their force or status as reasons.' (pp. 50-51) More specifically on moral justification, '"being moral" involves seeing certain considerations as providing no justification for action in some situations even though they involve elements which, in other contexts, would be relevant.' (p. 156) Or, more briefly, 'certain reasons for action [...] are morally inadmissible.' (p. 201)

Scanlon says that considerations *lose their status* as reasons. Dancy at one point says that a reason that is silenced is in fact 'no reason at all' (p. 60).[11] It is common to talk this way, to say that something that would influence the moral status of actions were it not for some normative blocking is not a reason. It is also common to say with Scanlon and Dancy that these things are not relevant. This is confusing. There is an interesting distinction to be made between such effects of actions as people being harmed and people sneezing. The distinction is due to the great difference in the importance or value of these effects. There is also an interesting distinction to be made between things that are effects of an action and things that are not. If we say that deontic principles or particularist 'deontic circumstances' make reasons irrelevant non-reasons, we would be hard pressed to invent a new vocabulary to express these important distinctions. Better to say that these deontic entities block reasons from having influence on the moral status

[9] I prefer this terminology mainly because I want to understand deontic principles as *regulating* the influence of reasons, rather than dispelling reasons to the realm of the merely potential. However, this could perhaps be accomplished by staying with Broome's terminology and making room for deontic principles in the explanation of ought facts. I also prefer my terminology because I prefer to avoid talk of facts in morals, and because I am not as comfortable as Broome is with the primitive notion of explanation.

[10] In *Ethics Without Principles* (2004, p. 41) Dancy talks of 'disablers' of reasons that seem to have the same silencing function. The later work, however, includes few moral examples and those given seem motivated by practical constraints.

[11] Dancy's position has not changed in its fundamentals. In his entry on 'Moral Particularism' to *The Stanford Encyclopedia of Philosophy* (2008), he states that 'what is a reason in one case may be no reason at all in another'.

of actions, though effects of an action that concern things of value are in a very basic sense relevant and when we can produce such effects we have some reason to do so. This is only a terminological matter, but since the notion of a reason is fundamental it is a relatively important one.

Dancy (1993) argues in support of moral particularism that the fact that an action causes pleasure normally counts in favour of performing it, but not if the pleasure is malicious; that the fact that we have borrowed a book normally counts in favour of returning it, but not if the book was stolen from the library; and that the fact that an action is a lie normally counts against performing it, but counts in favour of performing it if lies are required as part of a game we play (pp. 56, 60-61). It is not clear if lack of value or normative regulation of the influence of reasons is doing the work here. What is clear is that if we should disagree about these claims, it would be informative to learn whether the disagreement concerns value or influence. Myself I would say that it has no value to avoid lying in lying games, but that it has some value to return borrowed books, even to thieves, though perhaps more value to return them to their rightful owner. I would therefore be curious to know why exactly Dancy thinks that there is no valid reason to return the book.

5.2 ANTI-PATERNALISM AS AN INFLUENCE-REGULATING PRINCIPLE

There are many forms of anti-paternalism. Some are straightforwardly consequentialist, pointing to the likely counter-productive effects of certain institutional arrangements. Others are more sophisticatedly consequentialist, questioning the value of interference-generated personal good (such versions are briefly considered in paper two). This thesis is mainly concerned with the more deontic, more principled, normative anti-paternalism that I find in Mill's liberty principle and Feinberg's soft paternalism. Such anti-paternalism can most favourably be understood as the position that *reasons that concern a person's good are invalid for actions or effects that limit her liberty*. That a reason for an action is invalid means, on my understanding, that the reason does not influence the determination of the moral status of the action. Invalidation of reasons is one form of influence-regulation. Other forms include side constraints and lexical orderings of reasons. The distinctions between these forms of regulation are discussed in paper two. That we sometimes need to consider not only actions but also various effects of actions is explained in paper one.

5.3 AGAINST ANTI-PATERNALISM

The case against normative anti-paternalism is made most thoroughly in paper two and three of this thesis. One way to capture the general problem with anti-paternalism is to point to the difficulty in specifying the notion of limiting liberty in such a way that any reason that concerns a person's good is invalid for any action that limits her liberty. A person's good includes such major goods as survival and preservation of autonomy. That there should be some specification of limiting liberty such that these major goods do not provide any valid reasons for any action that limits liberty is simply unlikely. There are of course very narrow specifications that would do the job, but not without making anti-

paternalism redundant (since the allegedly invalid reasons are always outweighed anyway). This difficulty implies at least three arguments against anti-paternalism.

First, there is no action type such that the fact that an action of that type promotes personal good is always morally problematic (this is the argument of paper two). Second, there are cases where it is obviously justified to limit a person's liberty (partly) for her own good, including cases where the person is perfectly informed and rational and all agree on the relative strength of the relevant reasons as far as they concern her interests (this argument is developed in paper three). Third, drawing a line for when reasons that concern a person's good are valid and when they are invalid gives rise to peculiar jumps in justifiability, such that one action is overwhelmingly justified while another, very similar action, is overwhelmingly unjustified (this argument too is developed in paper three).

Admittedly, we may have a special need to be suspicious of reasons that refer to our good. Such reasons are more likely than other reasons to be invoked in rhetorical attempts to persuade us: 'It will benefit you too.' Such reasons are also commonly invoked for actions that mould us in the interests of others, especially when we are young: 'It is just for your own good.' We do not have the same need to develop a mistrustful attitude towards reasons that refer to the interests of others. This may possibly explain intuitions that tend towards anti-paternalism. It does not, of course, provide an argument for the doctrine.

On a more general level, I find deeply problematic the idea that reasons to effectively promote something of value can be invalid. Sometimes we cannot consider all relevant reasons because of practical constraints, including limited time, information and processing capacity. This does not mean, however, that they should be disregarded on normative grounds. I propose that secular morality is based on judgements of value. For any moral demand, we can ask whether it has value that we meet it. This is not simply to endorse consequentialism, or if it is this is consequentialism in a very broad sense (Cf. Sen 1982). It may have value that we respect rights, do our duty or are virtuous. More to the point, that I respect a certain right or fulfil a certain duty or attain or display a certain virtue may have value regardless of consequences. However, it cannot be morally prior to some other moral imperative in any other way than having more value. Values will conflict of course, and one of the main tasks of moral philosophy is to investigate the nature of various values and how values relate to each other, in order to understand and if possible resolve such conflicts.

I boldly propose that any non-redundant doctrine of invalidation will confront problems analogous to those that face anti-paternalism – difficulty specifying the relevant action type, wrong answers to moral questions in certain cases, and peculiar jumps in justifiability. Furthermore, if a reason has the special power to invalidate other reasons, this has the unfortunate consequence of putting within brackets the fundamental question of the relative strength of different reasons (this is a further argument against anti-paternalism developed in paper two).

In addition to the substantial arguments against anti-paternalism developed in the papers, there are also methodological reasons to prefer value conflicts over influence-

regulating principles. Investigation of what things have value and how they generate reasons for action can contribute to our moral outlook regardless of what exact relative value we attribute to those things or what other things we think have value. This is true both on an abstract and on a very concrete level. Say that you and I argue about whether or not to recycle our household trash. You explicate the value of a clean environment and I the value of a clean kitchen. You may point to the deep feeling of harmony and connectedness that can only be experienced immersed in untouched wilderness, while I may point to the elegance and convenience of putting all the trash in one single container. Our respective arguments can contribute to each our worldviews even if we disagree on the relative importance of these values.

In contrast, if you say that we should recycle as a matter of principle (perhaps because reasons that only concern convenience or aesthetics are invalid) and I disagree, then this principle may tell me nothing at all. More generally, if there is a principled rule that we think is incorrect in one instance, then the rule is falsified and it is unclear how we should relate to it – whether we should, for example, amend it or give it up completely. The rule itself gives no guidance in such matters. This is especially problematic since moral principles always end up being very complex, in order to avoid unreasonable prescriptions (one can of course bite the bullet and for example hold with Kant that we should simply never lie, but to avoid giving the wrong answer to moral questions, more complex treatments are needed, such as Korsgaard's (1986) qualified defence of this Kantian principle).

If following a rule is not a strict deontic requirement but simply something that has value, then the rule can find its place among other values. Its value can be contrasted with that of other rules and other things. Further specification of what exactly is valuable in following the rule will tell us whether it has value to follow it as often as possible or if it only strict compliance has value, or whatever.

The methodological case against principles is even stronger in politics. As individuals it has value that we interact with each other on equal terms. If the value of equal standing is very great, it may entail something akin to deontic principles regulating how we may behave towards each other and what considerations we should and should not take into account. However, society is not a moral agent that needs to foster a sense of its own proper footing. While it is certainly of value that people are not coerced, for example, it is hardly of further value that certain institutions do not coerce. It may be a bad that a person is an agent of coercion, independently of the effect that someone is coerced. Even failed attempts to do bad may possibly taint a person's character. However, institutions have no character. It is not bad in itself that the state is involved in coercion. If, implausibly, designing certain state institutions to attempt to coerce people would have some positive side effect, yet not have the effect that people were coerced, nor the agents of coercion, and if nothing else of value was affected, then the quasi-coercive nature of these institutions would not be a reason against this order. Individual deontic restrictions are, from the perspective of political philosophy, like practical constraints, part of the background against which institutions are designed.

My claim that principles are especially ill suited for politics may be surprising, given that they are so common in that area. This is partly because I do not accept the popular premise that state actions stand in need of some special justification (legitimacy) that is not required of private actions. I take it that, *ceteris paribus*, coercion by the state is not worse than coercion by some private party (Cf. Pettit 2002). Partly, the explanation is that I am abstracting from practical constraints and empirical circumstances such as the tendency of power to corrupt etc.

5.4 CONSTANT OVERRIDING

If, for a certain type of action, con reasons of a certain type are always stronger than any possible pro reasons of a certain type, the con reason will always override these pro reasons. Such regularities can perhaps be codified into principles. For example, for anti-paternalist constant overriding, the con reasons would concern liberty and the pro reasons would concern personal good. The effect would be very similar to that of invalidation – paternalism would be prima facie wrong. If my arguments against invalidation above and in the papers are convincing, a principle of constant overriding may seem an obvious alternative.

There are two important weaknesses with a principle of constant overriding. First, such a principle has nothing to say about cases where we have both personal good reasons (etc.) and other reasons for limiting liberty (etc.). All the principle says in such cases is that there are con reasons stronger than the personal good pro reasons. We do not know how much stronger and we do not know what will be the balance of reasons when we consider also other pro reasons. This weakness in itself is severe enough to make constant overriding a rather uninteresting interpretation of anti-paternalism, or principle more generally.

Second, it is simply difficult to see how one kind of reason can always be stronger than another kind, given that they are both valid. I may hold, for example, that in general liberty is more important than life and health, but how can it be that even the most trivial limiting of liberty will override the most enormous damage to health? If I can save you from certain death by pushing you out of the way of an approaching bus, though this will limit your liberty to move about without being pushed, how can the former reason be overridden by the latter? If your standing in the way of the bus is a deliberate suicide attempt, my pushing you might interfere with your right to decide over your own life and death. This right may be a particularly important aspect of liberty. However, being free to move around without being pushed is arguably another aspect of liberty and the one the example is concerned with. Assume therefore that your standing in the way is not a suicide attempt.

One might try to define the overriding reason type in such a way that reasons of this type are always very strong. The reason we have not to push people around is perhaps not a *significant* liberty reason, while such things as the freedom to vote or to marry do provide significant liberty reasons against interference. This strategy of course makes for a rather narrow doctrine, leaving many liberty-limiting actions outside of its domain. Similarly, other narrowly defined values could give rise to other prima facie

wrongs. One example may be the destruction of one culture for the expansion and material gain of another – call this colonialism. If colonialism is prima facie wrong this is not because of some special relationship between the value of the non-destruction of a culture and the value of expansion and material gain, but simply because anyone's reasons against destroying a culture are always stronger than her reasons for achieving expansion and material gain.

Even more narrow principles are hard to formulate in ways that make them reasonable, unless phrased in terms of great disvalues such as the destruction of a culture. Liberties of kinds that are normally very important may on occasion be relatively unimportant. Suppose that you lack money for urgently needed medical attention and you plan to marry a certain person on a whim. You reason that you won't live long anyway because of your predicament. Now someone (a demon perhaps) pays me a large amount of money to stop you from marrying this person. I do so and use the money to save your life. This seems to me justified or at least close enough to being justified to raise doubts that non-limiting of the freedom to marry always overrides health concerns. There is of course no end to how narrowly the principle may be defined. At the extreme, anti-paternalism would be the principle that personal good reasons are always overridden by our reasons not to eradicate every single trace of self-determination from a person's life. Somewhere short of this point anti-paternalism will no longer merit the name of principle, being simply a (short) list of certain reasons that override certain other reasons.

5.5 LIBERALISM WITHOUT ANTI-PATERNALISM

Anti-paternalism is an integral part of the liberal tradition. However, one may well affirm the values of liberty and autonomy without embracing anti-paternalism. For one thing, it is not obvious that these values are best promoted or protected by non-interference. My reason for limiting your liberty may well be that if I succeed, you will have more valuable options to choose from in the future and be a more self-reliant person (or whatever similar properties are taken to embody liberty and autonomy). This point has been made again and again in the literature (since it was made by Dworkin 1972). Paternalism does not entail a net loss of liberty over time; it merely entails a limiting of liberty in some more specific sense, a sense that is not easily distinguished (as argued in paper three).

Let us assume for a moment that limiting liberty can be clearly distinguished so that it implies a loss of value. Now, why should we not simply compare this loss or disvalue to the positive value that will result from limiting liberty? Stopping you from taking great risks may involve a disrespect of your autonomous will and your decision to take those risks. This is a bad thing. That you stay alive and healthy, on the other hand (and that your long term autonomy is thereby preserved) is arguably a good thing. How the situation should be evaluated apparently depends on such things as the significance the choice of taking the risks has for you, how my stopping you will affect your autonomy and sense of self-respect, what your life will be like if I save you etcetera. These specific matters may indeed depend on more general considerations concerning such things as the value of respect for autonomous choice and the value of human life.

However, there is no apparent need for the liberal to simply disregard some aspects of the situation.

It may seem that I am unnecessarily creating a theoretical problem. In practice, there must be some point were we should let people take their own risks. But this is not the issue. The issue is what grounds there are for allowing people to take their own risks. Why should people sometimes be free to take their own risks? A plausible answer is that they should be so when the value of (some kind of) liberty outweighs the ensuing disvalue of harm or risk of harm. Anti-paternalism, however, implies that in cases of conflict between values, we should first consider whether some person's liberty is limited (in some sense). If it is, we must remove reasons that concern the good of that person from consideration. Only after this weeding out of personal good reasons may we move on to consider the reasons that are still valid, in order to reach an all things considered judgment. The need to make practical judgements does not in itself call for weeding out reasons in this way.

Absolute prohibitions may be appropriate as rules of thumb, as constitutional restrictions on government, as laws or as codes of conduct. If they are, however, this is because people can not be trusted to weigh values properly, so that the consequences of insisting on such prohibitions are better than the alternatives, in terms of the overall promotion and protection of values. Our general inability to properly consider reasons should not, however, circumcise theoretical discussions of morality, whether it concerns private action or public policy.

5.6 MORE GENERAL THEORIES

Anti-paternalism, as understood here, can be a consequence of various more general normative positions. State anti-paternalism could be a requirement of state neutrality or anti-perfectionism in the sense that no government action should be justified with reference to the good of anyone. Another form of anti-paternalism follows from libertarianism, with its opposition to infringement of ownership rights, whether benevolent or not (though it is presumably consistent with libertarianism for the owner of a road to require that people only drive on it with seatbelts, and helmets, and gum shields, or for the owner of an apartment complex to require that the tenants submit to drug testing, or adopt strict diets and exercise regimes). Another ground for anti-paternalism is a Kantianism according to which treating others as ends requires that benevolence should only manifest itself as aid to people in promoting their ends as defined by themselves, assuming that duties to oneself may not be enforced by others (e.g. Baron 1997, pp. 13-15) (the central notion here is 'aid' rather than 'as defined by themselves', since the latter is consistent with some forms of paternalism, as noted above).

With the exception of some exotic versions, consequentialism is incompatible with anti-paternalism on the level of rightness, since its central claim is that all consequences of an action that have value should be considered. However, consequentialism can incorporate anti-paternalism as a rule of thumb if the consideration of some reasons for some actions is typically counter-productive or simply not worth the

effort. More univocally, anti-paternalism is incompatible with many forms of communitarianism. For example, a person's identity may be understood to necessarily involve certain ideas of the good, implying that it makes little sense to give priority to non-interference over promoting those goods (Taylor 1989, pp. 26-27).

In the literature commented on in the first two chapters, the detailed discussion on the proper definition and moral status of paternalism has to some extent been kept separate from more general ideas on morality and politics (though commonsense consequentialist arguments are common). Perhaps this is unavoidable in dealing with any more specific moral problem. In our case the tendency is strengthened by the fact that important figures such as Mill (in *On Liberty*) and Feinberg do not bother much with the most abstract level of rightness, but are more concerned with values or reasons on an intermediate level – between final value and purely empirical considerations – and the principles that allegedly govern them. Mill of course was a utilitarian and there has been much discussion on how his anti-paternalism fits into his more general understanding of morality. I tend to side with C.L. Ten (1980) in thinking that Mill's liberalism and his utilitarianism simply do not add up, but that he took them both very seriously. Feinberg (1984) explicitly avoids questions of final value, calling his four volume work on the moral limits of the criminal law an 'extended essay in applied moral philosophy' (p. 4).

This thesis follows the custom of keeping some distance between the local issue of paternalism and global issues of morality and politics. On its most theoretical level, the thesis concerns the proper understanding of paternalism and the reasonableness of rejecting it on normative grounds. More general theories are not considered in the papers. Assume that anti-paternalism follows from some more general theory. This is how I would understand the relationship between the arguments against anti-paternalism presented here and such a theory: 1) If there are decisive arguments for the theory independently of the reasonableness of anti-paternalism, then my arguments are in a sense irrelevant, though they may still provide some insight and may contribute to undermine the decisiveness of the decisive arguments in some (future) contexts. 2) If the theory gains support from the alleged reasonableness of anti-paternalism, my arguments, to the extent that they are sound, weaken this support, possibly undermining the theory. To exemplify: It could be argued against my approach that everyone has an absolute right to control her own body and property, and so that it can never be justified to limit that control for whatever reason other than to protect the like rights of others. If the premise of this claim is some idea of ownership that is independent of intuitions or arguments to the effect that people should not be interfered with for their own good, then the claim must be evaluated according to whatever standard is appropriate and if sound will imply anti-paternalism. However, if the idea of absolute ownership is to some extent based on anti-paternalist intuitions and arguments, then to that extent it is undermined by the arguments presented here, if sound.

6. LIBERTY OR HEALTH?

With its focus on arguing against anti-paternalism, this thesis is largely critical, though in interpreting paternalism and anti-paternalism, and in proposing aggregation rules for individual into group consent, as well as in taking a stand on certain public health policies, it is more constructive. In any case, something more positive should be said about the central values of health and liberty. This last proper chapter of the introduction, therefore, contains some rather undeveloped thoughts on the nature of these values. In connection, I present some views on libertarian paternalism, which has received so much attention lately.

In essence, I propose that there are many kinds of liberty and of health and that they are all important. As with the concept of paternalism, I think that interesting explications of these notions are generally normative. I follow Rawls (2001) in preferring an account of liberty where 'no priority is assigned to liberty as such, as if the exercise of something called "liberty" had a preeminent value and were the main, if not the sole, end of political and social justice.' (p. 44) The liberty values I shall soon identify, however, are rather more abstract and general than Rawls list of basic liberties. They are more on the level of abstraction of Berlin's (2002 [1969]) concepts of positive and negative liberty, of which he states: 'Both are ends in themselves' (p. 42).

6.1 HEALTH

In public health policy, what the government and public health authorities are aiming to promote (if uncorrupted) is not people's good per se, but their health. Health may be defined biomedically as the absence of disease and infirmity. This is the common definition in medical practice. The biomedical account could be complemented with the idea of survival and reproduction as basic functions of organisms (Boorse 1975). Alternatively, health may be defined biopsychosocially, which is common in theoretical contexts. The constitution of the World Health Organization (1946) famously states that health is 'a state of complete physical, mental and social well-being'. Several recent definitions of health aim to avoid the somewhat utopian character of the WHO definition and to shift focus from outcome to opportunity, by defining health in terms of potential, ability, or resources, rather than well-being (Bircher 2005; Law & Widdows 2007). On these wide definitions of health, the distinction between health and good more generally is not sharp, though such things as finances and achievement that could be considered part of the good are presumably excluded.

Health is obviously a contested concept and the debate over its nature is in part normative – what kind of well-being or functioning, if any, has value for all people. This is an important question which does not necessarily have only one answer. There could be several concepts of health which all have value for all, or most, people. I believe that both well-being and potential have value, for example, and perhaps even normal functioning independently of the first two. In any given context, health can quite properly be as broad a concept as is coherent with some particular policy promoting health so defined, as long as it captures something of value.

6.2 LIBERTY

Anti-paternalists seem most eager to protect the freedom of the capable (informed and rational) to do what they want to do. They tend to think that this value must be balanced against other values, including the freedom of other capable people to do what they want and the prevention of harms to others, but that it trumps (other) personal goods or values in the lives of the capable. While I agree that the freedom of the capable is a value, I do not believe in trumping, as should by now be clear. I also believe that the freedom of the capable is not a value in itself but that it can be derived from two other values. First, it has value that people are free to do what they want to do, here and now. This is important for anyone, also for less capable judgers such as minors, the ignorant, the immature, the intoxicated and the emotionally upset. It is important even for small children and the outright insane, at lest in some areas of their lives.[12] Second, it also has value that people preserve their integrity or coherence as they exist over time. In the interest of this latter value, what a person wants to do, here and now, can sometimes be overridden by consideration of what she would have wanted to do if she had been more sober, emotionally stable, rational, mature, informed etc. These two values will of course conflict more often for less capable people. Note that the value of integrity can motivate overriding the present will of a person, regardless of whether or not harm threatens. We may have reason to stop a person from making a fool of herself in public, simply because this would radically change the impression other people have of here and so how they relate to her and so her existence as a social being. This may be true even if the change would not be harmful, but rather good for her, because, for example, it would cause her to be more relaxed, to 'lighten up'.

The freedom to do what one wants to do is not, on my understanding, restricted to those things that are under one's legitimate control. It has some value for me to use your property or to decide what you should do. *Ceteris paribus*, it is better that I can live in your house and drive your car. If you are away and don't mind me using your things (with or without your permission), there are no reasons against my doing so and some for. However, quite often (though not as often as most would expect) it is more important that people have long term control over certain resources (their bodies, property and belongings), in order that they be able to shape their life according to plans and goals, and perhaps in order to stimulate economic activity and development.

Identifying as a value this freedom to do what one wants to do, quite generally, avoids the need to attribute value to self-determination in the sense of freedom to do what one wants to do in some more limited sphere of one's life (the self determining the *self*). What each person should have control over, all things considered, depends on some balancing of this general liberty value as it is manifested in each person's life, and against other values. I am aware that many have held that freedom of this general form can have no value, but I disagree. However, the value of some options is negligible and that of

[12] This bias in value is distinct from the important point made by Arneson (1989; 2005, section V) that anti-paternalism has unfortunate distributional effects since bad choosers have more to gain from paternalism than good choosers.

others great. Therefore, it may make sense to distinguish as independent values, in a Rawlsian manner, doing what one wants in particular areas (worship, movement).

The anti-paternalist insistence on non-interference naturally invokes the value of negative liberty – freedom from constraints. However, in explaining why we are nevertheless justified in interfering with people who act insufficiently voluntarily, anti-paternalists tend to invoke the hypothetical, more informed and rational self. Since they oppose benevolent limiting of liberty on principle, they tend to argue that restraining a person acting insufficiently voluntarily is not really restraining her at all. Indeed, this is 'no more illiberal than interference to prevent him from harming or offending an unwilling second party.' (Feinberg 1986, p. 12) This amounts to an appeal to that most suspicious form of positive liberty – the real will or real self, quite separate from the will of the actual person. Berlin (2002 [1969]) rightly warned against this 'monstrous impersonation, which consists in equating what X would choose if he were something he is not, or at least not yet, with what X actually seeks and chooses' (p. 180). It is tempting to reject such tendencies completely and attribute value only to the will of the present, actual self. However, there is no escaping considering the person as a more or less abstract entity existing over time. For how else would we attribute some value to treating people in sleep or unconsciousness according to their presumed wishes, wishes that they may not explicitly have formulated but that are easily inferred from their earlier wishes, choices and behaviour. It is partly to account for such situations that I suggest that integrity or coherence over time is a value. If this value is acknowledged, we need never appeal to the will of the person's more informed and rational self (if we still do that self will often sanction the same kind of involvement).

If we shall respect and preserve the integrity of a person, we better think hard on what this is exactly. There is an obvious risk that we take a person's everyday social persona to be her core self. Are people most truly such as they are most often? Or are their core selves better revealed in some circumstances than others? These questions seem to invite an outside, second person view of who a person most truly is, which may of course be influenced by normative assumptions of what aspects of the person are most valuable. One option is to go with a subjectivist account, according to which what is most central to any particular individual's self is what she herself identifies with. However, here we have the definiendum in the definiens. So we would have to specify what aspect of her it is that identifies who she most truly is. One way or other, we need some more objective account to answer what it means to preserve integrity.

I have focused on those liberty values that I take to be most central to paternalism. This far from exhausts the complete list. There is of course a rich literature on liberty and freedom that may inform our investigation of liberty-values. Some focus on available options or capabilities (Raz 1986; Sen 1992). Others focus on what causes us to have a certain bundle of options or capabilities, making distinctions between options being constrained by non-human nature or by human action, perhaps in turn divided into intentional constraint and unintentional (Berlin 2002 [1958]), or culpable and non-culpable (Nozick 1974), or according to whether or not someone can be held responsible for the constraint (Kristjánsson 1996). Yet others focus on the possibility that others will

interfere to limit our options (Pettit 1997). To me, it seems important that we have a wide range of good options to choose from, particularly when it comes to life-defining choices such as what careers to pursue and where to live and with whom. It does not seem important what has caused us to have some rather than other options, or whether someone is culpable or responsible for us having the options we have. Leaving be is not leaving free. I am undecided on whether it is important that others cannot interfere, independently of whether or not they actually do. Questions of what has value are, however, notoriously difficult to argue.

There is also a partly overlapping discussion on autonomy and its value. I believe that autonomy is an (or perhaps several) independent value(s) in its own right. Dworkin (1988) defines autonomy as 'a second-order capacity of persons to reflect critically upon their first-order preferences, desires, wishes, and so forth and the capacity to accept or attempt to change these in the light of higher-order preferences and values.' (p. 20) I agree with Dworkin that this is an important value. While it is important that we can do what we want, and choose from a wide and rich range of options, it is equally important that we can reflect on our wants and, to some degree, change them. Mill's anti-paternalism is in part based on his conjecture that the doctrine will promote the value of individuality, in the sense of living an original or even eccentric life, performing 'experiments of living' (1991 [1859], chapter III). Perhaps there is a special value in this too (no doubt it serves to enrich our view of the possible, which is perhaps Mill's main reason to favour it).

Capabilities can be interpreted as a super liberty value, encompassing all other liberty values. We should distinguish, however, between what is important in itself to have or be, and what is important that one has the option of having or being. The concepts of liberty and autonomy identified are of the former kind and so have value independently of their possible incorporation under capability.

6.3 LIBERTARIAN PATERNALISM

Thaler & Sunstein (2003a, 2003b) have established the term libertarian paternalism for the promotion of healthy choice through non-intrusive influence. Libertarian paternalism is paternalism in the sense that 'it attempts to influence the choices of affected parties in a way that will make choosers better off.' (2003b, p. 1162) It is libertarian in the sense that 'people should be free to opt out of specified arrangements if they choose to do so.' (p. 1161) That is, it is libertarian in the weak sense that it does not conflict with the freedom of choice of the parties. Thaler & Sunstein accurately and importantly note that such freedom is consistent with a great deal of influence, since, because of the great impact of default rules, starting points and framing effects, there is no way to track people's preferences as they exist independently of choice situations. These observations support the attribution of value to being free to do what one actually wants to do, rather than to realize some more hypothetical kind of preference. People often do not have well-formed preferences independently of choice situations, but once they are in one, they usually quickly form a preference, and then it is important that they be free to act on it. Since Thaler & Sunstein are not particularly impressed with natural ownership rights,

and since their doctrine does not positively promote libertarianism (one could imagine influence on choice that would promote autonomy, self-reliance, or a libertarian lifestyle of non-interference), it might perhaps more appropriately be called *non-intrusive paternalism*.

I think that it is obvious that we should, when possible, ensure that the choice situations that we face are as good as possible. This means in part that the alternatives available should be good ones in the sense that they tend to promote what has value (but variation also has value – independently of outcomes). If there must be an ordering of the alternatives such that some are more easily accessible than others, as is most often the case, this ordering too should promote what has value, such as health. I therefore subscribe to a version of the doctrine of libertarian or non-intrusive paternalism. Choice situations should in large part be designed so that we tend to make good choices. Such design is an important task for public health policy.

Thaler & Sunstein's most basic example is the placement of the desert in a cafeteria, where the director simply has to make a choice one way or other where to place the desert – early or late in the line. However, they do not shrink from much more far-reaching conclusions:

> But governments, no less than cafeterias (which governments frequently run), have to provide starting points of one or another kind; this is not avoidable. As we shall emphasize, they do so every day through the rules of contract and tort, in a way that inevitably affects some preferences and choices.' (p. 1165)

Concrete examples include US labour regulation such as the 40 hour working week and the requirement that employees be discharged only for a cause, or else be compensated (p. 1187).

Policy makers must normally choose, or act as if they had chosen, between regulating and not regulating. Regulation affects people directly and through the regulation of other-regarding behaviour such as desert placement in cafeterias. Concerning the latter, not regulating means that people will make other-regarding choices based on their own self-interest, or on some other reason or influence. In the cafeteria case, the director might just place the desert where it is usually placed, or roll a dice. I propose that we should not passively stand by and let the society and environment we live in be shaped by habit, chance and special interests. We should actively design our environment to further the kind of life that we value. If we disagree on what has value, this is a reason for discussion and compromise. Policy makers should, directly and indirectly, promote good choice situations. If some choice-affecting policy is of great consequence, this is no reason not to enact it, though it is perhaps a reason to submit it to public scrutiny and debate.

Thaler & Sunstein's focus on employee savings is instructive. Governments can regulate employers to ensure that their employee policies are conducive to good health. Governments can also enact welfare programs, impose taxes, offer subsidies, and design public spaces and the infrastructure more generally in ways that promote public health.

Sometimes popular resistance to regulation is an obstacle to public health policy. Such norms are partly shaped by advertising and private propaganda, which can be countered by regulation. The elimination of advertising that stimulates unhealthy behaviour and resistance to health-promotion would likely have very positive effects on public health and would most likely be non-intrusive.

Thaler & Sunstein recognize that people may prefer not to choose and that requiring them to choose may therefore be paternalistic (in some sense, presumably not their own). However, they do not consider the possibility that people may more generally prefer certain choice situations, even if they have no settled preference for how they would choose in that situation. There is a tendency in Thaler & Sunstein to assume that liberty is protected as long as people have a chance to opt out of any program. Opt-out freedom is liberty in a very narrow sense. Thaler & Sunstein's focus on promoting welfare should be balanced against other liberty values as well.

7. OVERVIEW OF PAPERS

'The Normative Core of Paternalism' defines anti-paternalism and 'Anti-paternalism and Invalidation of Reasons' rejects it. 'Paternalistic Interference' drives home the point that paternalism is not prima facie wrong. The focus of the first paper is on the complex interconnections between reasons and actions that must be sorted out in order to provide the most generous understanding of anti-paternalism. The second paper builds on this definition to characterize anti-paternalism as an instance of an influence-regulating principle of invalidation of reasons, and three arguments against such principles are presented. In the third paper, five accounts of the action component of paternalism are surveyed and found inadequate as specifications of something that it should be prima facie wrong to do out of benevolence. These three articles are the most general, most abstract and to my mind most fundamental to this thesis. It is no accident that I have most frequently referred to them in this introduction.

'Liberalism, Altruism and Group Consent' is an attempt to investigate how liberalism, in the spirit of anti-paternalism, can deal with regulation of groups where some members consent to regulation and some do not. In contrast to the rest of the thesis, I am here completely devoted to interpreting and refining anti-paternalism, not to criticise it. However, in this process its limits become apparent, including the impossibility of justifying the regulation of groups by partial consent of its members only, with no regard for their good. The proper regulation of divided groups seems to me a central issue in public health ethics, as opposed to traditional medical ethics.

This issue is carried over to the first of three more practically oriented articles, where my theoretical understanding of anti-paternalism is put to use in various policy areas. 'Anti-paternalism and Public Health Policy: The Case of Product Safety Regulation' investigates how anti-paternalism can be interpreted in that specific context, and suggests that the liberal does not need it. 'Responsibility, Paternalism and Alcohol Interlocks', co-authored by Jessica Nihlén Fahlquist, investigates the empirical and economical issues relevant for assessing a policy of mandatory alcohol interlocks in all

cars, as well as considers possible liberal resistance to such a policy based on the idea of individual responsibility as well as on anti-paternalism, finding the objections unconvincing.[13] Finally, 'Epistemic Paternalism in Public Health', co-authored by Sven Ove Hansson, considers the case for withholding of information of uncertain threats to public health – a paternalistic policy designed to protect people from allegedly unnecessary anxiety and depression. Based not on principled anti-paternalism, but on broadly consequentialist considerations, such a policy is found unwarranted and counter-productive.[14]

[13] In writing this paper, I was responsible for the overall structure and coherence of the text. I also wrote most of the section on paternalism, while Jessica wrote most of the section on responsibility. That said, we were both actively involved in all parts of the essay.
[14] While I did most of the writing for this paper, both authors contributed equally to all parts.

REFERENCES

Archard, David. 1990. Paternalism defined. *Analysis* 50(1): 36-42.

Archard, David. 1994. For our own good. *Australasian Journal of Philosophy* 72(3): 283-93.

Arneson, Richard. 1980. Mill versus Paternalism. *Ethics* 90(4): 470-489.

Arneson, Richard. 1989. Paternalism, Utility, and Fairness. *Revue International de Philosophie* 43: 409-437.

Arneson, Richard. 2005. Joel Feinberg and the Justification of Hard Paternalism. *Legal Theory* 11: 259-84.

Baron, Marcia W., Pettit, Philip & Slote, Michael A. 1997. *Three Methods of Ethics: A Debate*. Wiley-Blackwell. Oxford: Blackwell.

Berlin, Isaiah. 2002 (1969). Five Essays on Liberty. In Henry Hardy (ed.) *Liberty*. (Originally published as 'Four Essays on Liberty' with one less essay.) Oxford: Oxford University Press.

Bircher, J. 2005. Towards a dynamic definition of health and disease. *Medicine, Health Care and Philosophy* 8: 335–341.

Boorse, C. 1975. On the Distinction between Disease and Illness. *Philosophy and Public Affairs*, 5(1): 49–68.

Broome, John. 2004. Reasons. In R. Jay Wallace et al. (eds.) *Reasons and Value: Themes from the Moral Philosophy of Joseph Raz*, pp. 28-55. Oxford: Oxford University Press.

Dancy, Jonathan. 1993. *Moral Reasons*. Oxford: Blackwell.

Dancy, Jonathan. 2008. Moral Particularism. *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition) ed. Edward N. Zalta. URL = http://plato.stanford.edu/archives/fall2008/entries/moral-particularism/.

Devlin, Patrick. 1965. *The Enforcement of Morals*. Oxford: Oxford University Press.

De Marneffe, Peter. 2006. Avoiding Paternalism. *Philosophy and Public Affairs* 34(1): 68-94.

Dworkin, Gerald. 1972. Paternalism. *Monist* 56(1): 64-84.

Dworkin, Gerald. 1983. Paternalism: Some Second Thoughts. In Rolf Sartorius (ed.), *Paternalism*, pp. 105-111. Minneapolis: University of Minnesota Press.

Dworkin, Gerald. 1988. *The Theory and Practice of Autonomy*. Cambridge: Cambridge University Press.

Dworkin, Gerald. 2008. Paternalism. In *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition), ed. Edward N. Zalta. URL = http://plato.stanford.edu/archives/fall2008/entries/paternalism/.

Feinberg, Joel. 1971. Legal Paternalism. *Canadian Journal of Philosophy* 1: 105 -24.

Feinberg, Joel. 1984. *Harm to Others*. Oxford: Oxford University Press.

Feinberg, Joel. 1986. *Harm to Self*. Oxford: Oxford University Press.

Gert, Bernard & Culver, Charles M. 1976. Paternalistic Behavior. *Philosophy and Public Affairs* 6(1): 45-57.

Gert, Bernard & Culver, Charles M. 1979. The Justification of Paternalism. *Ethics* 89(2): 199-210.

Hart, H.L.A. 1963. *Law, Liberty and Morality*. Oxford: Oxford University Press.

Häyry, Heta. 1992. Legal Paternalism and Legal Moralism: Devlin, Hart and Ten. *Ratio Juris* 5(2): 191-201.

Hodson, John D. 1977. The Principle of Paternalism. *American Philosophical Quarterly* 14: 61-69.

Husak, Douglas N. 1981. Paternalism and Autonomy. *Philosophy and Public Affairs* 10(1): 27-46.

Husak, Douglas N. 2003. Legal Paternalism. In Hugh LaFollette (ed.), *The Oxford Handbook of Practical Ethics*, pp. 387–412. Oxford: Oxford University Press.

Kleinig, John. 1983. *Paternalism*. Manchester: Manchester University Press.

Korsgaard, C. M. The Right to Lie: Kant on Dealing with Evil. *Philosophy & Public Affairs* 15: 325-349.

Kristjánsson, Kristján. 1996. *Social Freedom: The Responsibility View*. Cambridge: Cambridge University Press.

Law, Ian & Widdows, Heather. Conceptualising Health: Insights from the Capability Approach. *Health Care Analysis* 16(4): 303-314.

Mill, John Stuart. 1991 (1859). On liberty. In *On Liberty and Other Essays*. Oxford: Oxford University Press.

Nikku, Nina. 1997. *Informative Paternalism*. Linköping: Linköping Studies in Arts and Science.

Nozick, Robert. 1974. *Anarchy, State, and Utopia*. Malden MA: Basic Books.

Parfit, Derek. *Climbing the Mountain*. Manuscript.

Pettit, Philip. 1997. *Republicanism: A Theory of Freedom and Government*. Oxford: Oxford University Press.

Pettit, Philip. 2002. Non-Consequentialism and Political Philosophy. In David Schmidtz (ed.), *Robert Nozick*, pp. 83-104. Cambridge: Cambridge University Press.

Rawls, John. 2001. *Justice as Fairness: A Restatement*. Cambridge MA: Harvard University Press.

Raz, Joseph. 1986. *The Morality of Freedom*. Oxford: Oxford University Press.

Raz, Joseph. 1990. *Practical Reason and Norms*, 2d ed. Princeton: Princeton University Press.

Scanlon, Thomas. 1998. *What We Owe to Each Other*. Cambridge MA: Harvard University Press.

Sen, Amartya. 1982. Rights and Agency. *Philosophy and Public Affairs* 11(1): 3-39.

Sen, Amartya. 1992. *Inequality Reexamined*. Cambridge MA: Harvard University Press.

Shiffrin, Seana. 2000. Paternalism, Unconscionability Doctrine, and Accommodation. *Philosophy and Public Affairs* 29(3): 205-250.

Sneddon, Andrew. 2001. What's Wrong With Selling Yourself Into Slavery? Paternalism and Deep Autonomy. *Crítica* 33(98): 97-121.

Thaler, Richard H. & Sunstein, Cass R. 2003a. Libertarian Paternalism. *The American Economic Review* 93: 175-179.

Thaler, Richard H. & Sunstein, Cass R. 2003b. Libertarian Paternalism Is Not an Oxymoron. *The University of Chicago Law Review* 70: 1159-1202.

Ten, C.L. 1971. Paternalism and Morality. *Ratio* 13: 56-66.

Ten, C.L. 1980. *Mill on Liberty*. Oxford: Clarendon Press.

Van de Veer, Donald. 1986. *Paternalistic Interference*. Princeton: Princeton University Press.

Res Publica 13: 441-458

The Normative Core of Paternalism*

Kalle Grill

ABSTRACT: The philosophical debate on paternalism is conducted as if the property of being paternalistic should be attributed to actions. Actions are typically deemed to be paternalistic if they amount to some kind of interference with a person and if the rationale for the action is the good of the person interfered with. This focus on actions obscures the normative issues involved. In particular, it makes it hard to provide an analysis of the traditional liberal resistance to paternalism. Given the fact that actions most often have mixed rationales, it is not clear how we should categorize and evaluate interfering actions for which only part of the rationale is the good of the person. The preferable solution is to attribute the property of being paternalistic not to actions, but to compounds of reasons and actions. The framework of action-reasons provides the tools for distinguishing where exactly paternalism lies in the complex web of reasons and actions.

Keywords: actions, action-reasons, anti-paternalism, harm to others, interference, paternalism, reasons

*Spelling and reference errors in the published *Res Publica* version have been corrected.

INTRODUCTION

The normative core of paternalism is the invocation of the good of a person as a reason for interference with her.[1] In order to clearly distinguish this normative core, we must resist the temptation to define paternalism in terms of actions and instead accept a somewhat more complex analysis. There are two distinct components involved in paternalism: an action component, but also a reason component. The property of being paternalistic should be attributed not to any one of these components, but only to action-reason compounds. Only then can we describe and evaluate the paternalistic content of a situation independently of other aspects of that situation.

This article concerns the conceptual issue of how paternalism should be defined. A methodological premise of the discussion is that we want to define paternalism in a way that will let us evaluate claims about its moral properties. The most common attitude towards paternalism is to reject it, absolutely or conditionally. A normatively useful definition of the concept should therefore accommodate different forms of anti-paternalism. Discussions of the justifiability of paternalism often simply assume that the object of discussion is liberty-limiting or interfering actions (or omissions) that are

[1] Or the invocation of the good of a group of people for interference with them. I will for the most part talk of single persons, though the analysis fits equally well for groups.

supported by one reason only – the good of the person interfered with.[2] Some authors even claim explicitly that only such actions can involve paternalism.[3] In fact, however, actions most often have mixed rationales: they are supported by more than one reason. Interferences are no exceptions. The good of a person can be a greater or smaller part of the rationale for an interference with her; it can be a sufficient reason in and of itself, it can be a necessary part of any sufficient set of reasons, or it can be a non-sufficient but contributory (possibly redundant) reason. I propose that allowing a person's good to count as a valid reason for interference with her is paternalistic regardless of the (relative) strength of that reason.

In order to distinguish the invocation of one particular reason for some action with a mixed rationale, we need a way to talk about the compound of a certain reason for a certain action. I propose that we simply adopt the term 'action-reason' to refer to such compounds. As a definition of paternalism, I propose that only action-reasons can be paternalistic and that an action-reason is paternalistic if and only if the reason is one referring to the good of a person and the action is an interference with the same person.[4] This definition concerns the structure of the concept. More specific conceptions of paternalism, corresponding to different normative views, should define the action and the reason components in greater detail.

Interpreting paternalism in terms of action-reason compounds coheres perfectly with the target of classical, Millian anti-paternalism.[5] According to this doctrine, interference as such may be quite acceptable so far as it is justified by the protection of people from *each other*.[6] Conversely, a person's good may be quite acceptable as a reason

---

[2] A prominent example is Gerald Dworkin, *Paternalism*, in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2002 Edition), http://plato.stanford.edu/archives/win2002/entries/paternalism/. The third condition in Dworkin's analysis of 'X acts paternalistically towards Y by doing (omitting) Z' is: *X does so just because Z will improve the welfare of Y* (emphasis added).

[3] John Gray claims that paternalism can only be the 'genuine moral dilemma as to whether it is proper to coerce an individual solely in his own interest' – Gray, *Mill On Liberty: A Defence* (London: Routledge & Kegan Paul, 1983), p. 90, emphasis added.

[4] Actions can be quite complex, as in the case of such state 'actions' as the formulation, adaptation and implementation of policies; policies that can involve legislation, law enforcement, taxes, information, direct aid and infrastructural adjustments.

[5] J.S. Mill's liberty principle states 'that the sole end for which mankind are warranted, individually or collectively, in interfering with the liberty of action of any of their number, is self-protection. That the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others. His own good, either physical of moral, is not a sufficient warrant'. See Mill, *On Liberty*, in On Liberty and Other Essays (Oxford: Oxford University Press 1991), p. 14. Part of the thesis of the present article is that, interpreted generously, Mill claims not only that the good of a person is not a sufficient warrant, but more generally that it is not an acceptable reason, whether or not it is sufficient. C.L. Ten has interpreted Mill along these lines: 'There are certain reasons for intervention in the conduct of individuals which must always be ruled out as irrelevant' – see his *Mill on Liberty* (Oxford: Clarendon Press 1980), p. 40. Joel Feinberg's professed methodology in *The Moral Limits of the Criminal Law* is to investigate 'what kinds of reason can have weight when balanced against the presumptive case for liberty' – see his *Harm to Others* (Oxford: Oxford University Press 1984), p. 9. On Richard Arneson's interpretation of Feinberg's anti-paternalism, 'Antipaternalism says that harm or risk of harm to a person who voluntarily consents to absorb the harm or stand the risks is never a good reason for criminal prohibition' – see Arneson, 'Joel Feinberg and the Justification of Hard Paternalism', Legal Theory 11 (2005) 259-84, p. 263.

[6] Cf. Mill, *On Liberty*, p. 83: 'As soon as any part of a person's conduct affects prejudicially the interests of others, society has jurisdiction over it'.

for *non-interfering* actions such as giving gifts and offering advice and support.[7] It is the combination of acting for a person's good while interfering with her that is rejected by anti-paternalism.

REASONS AND ACTIONS

The great majority of proposed definitions of paternalism attribute the predicate *paternalistic* to actions.[8] Such action-focused accounts do include a reason component, but only as a qualifier on what actions count as paternalistic. Interferences are usually said to be paternalistic only if they are motivated solely, or mainly, by the good of the person interfered with, or to the extent that they are so motivated. As I will try to show, neither these nor more complex conditions capture the normative core of paternalism.

On the action-reason account here put forth, there are two components of paternalism – the interference with a person, being some kind of action, and the good of the same person, being a reason for that action.[9] The term *interference* is here used generically. I do not wish to claim that this term in and of itself contributes much to our understanding of paternalism. Rather, *interference* functions as a place-holder, to be fleshed out by more detailed conceptions of paternalism.[10] Such conceptions must specify a class of actions, the members of which can pair up with reasons referring to the good of a certain person, to form paternalistic action-reasons. More detailed conceptions of paternalism should be based on substantial normative views about what reasons are invalid or problematic when invoked for what actions, or on attempts to describe such substantial normative views.

In the following survey of action-focused definitions of paternalism, I will try to show that regardless of how interference is fleshed out, actions by themselves cannot be paternalistic, but can only form parts of paternalistic action-reasons. I will not in this section distinguish between e.g. motivational, explanatory and justificatory reasons, but will return to the nature of reasons later on. For expository convenience, I will call reasons for an interference that refer to the good of the person interfered with

---

[7] Cf. Ibid., p. 84: 'Human beings owe to each other help to distinguish the better from the worse, and encouragement to choose the former and avoid the latter.'

[8] These include Gerald Dworkin, 'Paternalism', *The Monist* 56 (1972) 64-84; Bernard Gert and Charles M. Culver, 'Paternalistic behavior', *Philosophy and Public Affairs* 6(1) (1976) 45-57; John Kleinig, *Paternalism* (Manchester: Manchester University Press 1983); Donald Van de Veer, *Paternalistic Intervention* (Princeton: Princeton University Press 1986); David Archard, 'Paternalism Defined', Analysis 50(1) (1990) 36-42; Gerald Dworkin, 'Paternalism', *Stanford Encyclopedia*; Simon Clarke, 'A Definition of Paternalism', *Critical Review of International Social and Political Philosophy* 5(1) (2002) 81-91; and Peter De Marneffe, 'Avoiding Paternalism', *Philosophy and Public Affairs* 34(1) (2006) 68-94.

[9] On some action-focused accounts of paternalism, lack of *consent* is listed as an independent condition on paternalistic actions. However, we may incorporate consent into the action component by assuming that whether and what kind of consent is given affects what counts as interference. This approach has the advantage of allowing for different versions of paternalism to attribute to consent as central or as marginal a role as its proponents would like in specifying interference.

[10] The action component of paternalism traditionally goes by such names as 'interference with choice' (Van de Veer, p. 19), 'imposition' (Kleinig, p. 7), 'violation of autonomy' (Gerald Dworkin, 'Some Second Thoughts', in Rolf Sartorius (ed.), *Paternalism* (Minneapolis: University of Minnesota Press 1983) 105-11, p. 107), 'limiting liberty' (Joel Feinberg, *The Moral Limits of the Criminal Law Vol. 2 Harm to Self* (Oxford: Oxford University Press 1986), p. ix), or 'interference with the liberty of action' (Mill, *On Liberty*, p. 14).

'paternalistic reasons', though I do not intend to suggest that reasons can be paternalistic in themselves any more than actions can. As mentioned in the introduction, a common assumption is that an interference is paternalistic only if it is performed *solely* for paternalistic reasons. This reason-condition is far too narrow. Indeed, on inclusive accounts of what counts as a reason for an action this condition renders the class of paternalistic actions empty. As soon as there is some non-paternalistic reason for an action, the action is not paternalistic. Even actions for which paternalistic reasons by themselves provide a sufficient rationale do not qualify as paternalistic if there are other, redundant, reasons for the action.

Normatively, this is very strange. The mere presence of another reason, such as one referring to harms to others, should not erase the paternalistic content of a situation, especially not where an interference with a person is performed mainly for her own good. The inadequacy of the *solely* condition may suggest that we should relax our requirement and say that an action is paternalistic only if it is performed *mainly* for paternalistic reasons.[11] This condition is, however, also too narrow, in much the same way. Reasons that are not the main reason for an action may still be very important. They may for example be necessary parts of any set of reasons that provides a sufficient rationale for a certain action. The *mainly* condition entails that there is no paternalism involved when an interference is performed mainly for non-paternalistic reasons, even if it would not be motivated, *all things considered*, if it were not for the impact of paternalistic reasons. This does not accord well with the concern of anti-paternalists, who typically think that something has gone wrong when paternalistic reasons are allowed to tip the balance in favour of interference. More generally, the moral question raised by paternalism does not hinge on what reason is the main reason for an action, but on whether the good of a person may contribute to the rationale for an action – on whether this kind of reason is a valid reason for interference at all. It makes no sense to allow that a reason R contributes to the rationale for an action, making it motivated or justified, as long as R is weaker than some other reason, while rejecting the same action if R is (or becomes) stronger.[12]

The *mainly* condition is also too wide, since it fails to appreciate the importance of non-paternalistic reasons. Again, reasons that are not the main reason for an action may be very important. Even if the main reason for an interference is indeed a paternalistic one, other reasons may quite independently offer a sufficient rationale for the action. On the *mainly* condition, if the main reason for an action is a paternalistic reason and if paternalism should be rejected, then the action should consequently be rejected, regardless of what other reasons there are for the action. The non-paternalistic reasons are simply ignored. This is surely not intended by antipaternalists; nor is it reasonable.

Alongside the *mainly* condition, there may be any number of different conditions, demanding that some larger or smaller fraction of the rationale of an interference be paternalistic in order for the interference to be paternalistic. The larger the fraction, the

[11] This is proposed for example by Archard, 'Paternalism Defined', pp. 38-9.

[12] Clarke explicitly renounces the 'solely' and 'mainly' conditions, with the argument that also 'minor' reasons can make an action paternalistic – 'A Definition of Paternalism', p. 2, esp. n.1.

more vulnerable the condition is to the narrowness objection. The smaller the fraction, the more vulnerable it is to the wideness objection. Nearly all fractions face both objections, and no fraction avoids both.

The problems with the *solely* and *mainly* conditions stem from the fact that interferences can have, and often do have, mixed rationales. One attempt to deal with this complexity is to propose with John Kleinig that 'impositions are paternalistic to the extent that they are motivated by consideration of the welfare, etc., of the person imposed upon'.[13] Presumably, this is also how other action-focused accounts that allow that there are cases of *mixed* paternalism should handle the question of what actions count as paternalistic.[14] What, however, is the moral import of an action being paternalistic to some extent? What is it, for example, to reject paternalism on this account? If it is to reject those interferences that are paternalistic to a certain extent, the severe problems faced by the *mainly* condition reappear. Depending on exactly to what extent an action must be paternalistic in order to warrant rejection, anti-paternalism so defined is to varying degrees both too wide and too narrow.

The most generous interpretation of how the *extent* condition could accommodate anti-paternalism is arguably to allow that interferences be rejected to the extent that they are supported by paternalistic reasons. This interpretation presupposes that rejections of actions come in degrees, which may be problematic. Supposedly, one must consequently allow that actions are sometimes partly wrong and so on, thereby complicating any more comprehensive theory of the rightness and wrongness of actions. However, let us for the sake of argument disregard these complications. If nothing else, counter-measures taken towards actions can certainly come in degrees of severity.

That interferences are rejected to the extent that they are supported by paternalistic reasons can be understood either in an absolute or in a relative sense. Either the force of the rejection depends solely on the strength of the support from paternalistic reasons, or it depends on the strength of that support in relation to the strength of other supporting reasons. According to the first interpretation, paternalistic reasons are the only reasons to have any influence on what interferences are rejected. The stronger these reasons are, the stronger the rejection, independently of what non-paternalistic reasons there are for the action and the strength of these reasons. This surely amounts to an unacceptable disregard for non-paternalistic reasons. According to the second interpretation, non-paternalistic reasons may have an impact, but only through their relative strength. What matters is not the strength of the non-paternalistic reasons as such, but only their strength in relation to the strength of the paternalistic reasons. This implies, for example, that an interference supported by strong reasons referring to harm to others and equally strong paternalistic reasons should be rejected more forcefully than an interference supported by weak reasons referring to harm to others and somewhat weaker paternalistic reasons. Such disregard for harm to others is unreasonable. Under

[13] Kleinig, p.12 (emphasis in original). The more formally stated definition of paternalism on p. 13 suggest a strict either or-account of the rationale for an action. The surrounding discussion, however, makes it clear that the quoted passage is more in line with Kleinig's intentions.

[14] E.g. Feinberg, *Harm to Self*, p. 8. That there are such cases is often acknowledged, but the problems they raise for the analysis of paternalism are not.

both interpretations then, the *extent* condition discounts non-paternalistic reasons in a way that is entirely unwarranted.

In sum, all three main attempts to incorporate the reason component of paternalism as a qualifying condition on what actions count as paternalistic fail to make sense of the attributing of a special moral status to paternalism, and, especially, fail to make sense of anti-paternalism. They therefore fail to capture the normative core of paternalism. More complex accounts of paternalistic actions are admittedly possible. Peter de Marneffe considers the possibility of counting an action as paternalistic 'only if it cannot be fully justified unless paternalistic reasons are counted in its favour' and the action 'would be fully justified if paternalistic reasons were allowed to count in its favour'.[15] This account avoids the unfortunate disregard for non-paternalistic reasons. The rejection of all actions that are paternalistic on this account leave us with the same class of justified actions as does the rejection of all paternalistic action-reasons, to be further explored below. However, anti-paternalism will deliver no judgement as to the moral status of paternalistic reasons for interferences that are not justified, all things considered. It follows from the definition that no interference is paternalistic if it would not be fully justified even if paternalistic reasons were accepted as valid. Unjustified interferences that are performed for the good of the person interfered with are thus not paternalistic. This curiosity does not affect the judgement of what actions are justified, but it does affect the judgement of how we should reason. De Marneffe recognises this consequence of his account and proposes as a remedy that we add the further condition that the agent (or some of the agents in the case of government policy) is (are) psychologically motivated by paternalistic reasons. This solution has the unfortunate consequence that no paternalism is involved unless there is both a paternalistic motive and a paternalistic justification, thus excluding from the realm of the paternalistic cases where there is one or the other but not both. The preferable solution is simply to define paternalism not as the performance of certain actions, qualified by complex reason conditions, but rather as the invocation of certain reasons for certain actions.

EMPIRICAL CONSIDERATIONS

The strong tendency to define paternalism in terms of actions is perhaps partly due to the prevalence of empirical arguments in the literature. Mill set the stage for this mode of discussion. His classical arguments against state involvement in the market include state incompetence and corruption, and the inability of society to adhere to individual circumstances.[16] While Mill points out that these arguments are independent of his anti-paternalist liberty principle, his argument for that principle includes more subtle empirical considerations such as every person's greater interest in her own well-being, her greater knowledge of how to improve her own wellbeing, and the tendency of *vigorous and independent characters* to rebel against benevolent interference.[17] In ethics our main interest

[15] de Marneffe, p. 72.
[16] Mill, pp. 122-8.
[17] Ibid., pp. 84-5, 92.

is perhaps the moral status of actions – which actions are, for example, permissible, required or forbidden. This matter is in part decided by the likely consequences of actions, which are determined by empirical circumstances. We want to know, for example, whether a state should prohibit duelling or professional boxing, or the use of tobacco or LSD. Whether it should do so depends to some extent on the likely consequences of prohibition. If, for example, prohibition is counter-productive for some reason, so that it would lead to a surge in the prohibited activity, few would favour prohibition as a matter of principle.

This is all very well, but it has little to do with paternalism. Attempts to promote or protect a person's good that are likely to fail are obviously not very desirable, especially if they entail a cost in terms of liberty. There is no need to invoke the idea of paternalism to make that point. Empirical circumstances that determine the likelihood of successful promotion or protection of a person's good are instrumental to deciding the moral status of an action with that aim. However, quite a separate idea runs through the liberal tradition from Mill through Joel Feinberg and onwards. This idea is that it is somehow *illegitimate* to interfere with a person for her own good. Regardless of whether an interference with a person does in fact promote her good, there is something morally wrong with such interference. This moral wrong may not have the status of an ultimate principle, but neither is it the mere belief that benevolent interference is always counter-productive. Anti-paternalism as a normative doctrine is in this respect independent of empirical circumstances.

One may of course be a thoroughgoing consequentialist, and have no direct concern with normative doctrines beyond the principle of utility. For such consequentialists, paternalism is not really an issue. It does not matter what reasons there are, only how we act. Reasons may however enter the stage with the introduction of rules of thumb, action-guiding rules abiding by which promotes utility in the long run. Such rules of thumb may apply to our mental actions, to our mode of reasoning. They may, for example, tell us how to reason with regard to paternalistic reasons for interference. If they do, they do in fact entail a position on paternalism, though indirectly. The mental action of considering or attributing weight to a certain reason is one kind of invocation of a reason for an action and may properly be described as an action-reason. Positions on paternalism based on consequentialist rules of thumb therefore concern action-reasons rather than (physical) actions, just like positions on paternalism based on less empirically focused normative doctrines.

Though psychologically motivating reasons may perhaps have consequences, justificatory reasons do not. Empirical considerations on the whole, therefore, connect most naturally with actions, rather than with reasons or action-reasons. As shown in the previous section, however, defining paternalism in terms of actions fails to make sense of the normative core of paternalism. We should not let the widespread habit of mixing empirical and normative arguments lure us into accepting a flawed definition of paternalism. The moral problem of paternalism concerns the invocation of paternalistic reasons for interference. The moral status of an interference will in the end depend both

on our (correct) normative position on paternalism and on the consequences of interference.

ACTION-REASONS AND EFFECT-REASONS

Making use of action-reason compounds to describe the interplay between reasons and actions accommodates the fact that actions may be supported by many different reasons and that reasons may support many different actions.[18] Based on normative considerations, we may distinguish certain such (kinds of) compounds and attribute moral properties to them. I have proposed that an action-reason be counted as paternalistic if and only if the action is an interference with a person and the reason refers to the good of the same person. This is a very general definition of paternalism. Both the action and the reason component of paternalism can be further specified. Every paired specification delimits a different class of action-reasons, and so a different conception of paternalism.[19] As for the moral properties of these classes of action-reasons, the most common position to endorse regarding paternalism is to reject it – to hold it to be wrong or illegitimate or forbidden, at least under certain conditions. What is it then to reject an action-reason? Presumably, it is to hold that the reason in question is invalid for the action in question.[20] This is the form of anti-paternalism I will focus on in this section, though other moral properties than this relation of invalidity are certainly possible.[21]

---

[18] It may be that the reasons there are for an action determine what kind of action it is. This does not pose a problem for the action-reason account. Once an action is distinguished, whether by its actual effects or by some standard based on intentionality, it may be paired with different reasons, that are reasons for that action, to form action-reasons. The framework of action-reasons is independent of how exactly actions and reasons are individuated, though theories about individuation may perhaps be informed by this framework.

[19] This account of paternalism makes no direct reference to the attitude of the paternalist. To some extent, being motivated by or accepting as valid, paternalistic reasons for interference may be taken to constitute a paternalistic attitude. However, there is no reference to specific attitudes such as that of superiority or condescension, or the proper attitude of a parent towards her child. This is arguably an advantage of the account, since it is unclear whether paternalism necessarily involves any such attitudes.

[20] We may distinguish between on the one hand the relevance of a reason, determined by whether or not the value that the reason refers to is affected by the choice or action under consideration, and on the other hand the validity of a reason, determined by whether or not the reason should have any weight according to (the correct) normative principles. In order to be a reason for an action, it is enough that the reason is relevant. In order to actually give normative support to the action, the reason must also be valid.

[21] Moderate anti-paternalists may want to discount paternalistic reasons in some fashion, rather than reject them outright. Louis Groarke argues against absolute anti-paternalism and suggests that for any interference the value of *care* should be subtracted from the value of freedom – 'Paternalism and Egregious Harm', Public Affairs Quarterly 16(3) (2002) 203-30. However, he believes that '[p]aternalism would be permissible only in those cases where the net value was largely or perhaps very largely negative' – p. 219, emphasis added. On the other hand, extreme anti-paternalists may suggest that paternalism is so degrading that the fact that an interference protects or promotes the good of a person should not only not count as a valid reason for that interference but should actually count as a reason against it. The fact that (part of) the rationale for an interference is the good of the person interfered with *adds insult to injury*, so to speak (see Kleinig, pp. 70-2, for ideas along these lines). Interfering with a person for selfish purposes could thus be morally better than interfering in the same way out of benevolence. To encompass this view, the rejection of an action-reason may be taken to give rise to an anti-paternalistic reason against the action, with whatever weight necessary to account for the strength of the extreme anti-paternalism.

Importantly, the normative status of the components of an action-reason is independent of the normative status of the compound. In the case of anti-paternalism, what is rejected is neither the interference as such, nor the paternalistic reason, but only the combination of the two. A person's good may be a valid reason for some actions directed towards her, but not for interferences with her. Correspondingly, interferences with a person may be legitimately supported by some reasons, but not by her good.[22]

An example will illustrate the action-reason account: A seizes B's cigarettes in order to prevent B from smoking. This is presumably an interference with B.[23] A's motivating reason for interfering is concern for B's health (perhaps B has emphysema). C observes this incident and, being an anti-paternalist, rejects the action-reason 'seizing B's cigarettes – concern for B's health'. However, C is aware that unless A had seized B's cigarettes, B would later have smoked them in a confined space together with D (perhaps B's child). The action-reason 'seizing B's cigarettes – concern for D's health' is not paternalistic since the person interfered with and the person whose good is invoked are not identical.[24] C may therefore find the action 'seizing B's cigarettes' perfectly in order. C does not necessarily reject A's action, but only A's action-reason.

The analysis becomes somewhat more complicated if we take into account actions that have multiple effects, each of which may be an interference with a different person. It may not be paternalistic to invoke the good of a person A for an action which is an interference with A and with a second person B, if A's good is protected or promoted only through the interference with B and not through the interference with A. Public policy, for example, may interfere with all or many of those affected and may promote the good of all or some through the interference with others. In order to distinguish the paternalistic content of complex situations, we must extend our framework to cover separate effects of actions and allow that a reason for an action may be directed at one (some) of the effects of an action only, and not at others. We may call the invocation of a reason for an action which is directed at one of the effects of the action an 'effect-reason'. Strictly speaking, reasons are of course invoked not for effects but for actions. However, since one and the same reason (such as a person's good) may be directed at several different effects of an action, we must distinguish between a reason as it applies to one effect rather than another. In order to avoid dividing reasons into different aspects or subreasons, we may in the more formal analysis allow that reasons

[22] It follows from this interpretation that anti-paternalism cannot be dismissed with the simple observation that it is all but impossible to identify any action (and especially, perhaps, any state policy) that interferes with certain persons and promotes their good, without affecting the interests of others. This is as it should be.

[23] Whether or not it is an interference depends on how interference is fleshed out as part of a more detailed conception of paternalism. Most such conceptions would consider the seizing of another's property (against her will) an interference.

[24] In general, nothing prevents direct involvement with one person counting as interference with another. It could in some cases be an interference with P to seize Q's cigarettes. More commonly, it may be an interference with P to prevent Q from selling cigarettes to P. Such interference is an example of what Dworkin calls *impure* paternalism and Feinberg a 'two-party-case' – see Dworkin, *Paternalism*, Monist, p. 68; Feinberg, *Harm to Self*, ch. 22, e.g. p. 172.

are reasons for effects rather than actions.[25] In practice, the distinctions are quite intuitive.

An effect-reason is paternalistic if and only if the reason is one referring to the good of a person and the effect is an interference with the same person. This is intended as a specification or extension of the previously given definition of paternalism, rather than an alteration. The compound of an interference with a person and a reason referring to the same person's good is paternalistic, whether the interference is an action or an effect. Effects of actions may in this framework be individuated on the basis of normative concerns. In the case of paternalism we want to divide the total effect of an action into parts according to how the action interferes with different people. If possible, an action that interferes with several persons should be divided into one effect per person interfered with. It then straightforwardly follows where and how paternalism is involved in the complex web of actions, reasons and effects.

Consider the action of preventing A and B from fighting each other. Keeping A and B apart (or threatening them with punishment if they fight) will have the double effect of both preventing A from fighting B and preventing B from fighting A. Assume that the first effect amounts to an interference with A, and the second to an interference with B. The reasons for action in this kind of situation are typically A's and B's good. Both reasons may be invoked for both effects of the action. The effect-reasons 'preventing A from fighting B for A's good' and 'preventing B from fighting A for B's good' are paternalistic. It is paternalistic to invoke a person's good as a reason for preventing her fighting someone (or so we have assumed). Most typically, however, the main reason for preventing a fight is to avoid people being *fought with*. Now, if A and B do not want to be fought with (though they may perhaps want to fight back if attacked), neither 'preventing A from fighting B – B's good' nor 'preventing B from fighting A – A's good' are paternalistic. It may be, however, that A and B both want to have this fight, so that (let us assume) 'preventing A from fighting B' would be an interference not only with A but also with B, and 'preventing B from fighting A' would be an interference not only with B but also with A. Then both these effect-reasons are paternalistic, and the fight may only be prevented if paternalism is allowed (or if prevention is supported by other reasons than the good of A and B). Finally, it may be the case that A welcomes the fight while B does not. Then 'preventing B from fighting A – A's good' is paternalistic, while 'preventing A from fighting B – B's good' is not (and it may be that the fight can be prevented without paternalism, since preventing A from fighting B will prevent the fight from occurring).

Similarly, in the case of public policy, actions that amount to interferences with a number of people and are supported by the good of the same people may be given quite varied analyses depending on their finer structure. A policy involves little or no paternalism if no important reason for any interference effect refers to the good of the person interfered with. This is typically the case for standard criminal law such as the

[25] An alternative would be to introduce subreasons and 'effect-subreasons', which would be paternalistic if and only if the sub reason referred to the good of a person and the effect was an interference with the same person.

prohibition on theft and assault. Such policies protect the good of each person not through interference with her, but only through interference with others. Reasons for such policies that refer to the interests of the thief or the assailant may be relevant, but are negligible compared to the interests of others not to be robbed or assaulted. A policy involves a lot of paternalism, on the other hand, if an important reason for every interference effect is the good of the person interfered with. This might be the case for what we ordinarily think of as 'paternalistic policies', such as safety regulations and prohibitions of dangerous activities. It is for such policies that the framework of effect-reasons could do important work by distinguishing the paternalistic content of policy-making. The crucial question is to what degree each person's good is promoted or protected through interference *with her*, and to what degree the good of each person is rather promoted or protected through interferences *with others*. In other words – what reasons may be invoked for what effects?

Conceptually, the action-reason (effect-reason) account may be too complex to be in tune with everyday use of the term 'paternalistic'. For terminological convenience and out of consideration for common usage, actions may therefore be called paternalistic, in a derived sense, if they form a part of a paternalistic action-reason. In this matter, investigations of the *mainly*, *solely* and *extent* conditions may inform our terminology. Actions could perhaps be called paternalistic, in this derived sense, if their belonging to a paternalistic action-reason is a significant enough property of the action, or to the extent that it is significant. Importantly, these terminological choices have no bearing on the moral status of paternalism.

ACTUAL, BELIEVED AND INTENDED EFFECTS

The traditional focus on actions leads to problems not only with specifying in what way reasons make actions paternalistic (as discussed above), but also with what kind of reasons should be singled out as qualifiers of paternalistic actions. Though Joel Feinberg professes himself concerned with reasons for action and their legitimacy or quality, he is nevertheless absorbed by the search for the proper reason-qualifier for paternalistic actions. After noting that interferences (prohibitions) may be supported and opposed by several different reasons, in the introduction to Harm to Self, Feinberg goes on to discuss, at length, when actions and policies are properly called paternalistic.[26] He distinguishes between four kinds of reasons: 'conscious reasons', 'deep motivations', 'implicit rationales ', and 'true justifications'.[27] In the end, the deciding factor seems to be the 'implicit rationale' for an action, being the main or *true* reason for the action.[28]

Action-focused definitions of paternalism must tell us what kinds of reasons qualify actions as paternalistic. Is an interference paternalistic if it is psychologically motivated by the good of the person interfered with, or is the question rather whether or not the most reasonable justification for the action refers to the good of the person? Do

[26] Feinberg, Harm to Self, pp. 16-23.
[27] Ibid., p. 16.
[28] Ibid., p. 17.

the officially stated reasons have any impact on what actions count as paternalistic? On the action-reason account, there is no need to single out some (kinds of) reasons as more 'true' than others. The invocation of different reasons may be attributed different moral properties, independently of what other reasons there are. This account, therefore, can accommodate varied and complex moral positions and principles which cannot be accommodated by action-focused accounts.

A reason for an interference may, most saliently, refer to believed, intended or actual promotion or protection of good. The action component may similarly refer to actual, believed or intended interference. Compounds of paternalistic reasons and interferences, each either actual, intended, or merely believed, are paternalistic in different ways and concern different normative questions.[29] Beliefs and intentions are important primarily for matters of responsibility and blame, while actual effects are our prime interest in determining the desirability of different options.

Different kinds of paternalism can now be distinguished. We may for example look at intended and actual interferences that are meant to promote or protect the good of the person interfered with, but that in fact fail to do so. These interferences are obviously undesirable, but perhaps they are also especially immoral, at least according to some anti-paternalists. We may further consider whether intended and actual interferences that are not intended, nor believed, to promote or protect the good of the person interfered with, should be evaluated any differently than the first category. Or we may inquire as to the moral status of merely believed or intended interferences that do in fact, or are merely believed or intended to, promote or protect the good of the person believed or intended to be interfered with. However, the most important questions arguably concern actual interferences that actually promote or protect the good of the person interfered with, regardless of beliefs and intentions. Do the corresponding action-reasons or effect-reasons have a special moral status? Are they illegitimate somehow? Is there something that stops actual good-promotion or good-protection from generating valid reasons for action when it coincides with interference? These are questions I have not tried to answer. I have merely tried to defend an analysis of paternalism that allows them to be clearly stated.

CONCLUSION

Most accounts of paternalism assume that the entities that are potentially paternalistic are interfering actions. Such action-focused accounts do recognize that there is a reason component to paternalism – actions are said to be paternalistic only when performed for the good of the person interfered with. The reason component is thereby incorporated as a qualifier on what actions count as paternalistic. However, all such attempts fail to capture the normative core of paternalism, which is the invocation of reasons referring to

[29] On most action-focused accounts, the good-promotion or -protection is taken to be believed rather than actual. As for the action component, some authors focus on actual interference (e.g. Dworkin 'Paternalism', *Stanford Encyclopedia*), while others place the interference as well as the protection or promotion of good entirely in the head of the agent (e.g. Gert and Culver, pp. 49-50).

the good of a person for interference with her. Most importantly, action-focused accounts fail to make sense of the most common attitudes towards paternalism – anti-paternalism of various strands.

The failure properly to accommodate the reason component undermines all normative discussion of paternalism. Moral positions and principles cannot be properly formulated when the basic analysis of the concept prevents us from distinguishing between different reasons for the same action and attributing different moral properties to the invocation of these different reasons. We therefore need to forego the simplicity of action-focused accounts and allow that paternalism resides not in actions, but in reasons for action – action-reasons. In fact, we often need to complicate the analysis further and allow that effect-reasons lie at the heart of paternalism.

If we accept this analysis, we are in a better position to describe normative positions on paternalism, and to discuss their merits. The rather technical language of action-reasons offers a tool to help capture and explicate normative positions that are implicit in the liberal tradition. The account does not determine what counts as an interference or what counts as a paternalistic reason, nor does it determine what is the appropriate attitude towards different compounds of these two components of paternalism. It merely offers a framework in which these components and compounds can be given their proper place. Thus the road is paved for an important normative discussion of what reasons are valid for what actions.

ACKNOWLEDGEMENTS

Papers from which this article has developed were presented at the 2006 Society for Applied Philosophy conference on the Philosophy of Public Health and the 2006 Brave New World graduate student conference, both in Manchester, UK. For constructive comments, I am grateful to the participants of these discussions, to my fellow ethicists at the Royal Institute of Technology in Stockholm, to Torbjörn Tännsjö, and to an anonymous referee for this journal. A special thanks for inspiration and encouragement, as well as valuable comments, goes to my colleagues Lars Lindblom and Niklas Möller.

Anti-paternalism and Invalidation of Reasons

Kalle Grill

ABSTRACT: Anti-paternalism can fruitfully be interpreted as a principle of invalidation of reasons. That a reason for an action is invalid means that the reason is blocked from influencing the moral status of the action. More specifically, anti-paternalism blocks personal good reasons from influencing the moral status of certain interfering actions. Actions are only interfering in this sense if they target choice or action that is sufficiently voluntary. Anti-paternalism so interpreted is unreasonable on three grounds. First, it essentially entails that the degree to which a person acts voluntarily determines whether or not her good provides reasons for action. This leads to wrong answers to moral questions. Second, anti-paternalism entails peculiar jumps in justifiability at the threshold of voluntary enough. Third, anti-paternalism imposes a distinction in kind between the value of respecting choices that are sufficiently voluntary and choices that are not. This distinction is untenable and diverts our attention away from the relative strength of reasons. Invalidation in general is unreasonable on the same grounds.

1. Introduction

I have argued elsewhere that paternalism should be understood in terms of the invocation of certain reasons for certain actions, typically the invocation of a person's good for actions that limit her liberty (Grill 2007). On this understanding, the most salient form of absolute anti-paternalism is the principle that reasons that concern a person's good do not contribute to the justification of actions that limit her liberty. Such reasons may in some sense be relevant, but they are *invalid*. Moderate versions are possible where the strength of personal good reasons is discounted in some fashion when they are reasons for limiting liberty, or where exceptions are made for certain personal goods.

In this article, I will argue that absolute anti-paternalism is unreasonable if understood as a normative principle. I will briefly consider moderate versions and to what extent they are more reasonable. By a normative principle I mean a principle that is not motivated by contingent empirical circumstances such as the level of corruption or incompetence in certain governments, or the capacity of certain people to efficiently pursue their good. While normative and empirical arguments are often intertwined in the literature on paternalism, I find it fruitful to keep them apart.

Throughout I discuss moral reasons. Reasons that are invalidated as moral reasons may possibly remain valid as reasons of some other sort (e.g. prudential). I make no distinction between protection and promotion of good – that is between preventing harm and providing benefit. If the reader thinks that such a distinction is warranted, she may read my argument as concerning harm-prevention.

The second section of this article gives an account of invalidation of reasons and describes how anti-paternalism can reasonably be understood as a principle of invalidation. The bulk of the article is section three, exploring three arguments against anti-paternalism so understood. Section four discusses some strategies for mitigating the problems exposed. Section five sums it up.

2. ANTI-PATERNALISM AS INVALIDATION

The main representative for the kind of anti-paternalism I oppose is Joel Feinberg (1984; 1986), building on John Stuart Mill (1991 [1859]). Though Feinberg claims to be writing an 'essay in applied moral philosophy' (1984, p. 4), his main arguments against paternalism are founded on ideals of autonomy (1986, chapter 18) and personal sovereignty (1986, chapter 19), not as useful heuristic concepts, but as a deep moral values. He explicitly addresses his anti-paternalism to 'an ideal legislator' and claims to be on 'a quest not for useful policies but for valid principles.' (1984, p. 4) More generally, anti-paternalism of various sorts is typically defended with reference to liberal values. If not autonomy or sovereignty, these values may be for example Donald Van de Veer's 'the right to direct one's own life' (1986, section 2.4), John Kleinig's 'the ideal of human personality' (1983, pp. 24-27) or John Stuart Mill's 'liberty' or 'individuality' (1991 [1859], chapter III). Personal good reasons are on the other hand provided by such values as health, prosperity, achievement or well-being.[1]

There is no lack of empirical arguments against paternalism. Two classical ones from Mill (1991 [1859], pp. 84-85) focus on the target of paternalism – people are generally more interested than others in their own good and generally know better how to promote their own good. Other arguments focus on the paternalist – state incompetence and corruption are reasons against legal paternalism specifically. Empirical arguments such as these vary with circumstances: Some states are more corrupt than others, some people are better at promoting their good than others. It is overly optimistic to assume that people are always the best judges of what promotes their own good, and even more so to assume that limiting their liberty can never help to promote that good (unless true by definition, a possibility that will be considered at the end of section three). In some circumstances, of course, anti-paternalism may be a sensible rule of thumb. Regardless, I will put empirical arguments to one side and focus on the deeper normative aspect of anti-paternalism.

Naturally, we can be mistaken about reasons and there are practical constraints to our ability to consider reasons, including limited time, information and processing capacity. However, there is a theoretical, normative level of evaluation on which we can consider for any actual or hypothetical action whether or not it is justified, what reasons there are for and against it. On this level, practical constraints do not restrict what

[1] Seana Shiffrin (2000) argues against something that she calls paternalism but that need not involve personal good reasons. Exactly what reasons paternalism can or must involve according to Shiffrin is not clear, but her examples show that it can for instance involve reasons that concern the good of other people. My argument can be modified to accommodate such reasons.

reasons are eligible. Concrete situations that approach this theoretical level include leisurely discussions of important future alternative actions, both more private and more political. Once the time of choice is upon us, practical constraints may motivate limiting the scope of reasons to consider and may excuse or even justify mistakes and suboptimal or imperfect choices or outcomes. However, when the moral status of an action is discussed in the abstract, the fact that the actual decision situation contains (contained, will contain) practical constraints does not determine what reasons should be considered. On this level, some reasons concern things that we agree are important – that have some value, no mistake or confusion about it. Such reasons cannot be rejected on practical or empirical grounds.[2] They can, however, be rejected on normative grounds, they can be deemed *invalid*.

An interesting defence of invalidation must bear on reasons that concern things of value. To have value is, I will assume, a very basic property, approximately to be worthy of consideration ceteris paribus, regardless of what other considerations there are.[3] If we simply disagree about what has value in this basic sense, we need not invoke the idea of invalidation. However, philosophers subscribing to different moral theories might well agree on what has value. For example, non-consequentialists may agree with hedonistic utilitarians that preventing harm is important in a basic sense that preventing sneezing is not, even if they hold that this consideration is sometimes excluded from influencing what we ought to do. So for example, Thomas Scanlon (1998) in his critique of a morality based on well-being readily admits: 'It would be absurd to deny that well-being is important' (p. 141).

I will postulate that reasons that concern things of value have strength and so are pro tanto reasons for action. That a reason *concerns* some thing means, I take it, that this thing is somehow (expected to be) affected by the action and that the strength of the reason is based on this effect. I do not mean to exclude agent-relative values, but rather take the effects of an action to include that action being performed by that agent. I will not discuss what factors other than the value of the action and its (expected) outcome determine the precise strength of a reason (perhaps most obviously, important factors include the likelihood that the action will be effective and the outcome of not performing the action). For simplicity, I will accept the popular position that only facts, not mere beliefs, provide reasons for action.[4]

---

[2] Perhaps rejecting such reasons can also be the solution to coordination problems such as prisoners' dilemmas, when doing what we have most reason to do leads to suboptimal outcomes. Joseph Raz' (1990 [1975]) account of 'exclusionary reasons' is based on such coordination problems and on practical constraints, not on invalidation. I have chosen not to use the terminology of exclusionary reasons partly to mark this distinction.

[3] Basic as in simple, pre-theoretical, not as in metaphysically fundamental. I will remain neutral concerning what are the bearers of value, while expressing myself as if value can be had by both events/states/facts such as that a person's good is promoted, and by a person's good as an abstract, generic object.

[4] The popular position is defended by Derek Parfit in the manuscript to *Climbing the Mountain* (chapter 1 section 1) and by Raz: 'Only reasons understood as facts are normatively significant' (1990 [1975], p. 18). The main argument put forward by both Parfit and Raz is that we can have reason without belief. That argument of course establishes only that *not only* beliefs provide reasons, while leaving it an open question whether beliefs *can* provide reasons.

In order to make conceptual room for invalidation, I distinguish the *strength* of a reason from its *influence* in determining the moral status of an action (forbidden, permissible, obligatory etc.) and so what we ought to do (not do, may do etc.).[5] Invalidation is the blocking of a reason from influencing the moral status of an action according to its strength. To influence an action's moral status is, I take it, to have weight on the scales, to figure among the factors that should be considered when forming an all things considered judgement (under ideal conditions).[6] A reason may have influence in this sense even if it is ultimately overridden by other reasons. We could say with Broome (2004) that the reason figures in a 'weighing explanation' of an ought fact. However, while Broome says that deontic side constraints replace weighing explanations (making them merely potential), I prefer to say that deontic principles regulate the weighing, for example by making certain reasons invalid, banishing them from the scales.

Many, including Scanlon, prefer not to call reasons without influence 'reasons'. However, opting for this terminology inflates the moral distinction between such things as preventing pain and preventing sneezing. Scanlon (1998) claims that '"being moral" involves seeing certain considerations as providing no justification for action in some situations even though they involve elements which, in other contexts, would be relevant.' (p. 156) Similarly, Frances Kamm (2006) claims that a moral right 'excludes our considering certain other factors that would ordinarily be reasons.' (p. 237) In other words, Scanlon and Kamm claims that, oftentimes, neither the fact that an action will prevent pain nor that it will prevent sneezing is a reason for action – both facts are irrelevant to moral decision and evaluation. To explain the distinction between the two facts, Scanlon would presumably say that preventing pain is a consideration that involves elements which, in other contexts, would be relevant. Kamm would say that preventing pain is a factor that would ordinarily be a reason. I find this terminology biased, cumbersome and confusing. Better to reserve the terms relevance and reason for things that are important, that have value, and say that not all reasons have influence. I find that this terminology clarifies the distinction between consequentialism and non-consequentialism in this area. Regardless, this is the terminology I will use.

Doctrines or principles of invalidation can be placed within a larger family of influence-regulating principles – principles that strike a wedge between the strength of a reason and its influence. A more commonly recognized form of influence-regulation is that of side constraints or absolute reasons. An absolute reason entails that an action

[5] I aim to follow the standard use of 'strength' (sometimes 'weight') employed by e.g. Raz' (1990 [1975]). Joshua Gert (2007) has argued convincingly that reasons have two distinct dimensions of strength – requiring strength and justifying strength. If this is correct, there would correspondingly be two kinds of influence. I will disregard this complication since my argument holds for both kinds of reasons. I will however avoid talk of the 'balance of reasons' and stay with the more general 'determining the moral status of actions'. When I say that one reason is stronger than another, this should always be understood as concerning the same kind of strength, typically requiring strength.

[6] The picture of reasons on the scales is of course metaphorical. I admit that I do not have a theory of how exactly to derive an all things considered judgement from a set of reasons with influence. I can only say with Broome (2004) that such a judgement should be based only on consideration of the strengths of the relevant reasons (or more Broomian – the aggregated weights of the reasons determine whether or not you ought to perform a certain action) (p. 37-38).

should or should not be done, irrespective of other reasons.[7] This a priori-like property of a reason entails that other reasons have no influence – their strength need not be considered. Reasons that invalidate other reasons are distinct from and weaker than absolute reasons since they entail only that some reason(s) have no influence, leaving other reasons to potentially override the invalidating reason. In fact, absolute reasons can be defined as reasons that invalidate all other reasons.

Another form of influence-regulation is that of a lexical ordering of types of reasons. Such a hierarchy allows reasons to be absolute in relation to reasons lower in the hierarchy while not being absolute in relation to reason on the same or higher levels. While a single case of invalidation can be explained in terms of lexical priority, this framework cannot accommodate the properties of invalidation more generally. To see this, assume that reasons of type H are invalidated by reasons of type L. Now there may be other reasons, say of type O, which can override reasons of type L, but which can in turn be overridden by reasons of type H. This is exactly what anti-paternalists typically claim. H reasons (to prevent harm to a person) are invalidated by L reasons (to respect the liberty of the same person), while O reasons (to prevent harm to another person) can override L reasons, according to the harm principle. Further, H reasons can override O reasons since it can be right to prevent serious harm to one person instead of preventing less serious harm to another (when there are no liberty reasons either way). Lexical hierarchy therefore cannot explain invalidation. Instead, a lexically prior reason can be defined as a reason invalidating all reasons on lower levels (tough this is not the only possible understanding of lexically prior).

Feinberg states at one point that the anti-paternalist 'must argue that paternalistic reasons [...] are morally illegitimate or *invalid* reasons' (1986, pp. 25-26, emphasis added). My account of invalidation is an attempt to explicate this important aspect of the anti-paternalist position. A doctrine of invalidation should ideally be specified to a class of actions and a class of reasons. The doctrine then says that reasons in the relevant class do not influence the moral status of actions in the relevant class. This might have the further implication that being motivated by such a reason to perform such an action, or accepting it as justification for such an action, is inappropriate or condemnable in a way that warrants disapproval or punishment, perhaps because it manifests a bad attitude of some sort. However, the more basic idea is the normative invalidity of the reason. Whether failure to see this invalidity is immoral and if so what is the appropriate response to such immorality – these are secondary questions.

3. AGAINST ANTI-PATERNALISM

My scepticism of invalidation in general and of normative anti-paternalism in particular is at heart rather straight-forward: We may not want to reduce morality to a concern for well-being or welfare or interests, but surely these things, in one interpretation or other,

[7] The term from e.g. Raz 1990 (1975), p. 27. Such reasons are sometimes called 'decisive', but 'absolute' is preferable because 'decisive' other times refer to what Raz calls 'conclusive' reasons – reasons that are not overridden in a certain case (though they can be).

are important. Why, in the absence of practical constraints, should reasons, which by their nature concern things of value, nonetheless lack influence? Naturally, reasons may be overridden by other reasons, sometimes by much stronger reasons, but this does not mean that they are invalid. Since we can attribute whatever value we find reasonable to such things as the promotion of values, the fulfilment of duties, the non-violation of rights, and the development of virtues, invalidation has no role to play in determining the moral status of actions. On the contrary, the double power of certain reasons to both influence the moral status of actions with their strength, and to generate influence-regulating doctrines blocking the influence of other reasons, obfuscates the very value of that which these reasons concern. Focusing now on anti-paternalism, I will specify this scepticism into three distinct but related arguments against the doctrine.

First argument – voluntariness and (yet) disaster

Anti-paternalism is the influence-regulating principle that personal good reasons are invalid for a certain class of actions. Let us call these actions *problematic interferences*. Inclusion in the class is determined by two main criteria. On most accounts, the action must be liberty-limiting or interfering in some sense. Anti-paternalism typically does not entail that personal good reasons are invalid for innocuous involvement in other people's life, such as greeting them in the street or giving them small gifts.[8] Specifying this criterion of problematic interferences is difficult. However, I will not discuss these difficulties here.

I shall focus instead on what is arguably an inescapable criterion of problematic interference, namely what Feinberg calls *voluntariness* and has explored thoroughly (1986, chapter 20). Anti-paternalism protects only choices or actions that are sufficiently voluntary. It is not problematic interference to restrain people in rage or panic, people heavily influenced by drugs, or people with severe mental disorders. Or rather, restraining these people may amount to problematic interference if the action conflicts with their voluntary choices made previously, but not simply because it obstructs their current choice or action. Perhaps even more obviously, it is not problematic interference to benevolently restrain very small children.

As is widely agreed among anti-paternalists, voluntariness cannot be determined exclusively by the reasonableness of choices or actions.[9] In order to be protected by anti-paternalism a person need not make correct judgements, but only judgements that are sufficiently voluntary. For Feinberg, a person acts perfectly voluntarily if she is competent, there is no coercion or duress, no subtle manipulation, no ignorance or

[8] Sunstein and Thaler (2003) argues for what they call 'libertarian paternalism', postulating that 'a policy therefore counts as "paternalistic" if it attempts to influence the choices of affected parties in a way that will make choosers better off.' (p. 1162) A (private) policy of greeting people in the street because it makes them happy would thus be paternalistic. Some people may possibly oppose even such unproblematic benevolence. I will focus on the more reasonable because more restricted form of anti-paternalism that opposes only liberty-limiting benevolence, though the form of my argument is not affected by how exactly this notion is defined.

[9] Feinberg (1986) makes this point repeatedly, especially in chapter 20. For example: 'If we are to take extreme unreasonableness as the product of impairment, we must not do so *a priori*, or by definition.' (p. 133)

mistake and no distorting circumstances (such as excitement, strong emotion, or time pressure) (1986, p. 115). Of course, hardly any choice is perfectly voluntary. Feinberg explicitly proposes that a person should be protected from paternalism if her actions or choices are 'voluntary enough' (chapter 20). Mill similarly makes exceptions to his anti-paternalism when a person is 'a child, or delirious, or in some state of excitement or absorption incompatible with the full use of the reflecting faculty' (1991 [1859], p. 107).

My argument is independent of how exactly voluntariness is specified. I will assume more narrowly that *informedness* and *capacity* are two important factors.[10] I take capacity to include the ability to make judgements and act on them. In addition, it may or may not include the possibility to actually do so. The term capacity is meant to allow that persons who can make rational or reasonable choices may choose not to employ that capacity. I thereby wish to leave room for the position that voluntary choices need not to be at all rational or reasonable as long as they are in line with the person's character, or if they are actively or passively accepted by the person.[11]

Now my first argument is simply that it is unreasonable that the degree to which a person acts voluntarily shall determine whether or not her good counts in determining how we ought to treat her. It is unreasonable, for example, that if we can stop a person from harming herself, perhaps seriously, the fact that she is relatively informed and rational shall entail that her health and well-being has no bearing on whether or not we ought to stop her. I propose that if we ought not to stop such a person from harming herself, that is because there are more important things at stake than her health and well-being, not because reasons that concern these things are invalid. Voluntariness often has an indirect effect on the moral status of actions, of course, since a high degree of voluntariness normally makes for good decisions, which normally makes for good consequences. When things are not normal, however, and a voluntary choice leads to catastrophic consequences, the high degree of voluntariness is not a good reason to disregard these consequences.

If we accept invalidation to kick in at a certain degree of voluntariness, we end up giving the wrong answers to moral questions. Consider: (In this and coming cases, I am assuming as part of the hypothetical example what reasons there are and what strength they have. I take it for granted that the specifics of the case can be described so that these assumptions are reasonable. If the reader finds that this or some coming case fits no reasonably realistic situation, then we have identified a difference in our basic value ascriptions, rather than in our positions on influence-regulation.)

> **The bridge**. A person tries to walk over a bridge but we have a chance to stop her. We know that the person is relatively well informed about the general condition of the bridge, is acting in character, is calm and collected, attentive, mature and intelligent, under no duress or pressure, etc. Stopping her would interfere with her

[10] Van de Veer (1986) argues that anti-paternalism should not protect a person who would 'validly consent' if she were 'aware of the relevant circumstances' and if her 'normal capacities for deliberation and choice were not impaired' (p. 88).

[11] Richard Arneson (1980) argues that we should not restrain people just because they have a habit, or liking, of acting spontaneously, or even out of character.

> liberty to move around freely, which is a strong reason against doing so. However, having performed an extensive study of the bridge's durability, we know that, appearances to the contrary, the bridge is unsafe. Stopping the person would therefore most probably prevent her from falling to certain death, which is a much stronger reason for doing so.

Anti-paternalism seems to imply that, in the bridge, the reason to stop the person is invalid and so we should not stop her (unless there are other reasons to do so). That is the wrong conclusion – it is morally unreasonable. We should stop the person, because otherwise she will most probably die.

To accommodate this strong intuition without giving up their doctrine, anti-paternalists must claim that stopping the person does not in fact amount to problematic interference. I will assume that, in this case and in the coming cases, there is no time to argue and so stopping the person amounts to physically restraining her or threatening her with punishment or some such clearly liberty-limiting action. If stopping is not problematic interference, therefore, this must be because the degree of voluntariness is too low.

Mill argues in his original bridge case, from which the bridge is adapted, that the person may be coercively turned back 'without any real infringement of his liberty; for liberty consist in doing what one desires, and he does not desire to fall into the river.' Let us assume that it is true that the person does not want to fall into the river. Mill's reasoning has been thoroughly criticized by J.P. Day (1970). Its main fault lies in the false premise that '[i]f *Y* is a consequence of *X* and *A* does not want *Y*, then *A* does not want *X*.' (p. 181) Though the person does not desire to fall into the river, she does desire to walk over the bridge. The anti-paternalist could argue that even so, this desire is not sufficiently voluntary. However, we know that the person scores high on the standard aspects of voluntariness. She just happens to be wrong, as very able people sometimes are. Perhaps she lacks access to the best information, perhaps she did not bother enough to evaluate this information, or perhaps she miscalculated and so reached the wrong conclusion.

Very small imperfections in people's informedness or capacity can make for disastrous decisions. If the consequences of an imperfect decision are bad enough and if these consequences can be avoided without losing anything of comparable value, they should be avoided. We have the possibility to save the person in the bridge because, in this particular case, we understand the likely consequences of her actions better than she does. To avoid giving the wrong answer in the bridge and similar cases, anti-paternalism must require that, by definition, no one with a possibility to intervene can understand the consequences of a sufficiently voluntary action better than the agent. More generally, it must be required that the outcome of sufficiently voluntary actions can never be improved by intervention by someone other than the agent. But this is just to say that it is reasonable to claim that reasons are invalid only when they have no strength and so are in fact not reasons. Unless we have an advantaged position vis-à-vis the person, we have

no reason to stop her and so there is no reason that can be invalidated. Anti-paternalism, therefore, is either redundant or morally unreasonable.

A popular moderation of anti-paternalism is to make exceptions for personal good reasons that concern autonomy (e.g. Dworkin 1972; Kleinig 1983). On such a moderation, anti-paternalists could defend stopping in the bridge as autonomy-preserving (if they reject the argument that life and death cases like the bridge do not concern autonomy since autonomy is important only if someone is around to have it). Most any harm can diminish autonomy in some way to some extent and so this moderation can entail that some personal good reasons are always valid, making for a substantially weakened doctrine. However, given a more narrow conception of autonomy, cases can be constructed that are analogous to the bridge but where the harm does not diminish autonomy (financial ruin or disfigurement or horrific but temporary pain may be substituted for falling to certain death).

Another kind of moderation is to accept all personal good reasons as partially invalid, though discounted by some factor (e.g. Groarke 2002).[12] Depending on the discount factor, it may be that such versions entail that we should stop the person in the bridge. However, there will be similar cases, where the difference between the relative strength of the reasons is smaller, where such versions will entail the morally wrong answer. Of course, the larger the factor (the smaller the discounting) the fewer such cases, and the less convincing. The most convincing cases must be constructed to match the exact discount factor. I see no way to construct a generic case for all versions. The larger the factor, the weaker the doctrine, and so, from my perspective, the more reasonable. However, the basic argument above applies also to any moderate version – whether or not we should prevent harm to a person does not depend on her level of voluntariness, but on what is at stake.[13]

Perfectly voluntary action and many reasons

In the bridge, it is very likely that stopping the person promotes her good according to her own lights, though she fails to see that because of her mistake, whatever it is. Most probably, she will be happy that we stopped her and so stopping her does not go against her hypothetical, enlightened self-interest. In this sense, Mill does have a point. In other cases we should stop people even if they predict and understand perfectly the consequences of their actions. This may be because they misjudge or do not know to what extent those consequences contribute to their good. It may even be that, though people score high on voluntariness, they have a mistaken view of what their own good consists in. Or so I believe. However, since this is (surprisingly) controversial, I will not press the issue. My argument is independent of what is the correct theory of the good. Even if a person acts perfectly voluntarily, *and* we accept her own view of her good at face value, we should sometimes stop her. Consider:

[12] Groarke proposes that we should, in considering interference with an act, 'add together the freedom value (+something) and the care value (-something) to determine the net value resulting from such an act. Paternalism would be permissible only in those cases where the net value was largely or perhaps very largely negative.' (p. 219)

[13] I take Feinberg to be an absolutist in both regards.

> **The stunt**: A person tries to perform a spectacular stunt but we have a chance to stop her. We know that the person is acting perfectly voluntarily. Stopping her would interfere with her liberty and rob her of a life-defining experience, which is a strong reason against doing so. Stopping her would also eliminate a small but real risk of harm to an innocent and non-consenting passer-by, which is a strong but somewhat weaker reason for doing so. In addition, stopping her would eliminate a small but real risk of harm to herself, which is a relatively weak but non-negligible reason for doing so. Together, the two reasons for stopping her are slightly stronger than the one reason against.

The stunt illustrates the important and oft-forgotten fact that there are normally several reasons for and against any particular action. That some of these reasons are strong in no way excludes the possibility that a weak reason is decisive (tips the balance).

Anti-paternalism implies that we should not stop the stunt, since the personal good reason is invalid (unless there are other reasons to do so). Is this reasonable? –It is not. We should stop the person because of the risks to herself and the passer-by. This is less obvious than in the bridge, because our reasons to do so are only slightly stronger than those against. In particular, the personal good reason is rather weak and so blocking it does not seem so unreasonable. However, this could be adjusted by increasing the risk in the hypothetical case. The personal good reason is set to be weak only to make the point that the validity of weak reasons may be important and that their invalidation is no more reasonable than that of stronger reasons.

As in the bridge, moderations in terms of discounting can entail that we should stop the stunt. Again, the weaker the doctrine, the more reasonable. By modifying the stunt we can see how 'discounting versions' of anti-paternalism are in one way sharply distinct from absolute versions. Assume that the liberty reason and the harm to others reason are equally strong. Now the case is a sort of moral dilemma as long as the harm to self reason is invalid. However, as long as this reason has some influence, however small, we have stronger reasons to stop the stunt. Any discounting version, regardless of the discount factor, will give the same answer. Absolutist anti-paternalist will not like the idea that the personal good of the stunt artist should decide the case if favour of intervention.

Why should the personal good reason be blocked or discounted? As in the bridge, the person's high degree of voluntariness is no reason to disregard her health. To see this clearly, it helps to understand why she goes ahead with the stunt when we have overriding reasons to stop her. Unlike the bridge, the person in the stunt is perfectly informed and has perfect capacity of judgement. There is no mistake involved. Assume the following explanation: The person considers the risk of harm to herself a rather strong reason against performing the stunt. She considers the expected thrill and fame a somewhat stronger reason for performing it. Since these reasons are of similar strength, she would refrain from the stunt if she considered the risk to the passer-by a reason of some strength. However, she couldn't care less about the well-being of some passing stranger. Therefore, in her judgement the reasons for performing the stunt override the

reasons against. We, unlike the person, do feel that the risk to the passer-by is a strong reason against performing the stunt and so think that she should refrain. Alas, she goes ahead.

Our decision whether or not to stop the person is of course distinct from her decision whether or not to perform the stunt and so different reasons may apply, and reasons relevant to both decisions may have different strength in each case. Just as we can consider the person's reasons for performing the stunt, she could consider our reasons for stopping her. In both cases, we could argue with her about the relevant reasons and their relative strength. When it comes to stopping her, it could be that she agrees with us on the strength of our (weak) reason to eliminate her voluntarily assumed risk and our (strong) reason to let her take that risk. She may also agree that these reasons are valid. It might be, then, that she opposes our intervention only because she thinks that we exaggerate the strength of our reason to eliminate the risk to the passer-by. In other words, there is no mismatch between our views as they concern those reasons for and against stopping her that concern her liberty and her well-being. Anti-paternalism therefore implies that we should disregard as invalid (or discount) a reason that concerns the good of a person, even though she is herself in perfect and voluntary agreement with us concerning the strength and validity of this reason.

It could perhaps be argued that we have no reason to eliminate voluntarily assumed risks to start with. This is not a view about invalidation then but about basic value. It is an unreasonable view. People accept risks, voluntarily, because that seems to them the best option given the circumstances. If the circumstances are harsh, the risks may be great. People may regret having to choose between two risky options and still make a choice, voluntarily. If we can do something to lower these risks, without losing anything of comparable value, we should do so. If you voluntarily choose to risk your life driving to work by route A over risking your life by driving to work by route B, this in no way implies that I (being in charge of city planning) have no reason to try to decrease the risks involved in driving either route.

To sum up the first argument, anti-paternalism is unreasonable because it gives the wrong answer in two typical cases. The first case shows that people who are slightly less than perfectly informed and capable may make mistakes and that these mistakes may have disastrous consequences. When they do, the relatively high degree of voluntariness is no reason to disregard these consequences. In fact, doing so leads to morally incorrect conclusions. These conclusions can be avoided only by making anti-paternalism so narrow that it becomes redundant, or possibly by strong moderation of the doctrine. The second case shows that even people who are perfectly informed and capable, and so make no factual mistakes, may go wrong in the strength they attribute to reasons that concern things *other than* their own interests. This may lead them to do things they should not do and that we should stop them from doing if we can. We should stop them because of the value both we and they attribute to their own well-being, and because of the value of things that they do not acknowledge, such as protecting the interests of third parties. Anti-paternalism unreasonably entails that we should not stop such people.

Second argument – jumps in justifiability

The first argument is general and appeals directly to intuitions about what we should do and what kind of thing may reasonably make certain reasons invalid. The second argument draws on the fact that voluntariness is entirely a matter of degree (as noted by Feinberg 1986, p. 104). Anti-paternalism must set a threshold for how voluntarily a person must act in order that limiting her liberty qualifies as problematic interference. At this threshold, the justifiability of actions takes a peculiar jump.

I will focus on informedness: In Mill's original bridge case the man knows nothing about the condition of the bridge, so perhaps he is below the threshold. What then if we stop him and so have time to explain our findings about the hidden weaknesses of the bridge, but he dismisses these findings? Is he now sufficiently informed so that there is no valid reason to stop him? Or should we rather take his dismissal as an indication that he has not fully understood the risks involved? Exactly how well do we need to explain our findings in order that he should qualify as sufficiently informed? Must we force him through a course in the mechanics of bridge durability in order to ensure that he really has the relevant information? The point of these questions is just that informedness comes in degrees and that a threshold must be set, on normative grounds.

As noted by Arneson (2005), Feinberg's account makes it clear that anti-paternalism is committed to 'the enormous overriding importance of the line between self-harming choice that is not quite voluntary enough and choice that just passes the threshold of being voluntary enough.' (p. 268).[14] Sometimes lines have to be drawn and sometimes much depends on whether or not a threshold is reached. For example, it is very reasonable that if a bridge is safe enough we will walk over it, while if it is not safe enough we will take a long detour and lose valuable time. This is reasonable even if the threshold is somewhat arbitrary and small differences may shift the status of the bridge from safe enough to not safe enough. However, in the case of invalidation we are not considering two alternative paths of action, but two theoretical perspectives (invalidation or no invalidation), only one of which demands that we draw a line. If drawing a line leads to unreasonable consequences, this speaks against the perspective that demands that we do so.

The problem with a threshold is not only its unreasonably great importance in determining what we ought to do. In the case of anti-paternalism, the threshold-setting does not determine directly the moral status of an action, but rather determines whether certain reasons have influence on this status. A valid reason for an action makes the action more justified. This means that anti-paternalism implies that an action that affects a person just below the threshold of voluntariness may be much more justified than an

[14] As part of his argument against Feinberg, Arneson points to the fact that soft anti-paternalists, embracing the requirement of voluntariness, occupy a somewhat unstable position between on the one hand acceptance of paternalism and on the other hand hard anti-paternalists, who allegedly do not think that lack of sufficient voluntariness makes interference any more legitimate (pp. 266-68). However, even hard anti-paternalists must embrace some of the components of Feinberg's concept of voluntariness, minimally competence and lack of duress – infants and people at gunpoint may certainly be stopped from harming themselves. The only difference between hard and soft anti-paternalists is which components they embrace and to what degree. There is no (remotely reasonable) stable, 'end point' position to occupy.

otherwise similar action that affects a person just over this threshold. In other words, at the threshold the justifiability of the action takes a 'jump'. Such jumps in justifiability are quite another thing than the simple sorting of actions into, for example, the categories of the morally permissible and impermissible.[15] The line gives rise to a gap, and a very large gap if the reason is a strong one. This is unreasonable, and there is a ready alternative – to allow that reasons have influence according to their strength.

Consider:

> **The suicide** A person tries to kill herself but we have a chance to stop her. Stopping her would interfere with her liberty, which is a strong reason against doing so. Stopping her would also save her life, which is a much stronger reason for doing so.

Anti-paternalism implies that whether or not we should stop the suicide depends on the person's degree of voluntariness (unless there are other reasons to do so). If the voluntariness is under the threshold, we have much stronger reason to stop it than not, and so stopping is clearly and with good margin justified. If, on the other hand, the voluntariness is over the threshold, we have only a strong reason not to stop it, and so stopping is clearly and with good margin unjustified. I propose that this jump in justifiability is unacceptable. In fact I doubt that we can bring ourselves to even comprehend this jump, at least when it depends on infinitesimal differences – one further tiny piece of information, or a tiny improvement in decision-making capacity, will imply that an action that would have been overwhelmingly unjustified becomes overwhelmingly justified (in section four I will consider some strategies for smoothing out the threshold).

It may seem doubtful that we could have a strong reason to stop the suicide, considering that the person herself has obviously judged her life to be miserable enough to be worth ending. If the person's act perfectly voluntarily and she tries to kill herself because she thinks that her life lacks value or importance, we have reason to stop her only if we allow that something can be good for her even though she does not agree. As I indicated above, I think this is quite possible. However, there are other possibilities. The person might agree with us on the value of her survival, but still want to kill herself for some higher purpose, that we find lack value. If so, the case resembles the stunt – there is no mismatch of views concerning the value of the person's survival and liberty,

[15] It may be thought that this admission commits me to accept jumps in justifiability in other areas, since it is much more justified to punish a person who has performed an impermissible action, or committed a crime, than one who has not. However, I do not accept this assumption. It is less justified to punish a person who has just barely behaved immorally than one who has done so with good margin. This is why the law should distinguish between crimes of different degrees of aggravation. More to the point, that it is more justified to punish a person who has just barely behaved immorally or criminally than a person who has almost but not quite done so is almost entirely due to practical considerations such as the efficiency of the justice system, and of unrelated normative considerations such as that there should be some room for legal immorality in a liberal society.

but only concerning the value of some other thing, such as the independence of Tibet, or the glory of God.[16]

If the person acts less than perfectly voluntarily, there are further possibilities. She might simply feel compelled to end her life, perhaps out of despair, in conflict with her own preference. She might then agree both that she should not try to kill herself and that we should stop her. Consider a more everyday case of this kind:

> **The smoker** A person tries to go on smoking but we have a chance to stop her. Stopping her would interfere with her liberty and rob her of the pleasure of smoking, which are two strong reasons against doing so. Stopping her would also greatly improve her health prospects, which is a much stronger reason for doing so than the two other reasons combined.

It should not be far-fetched to assume that the smoker might realistically agree with us concerning the strength of our reasons. She values the pleasure of smoking and she would dislike being forcibly prevented from buying cigarettes. However, she values her health more. Still, she cannot bring herself to invite coercive means to stop her from smoking. As in the suicide, anti-paternalism implies that whether or not we should stop the smoker depends on her degree of voluntariness (unless there are other reasons to do so). If she is under the threshold, stopping her is clearly and with good margin justified. If she is over the threshold, stopping her is clearly and with good margin unjustified. Again, this jump in justifiability is unreasonable.

The second argument is independent of where exactly the threshold is set to be. If there was a neutral or obvious level at which to draw the line, however, the implied jumps in justifiability would be somewhat less outrageous. Anti-paternalists often draw a sharp line between competents and incompetents, between the healthy and the (mentally) ill, between adults and children.[17] However, the physical properties underlying these categories vary by degree. It may be thought that legal status provides a less arbitrary basis for a threshold. This is not so. Once bestowed, legal status may admittedly make a normative difference. It is perhaps worse to limit the liberty of a person of age, because this frustrates legitimate expectations not to be so treated that are induced by the legal system. Similarly, it is in one sense less objectionable to limit the liberty of a person if she has been committed to a mental hospital, regardless of the actual state of her psyche (though committing her is of course typically more objectionable if she is healthy).

[16] Another possibility is that there is perfect agreement on the strength of the reasons involved and that these reasons determine that the person should try to kill herself and that we should try to stop her. For example, a code of honour that we all accept may give the person most reason to kill herself, while we have most reason to stop her, since there is no dishonour in preventing or being forcibly prevented from suicide (the person might not openly admit as much if the code demands that one does not prompt prevention). On this scenario, stopping the suicide may not be contrary to anti-paternalism, since interventions with a person that are in line with her wishes may not be liberty-limiting. Arneson (1980) makes this point in connection with a prohibition of duelling intended to enforce a general preference not to be challenged (p. 471).

[17] For a critical discussion of actual and possible Millian arguments for blanket justification of paternalism towards children as a group, see Aviram 1991.

However, such legal circumstances can only reinforce an underlying moral principle, which must be spelled out in terms of non-legal, concrete physical or psychological properties of persons. It would be hopelessly vacuous to argue that the people that we must protect from benevolent interference are those that have been granted a legal right to be so protected. A moral principle like anti-paternalism should help us decide upon such matters as when people should reach lawful age and when they may be committed, and so cannot itself depend on the answers to such questions.

What we ought to do in real cases where someone tries to kill herself or tries to go on smoking will of course depend on many factors. Perhaps we should sometimes let people chose death or ill health, because their liberty or autonomy or self-determination is that important. This does not mean, however, that their survival or health provides no valid reason for action. Rather, reasons that concern these things are overridden. Perhaps we should sometimes stop people who act less voluntarily from harming themselves, but not those who act more voluntarily. If so, this will normally be because the cases differ in terms of some value. Perhaps the freedom of choice of people who act less voluntarily has less value. Perhaps, more reasonably, the fact that they make worse choices, with worse consequences, gives us stronger reasons to stop them. None of this, of course, implies that we have no valid reason to stop people who act more voluntary from harming themselves.

Third argument – liberty for all

The first two arguments both attack anti-paternalism for being too ambitious in saying that certain reasons are invalid. Together, these two arguments support the main claim of this article – that normative anti-paternalism is unreasonable. However, there is also a sense in which anti-paternalism is not ambitious enough. If liberal values such as liberty and autonomy invalidate reasons that concern such values as health and well-being, they need never be weighed against these things. Without an account of the relative value of liberty we have nothing to say about cases that do not fall under the anti-paternalist doctrine. Freedom is important, and it is important also for the demented, for minors, for the ignorant, and for people under time pressure. It is not as if the value of controlling (to some extent) ones own life kicks in only at a certain degree of informedness and rationality. Perhaps there is some level under which people *cannot* choose for themselves or cannot appreciate self-determination. Freedom, however, has value for people that are well above this level but that we should nonetheless sometimes coerce in their own interest, for example people in their lower teens. Why then does the protection of this freedom not activate the doctrine of anti-paternalism?

Three answers are possible for the anti-paternalist. First, she can insist that the freedom of young teens (and generally people acting under the sufficient degree of voluntariness) is of another type than the freedom of informed and rational adults and so does not qualify as true liberty (or whatever is the central value) and so does not activate the doctrine. This distinction in value is mysterious. Second, the anti-paternalist can lower the threshold and claim that the freedom of young teens (etc.) should never be limited in their own interest, and in fact that their well-being is not even a valid reason

for limiting their liberty. Even those who are prepared to stand back and let the people in the above cases kill or harm themselves (and others) would surely agree that this is unreasonable. Third, the anti-paternalist can say that the value of liberty is for young teens (etc.) appropriately reflected in the strength of such reasons as concern this value, and so there is no need for anti-paternalism in this area. This is the reasonable answer. But if this is the anti-paternalist's answer, the further question must be why the same is not true for informed and rational adults, and indeed for anyone.

Possible counter-arguments: Mistakes and anti-paternalist theories of the good

We may of course be mistaken concerning whether or not a certain action actually promotes a person's good. We may either be mistaken about the empirical outcome of the action, or about the value of that outcome. I have set aside contingent empirical considerations and I take it for granted that we fairly often produce those outcomes we intend to produce, also as they concern others, so that non-contingent empirical mistakes do not motivate anti-paternalism in the form of invalidation. Mistakes concerning value are normative but whatever their prevalence they do not, normally, support anti-paternalism. Imagine that we mistakenly hold that well-being or physical health is part of the good for a person, while in actuality the good consists only in moral excellence. This would certainly be relevant to many situations that anti-paternalism is typically concerned with, but has more general implications. For example, the fact that well-being is promoted by helping people in their life projects according to their own wishes, would not be a reason to do so. A critique of reasons that is based on a theory of the good is not, normally, an anti-paternalist critique. Rather, such a critique makes anti-paternalism concerning these reasons redundant. It is pointless to advance a doctrine that states that certain facts, which do not provide reasons because they concern nothing of value, do not provide valid reasons for the particular class of actions that amount to problematic interferences.

However, there is one exception to the norm. Theories of the good can in fact be designed to mirror anti-paternalism. It could be claimed that nothing that comes about as a result of problematic interference can be good, regardless of how closely it resembles actual good in other respects. Similar claims in other areas have some credibility. It is a common position that happiness caused by other people's suffering has no value, even though happiness in general has (e.g. Dancy 1993, p. 56). There may appear to be something evil, something immoral, about suffering-induced happiness, irrespective of, or in addition to, the evil of the suffering (which may not even be actual). The quality of the happiness, even the feeling of it, may be thought foul. More generally, it may be claimed that happiness is only of value when it is deserved.

I do not wish to challenge the position that malicious or undeserved happiness is more bad than good. However, these things may be seen as compounds. One could argue with Irwin Goldstein (1989) that 'malice could be intrinsically bad because of its pleasantness without the pleasantness in malice being intrinsically bad.' (p. 270) In other words, it could be a good thing to feel pleasure, but a bad thing to take pleasure in the suffering of others. It could be argued further that what is bad about malicious happiness

is not the happiness as such but the lack of compassion that is its source. Similarly in the case of undeserved happiness, what is bad is arguably the lack of retribution or the injustice of the situation. If so, malicious or undeserved happiness could be more or less bad depending on how the good of the happiness compares to the bad of the lack of compassion or the injustice. In general, it is advisable to place importance with separate things rather than compounds, so as not to mask invalidation and other operations on the influence of reasons behind purported judgements of basic value or goodness. If we agree on the normative influence of some fact, it does not matter whether or not it is a compound. In cases of diverging judgements of influence, however, we should look to the finer components to help us understand and possibly resolve the conflict.[18]

More importantly in the present context: Even if malicious and undeserved happiness should lack (positive) value, these things are significantly different from interference-generated good. We may be repelled by a coercive interference that seems to us not to respect the autonomy of the person. However, the resulting good, be it preserved life or health or some more mundane benefit, is not tainted by its history in any way similar to that of malicious happiness. If I learn that you avoided an accident, and so stayed healthy, only because you were involuntarily coerced, this does not affect my positive appraisal of your health, regardless of how unjust, immoral or even disgusting I find the coercion. We may admit then that some extreme theories of the good will have the same implications for what reasons have influence as will more standard theories of the good coupled with invalidation. Such theories then imply a form of anti-paternalism, but not a very reasonable kind and not the kind that is the topic of this article.

4. COMPROMISE AND MITIGATION

The argument from jumps in justifiability is strengthened by the fact that anti-paternalism seemingly implies that very small differences between two actions can make for large jumps, including jumps from overwhelmingly justified to overwhelmingly unjustified. The argument from wrong answers to moral questions is strengthened by the fact that the threshold degree of voluntariness is independent of what is at stake. In this section, I will first consider a compromise that links the threshold to benefit, or risk. I will then consider two attempts to smooth out the threshold and so avoid large jumps.

Requisite degree of voluntariness varying with how much personal good is at stake

An interesting aspect of Feinberg's anti-paternalism is that the threshold degree of voluntariness is made to depend on 'the nature of the circumstances, the interests at stake, and the moral or legal purpose to be served.' (1986, p. 117) In particular, the threshold depends in part on the severity of the risks involved – the higher the risks, the higher the threshold – and on whether the risks include irrevocable harm (pp. 118-121). In other words: The stronger the personal good reasons for an action, the higher the

[18] It could be argued that importance or goodness resides in compounds as a matter of metaphysical fact. I propose that the metaphysics of normative properties should answer to normative argument.

threshold for when that action qualifies as a problematic interference, and so the higher the threshold for when those reasons turn invalid. By this manoeuvre, the strength of personal good reasons is brought into the equation through the back door. Personal good reasons now give valid support, not to problematic interference of course, but to what would have been problematic interference were it not for the strength of those very reasons. The doctrine has to some extent been dismantled from within.

This is a very reasonable move to make for any anti-paternalist who, like Feinberg, is concerned not simply to 'prevent people from acting with low degrees of voluntariness', but rather to 'prevent people from suffering harm that they have not truly chosen to suffer.' (p. 119) The result is a version of anti-paternalism that is more reasonable because it takes into consideration how much personal good is at stake. Feinberg's compromise entails that the stronger the personal good reasons, the more likely they are to influence the moral status of actions according to their strength. However, for voluntary enough harms, the ban on personal good reasons still holds. *If* a choice is voluntary enough, all things considered, including the personal good factor, *then* personal good reasons are invalid and so do not influence the moral status of actions. The structure of the doctrine is thereby preserved, and with it the peculiar disregard for (a smaller but still substantial class of) personal good reasons.

As noted, very small deficits in voluntariness can make for disastrous consequences. Presumably, Feinberg would hold that disaster is morally irrelevant as long as the degree of voluntariness is high enough to match the risk. This is unreasonable. Furthermore, disaster can be the result of choices that are trivial from the point of view of liberty. Severe risk should perhaps be accepted for the sake of important liberties, but not simply because of a high degree of voluntariness. This point can be illustrated by comparing a well planned philosophical suicide with a five-party game of Russian roulette. What should count against stopping these activities is not so much the degree of voluntariness (for which Feinberg must presumably set a much lower threshold in the case of the five times less risky Russian roulette) but rather the sort of liberty at stake and its value. Moreover, a flexible threshold in no way avoids jumps in justifiability. The compromise is unstable. It points the way to the preferable solution of simply allowing both liberty reasons and personal good reasons influence according to strength.

Indeterminacy

Jumps in justifiability are particularly aggravating because they depend on infinitesimal differences. Moving along the scale from innocuous involvement to problematic interference, there would seem to be a point at which your health is no longer a valid consideration when evaluating my attempts to make you give up smoking. At the same point, public health stops being a valid consideration when evaluating government policy directed at decreasing drug abuse or increasing the use of seat belts. Two strategies of mitigation may be proposed to remove this point by replacing it with indeterminacy or with gradual transition.

It may be suggested that anti-paternalism does not in fact draw a sharp line between problematic interference and other actions, since there is an area in between in

which we simply do not know, or where it is genuinely indeterminate, whether an action is or is not an interference. The former, epistemic version of indeterminacy does not make the doctrine more reasonable. If any action either is or is not a problematic interference, and if whether or not it is depends on the properties of the action, then the fact that we cannot distinguish problematic interferences from other actions does not make it any more reasonable that the validity of personal good reasons depends on infinitesimal differences between actions.

On the other hand, the indeterminacy may be not only epistemic, but ontological. Some actions would then be problematic interferences, some would be indeterminate and some would not be problematic interferences. The indeterminate actions are simply beyond the scope of the doctrine. This account may be questioned if one demands that normative principles be action-guiding at least in theory. We could then demand an answer to the question of how to treat indeterminate actions – should we or should we not allow that they are supported by personal good reasons? If anti-paternalism does not reject the invocation of personal good reasons for indeterminate actions, it seems that the line has simply been moved from the point where innocuous involvements become interferences to the point where indeterminate actions become interferences. Similarly, if the doctrine does reject the invocation of personal good reasons for indeterminate actions, the line has been moved to the point where innocuous involvements become indeterminate.

Requiring action-guidance in this sense is perhaps overly demanding. Perhaps we must simply accept that it is sometimes indeterminate whether or not personal good reasons give valid support to an action. If so the threshold is turned into a hole. Consequently, some answers to moral questions are replaced by no answers. However, the cases in the first argument above (the bridge, the stunt) are designed to involve people of high and even perfect voluntariness and so it is hardly indeterminate whether stopping them amounts to problematic interference. On the other hand, the sudden jumps in justifiability in cases like the suicide and the smoker would indeed be replaced with a twilight zone of indeterminacy. Stopping would be overwhelmingly justified on one side of this zone, overwhelmingly unjustified on the other side, and indeterminate in between. We avoid jumps by giving up comprehensiveness. Some might possibly consider this an improvement, though certainly not a great one. Avoiding difficult boundary issues by making hard cases indeterminate is no great improvement.

Gradual transition

Another attempt to disarm the problem of arbitrariness reformulates the doctrine as follows: Personal good reasons are valid only *to the extent* that the action is not a problematic interference. Absolute anti-paternalists should be reluctant to accept this reformulation, since it replaces the absolute moral ban on benevolent problematic interference with a sliding scale of gradual acceptance of such interference. We may still consider, of course, whether it is reasonable. The reformulation seems to transform the impossible question of where to draw the line between valid and invalid reasons into the more agreeable though vastly more complex series of questions of when the validity of

personal good reasons starts to decline, how fast it declines, and at what point (if any) they are completely invalid. In other words, on this account the influence of a reason for an action is some function of two variables – strength and the degree to which the action is a problematic interference.

Another way to describe the account is to say that personal good reasons are discounted by a changing factor. Gradual transition is in this sense equivalent to moderation by discounting. Discounting or partial invalidity must be kept distinct from relative weakness. If the gradual rejection of paternalism only means that personal good reasons are not very strong relative to other reasons, then paternalism is in fact fully accepted over the whole range of gradual rejection. Assume as in section two that there are H reasons (to prevent harm to a person), L reasons (to respect the liberty of the same person) and O reasons (to prevent harm to another person). Discounting can be distinguished from relative weakness by pointing out that an H reason which is discounted by an L reason can still be the stronger in the sense that it would override an O reason that is stronger or equally strong as the L reason.

Gradual transition would not make the answers to the moral questions in the first argument noticeably more reasonable, since the degree of voluntariness is very high or maximal, and so the validity of the personal good reasons very low. The jumps in the second argument would be replaced with a sliding scale, and so stopping the suicide or the smoker would be more justified the lower the degree of voluntariness. What exactly the sliding scale would look like depends on the function from strength and 'problematic interference'-ness to justifiability. It is not clear to me what should specify this function. What is clear is that in a conflict between a liberty reason and a personal good reason of a certain strength, holding the latter constant entails that, normally, the higher the degree of voluntariness, the more liberty at stake, and the more liberty at stake, the smaller the relative strength of the personal good reason. In this sense, the influence of personal good reasons varies along a sliding scale regardless. I fail to see the advantage of an additional scale at the level of influence-regulation.

In sum, then, Feinberg's compromise makes anti-paternalism more reasonable, but only by moving in the direction of abandoning the doctrine entirely. Indeterminacy and gradual transition provide small and superficial improvements. The general structure of anti-paternalism remains problematic and vulnerable to the arguments in the previous section, or slight variations of those arguments.

5. CONCLUDING REMARKS

Regard for the good of others is a splendid thing. Sometimes we are so fortunate that there are no reasons against acting for the good of others. More commonly, however, benefits are only one among many considerations we need to take into account. Perhaps some will benefit and others lose, perhaps there are more intricate questions of justice to consider, perhaps non-personal values such as preservation of the natural wilderness or the continuation of the human species may be affected. Perhaps what benefits someone also limits her liberty or disrespects her autonomy. To oppose anti-paternalism is not to

disregard whatever reasons exist against problematic interference. Without influence-regulating doctrines, moral argument must consider the strength of each reason on its own terms and compare it to the strength of other reasons. In some cases the reasons against interference with a person are so strong that the benefit to her pales in comparison. This does not imply that the benefit does not provide one valid reason among others.

Voluntariness is a factor in many other contexts than paternalism. A claim that a person has a right to something often means that she may have or do this thing on the condition that her choice is sufficiently voluntary. The right to marry freely, to vote, to enter contracts – these rights are arguably conditional on voluntariness. Annulling or failing to accept a marriage, a vote or a contract does not violate a right if the parties concerned did not act voluntarily. We may want to annul or refuse to accept these things for personal good reasons, but also for the good of other people, or for whatever other reason. The arguments against anti-paternalism can therefore rather straightforwardly be reformulated as arguments against most any doctrine of invalidation.

To oppose influence-regulating principles is not to deny that there are intricate relationships between reasons. Things of value may be empirically related to each other in the sense that actions tend to affect many such things simultaneously. Things of value may also be conceptually related – value may for example be disjunctive in the sense that it is of value that one of several things happen. Investigations of the bearers of value and their empirical and conceptual relationships should help us clarify and handle conflicts of reasons. Without striking a wedge between the strength and the influence of reasons, potentially conflicting values such as liberty, autonomy, self-determination, physical and mental health, desire-satisfaction, happiness, and so on, may be investigated each in their own right.

When Isaiah Berlin tells us that '[t]he extent of a man's negative freedom is, as it were, a function of what doors, and how many, are open to him; upon what prospects they open; and how open they are' (2002 [1969], p. 41); when Amartya Sen develops his concept of freedom as capability (1992, chapter three) or when Joseph Raz develops his ideal of autonomy (1986); when Mill develops the notion of individuality (1991 [1859], chapter III), and even when he briefly states that '[t]he only freedom which deserves the name, is that of pursuing our own good in our own way' (p. 17) – in all these cases we see significant contributions to our system of values, to our views on what is important in life. When, on the other hand, Mill tells us that 'the sole end for which mankind are warranted, individually or collectively, in interfering with the liberty of action of any of their number, is self-protection' (p. 14), we are left confused as to what may then be the value of liberty and what other values there may be, that they should be related in this way.

ACKNOWLEDGEMENTS

This article has been in the writing for a long time and I so am sure I have some unrecognized debts. For very helpful comments, I wish to thank especially Sara Belfrage, Alon Harel and Lars Lindblom. An anonymous reviewer for *The Journal of Political Philosophy* exposed several weaknesses as well as the many ways in which my argument could be misunderstood. Hopefully it is now more transparent.

REFERENCES

Arneson, Richard. 1980. Mill versus Paternalism. Ethics 90: 470-89.

Arneson, Richard. 2005. Joel Feinberg and the Justification of Hard Paternalism. Legal Theory 11: 259-284.

Aviram, Aharon. 1991. The Paternalistic Attitude Toward Children. Educational Theory 41: 199-211

Berlin, Isaiah. 2002 (1969). Five Essays on Liberty: Introduction. In Liberty. Oxford: Oxford University Press.

Dancy, Jonathan. 1993. Moral Reasons. Oxford: Blackwell.

Day, J.P. 1970. On Liberty and the Real Will. Philosophy 45: 177-192.

Dworkin, Gerald. 1972. Paternalism. The Monist 56: 64–84.

Feinberg, Joel. 1984. Harm to Others. Oxford: Oxford University Press.

Feinberg, Joel. 1986. Harm to Self. Oxford: Oxford University Press.

Gert, Joshua. 2007. Normative Strength and the Balance of Reasons. Philosophical Review 116: 533-562.

Grill, Kalle. 2007. The Normative Core of Paternalism. Res Publica 13: 441-458.

Goldstein, Irwin. 1989. Pleasure and Pain: Unconditional, Intrinsic Values. Philosophy and Phenomenological Research 50: 255-276

Groarke, Louis. 2002. Paternalism and Egregious Harm. Public Affairs Quarterly 16: 203-230.

Kamm, F.M. 2007. Intricate Ethics. Oxford: Oxford University Press.

Husak, Douglas. 2003. Legal Paternalism. In The Oxford Handbook of Practical Ethics. Oxford: Oxford University Press.

Kleinig, John. 1983. Paternalism. Manchester: Manchester University Press.

Mill, John Stuart. 1991 (1859). On Liberty. In On Liberty and Other Essays. Oxford: Oxford University Press.

Nozick, Robert. 1974. Anarchy, State, and Utopia. Malden MA: Basic Books.

Raz, Joseph. 1986. The Morality of Freedom. Oxford: Oxford University Press.

Raz, Joseph. 1990 (1975). Practical Reason and Norms (2d edition). Princeton: Princeton University Press.

Scanlon, Thomas. 1998. What We Owe to Each Other. Cambridge MA: Harvard University Press.

Shiffrin, Seana. 2000. Paternalism, Unconscionability Doctrine, and Accommodation. Philosophy and Public Affairs 29: 205-250.

Sen, Amartya. 1992. Inequality Re-examined. Oxford: Oxford University Press.

Sunstein, Cass R. and Thaler, Richard H. 2003. Libertarian Paternalism Is Not an Oxymoron. The University of Chicago Law Review 70: 1159-1202,

Van de Veer, Donald. 1986. Paternalistic Intervention. Princeton: Princeton University Press.

Paternalistic Interference

Kalle Grill

ABSTRACT: In order that paternalism, understood as benevolent interference, should be a morally relevant category, there would have to be something morally significant about the combination of benevolence and interference that goes beyond that of interference as such. In the context of paternalism, interference is almost universally defined in terms of either liberal values such as liberty, autonomy or sovereignty, or a self-regarding sphere of life, or operationalizations of the liberal values in terms of choice, decision or agency, or coercion, or wronging. A survey of these five accounts shows that none provide a definition on which paternalism is morally relevant. We should therefore understand paternalism as a conflict between typical liberal values and other values for a person, which but must be resolved in different ways in different contexts.

INTRODUCTION

Paternalism minimally involves conflicting considerations regarding the same person, where some sort of personal liberty or autonomy value conflicts with some other value for the person.[1] In a sense, each conflict between two values presents a special moral problem. How does happiness compare to achievement, nature to art, social cohesion to the rule of law? Investigating such value conflicts with a focus on the nature of the values involved and their relative importance is fascinating and important philosophical work. Discussions of paternalism, however, are often carried out under the assumption of anti-paternalism – the principle that paternalism is prima facie wrong. From the perspective of value conflicts, anti-paternalism entails that the paternalism conflict is resolved once and for all, for all instances, in favour of one of the values. In other words, there is no conflict, but a principle.

Admittedly, there is in a sense a prima facie wrong involved in any value conflict where a value is diminished. However, if another value is promoted or protected, something is also prima facie right.[2] The value conflict as such identifies a prima facie wrong only if the diminishing always outweighs the promotion or protection. I propose that conflicts between liberty or autonomy and other personal values, such as health,

[1] Seana Shiffrin has defended the radical position that paternalism need not involve benevolence or the promotion of good ('Paternalism, Unconscionability Doctrine, and Accommodation', *Philosophy and Public Affairs* 29[3] [2000] 205-250). Pace Shiffrin, I will assume in the following that paternalism involves a benevolent reason. If it does not, it is all the more urgent to investigate the remaining core component – interference.

[2] I assume that the good of people in general have value and that it is in general right to promote or protect that value. If we have no reason to ever promote the good of others, and we have some reason to avoid diminishing of liberty or autonomy, then paternalism is trivially prima facie wrong.

happiness or achievement, have no fixed solution but should be resolved in different ways in different cases.[3]

In a value conflict, actions are important only to the extent that they affect values. There is no need to determine in the abstract what actions or action types affect what values in what way. Cutting off someone's leg normally diminishes her health but sometimes protects it – this does not complicate the value conflict, which is concerned with the relative importance of health, regardless of how changes in health come about.[4] In contrast, principles are more closely tied to actions. Negative or constraining principles such as anti-paternalism can operate in either of two ways: The principle may target actions, saying that certain actions, defined in part by what reasons are invoked for them, may not be performed, or may be performed only under certain circumstances. Or the principle may target reasons for action, saying that certain reasons may not be invoked for certain actions, or may be so only under certain circumstances. Correspondingly, on a principled approach, paternalism could be understood either as the performance of certain actions, defined in part by what reasons are invoked for them, or as the invocation of certain reasons for certain actions.[5] In either case, we can abstract from the role of reasons and focus on the actions that allegedly amount to paternalism when combined with benevolent reasons in one of these ways. I will call such an action an *interference*. This article argues against a principled approach to paternalism by scrutinizing the criteria for defining interference so specified.

If the categorization of something as paternalism shall be morally relevant, there must be something about interferences that makes these actions particularly problematic in combination with benevolent reasons – reasons that concern the good of the person interfered with. This property may be prima facie, or may be conditional on certain circumstances. Importantly, the moral relevance must not be parasitic on some other property or properties. To see this, assume that it is a valid moral principle that the killing of innocents is always morally wrong. Now to claim that *benevolent* killing of innocents is always morally wrong does not add anything to our criteria of moral evaluation (unless the benevolence makes the killing even more wrong). There may of course be more than one reason for a moral position. We may think, for example, that sin taxes are wrong

[3] Many contributions to the discussion on paternalism question the strictness or absoluteness of the anti-paternalist principle but remain true to the background anti-paternalist framework. I agree with Douglas Husak's one-time conclusion that 'philosophers should abandon the attempt to formulate general objections to paternalism, and should concentrate instead on assessing the justifiability of instances of paternalism on their individual merits.' ('Paternalism and Autonomy', *Philosophy and Public Affairs* 10[1] [1981] 27-46, p. 46) Husak's example is Gerald Dworkin, 'Paternalism', *Monist* 56 (1972) 64-84. A more recent example is Louis Groarke, 'Paternalism and Egregious Harm', *Public Affairs Quarterly* 16(3) (2002): 203–230. In more applied contexts, examples abound. Recent critics of anti-paternalism include Richard Arneson ('Joel Fienberg and the Justification of Hard Paternalism', *Legal Theory* 11 [2005] 259-84) who delivers a sharp critique of Joel Feinberg's anti-paternalism and concludes that 'there is no successful case against hard paternalism to be made' (pp. 259-60), and Peter de Marneffe ('Avoiding Paternalism', *Philosophy and Public Affairs* 34[1] [2006] 68-94) who similarly concludes that 'there is no compelling reason to think that paternalism is always wrong'. (p. 69)

[4] My thesis and argument are independent of what exactly is good for a person. Even on extremely subjectivist theories of value, according to which nothing has value for a person unless she thinks it does, other people can promote or protect her good so defined through more or less interfering actions.

[5] See Kalle Grill, 'The Normative Core of Paternalism', *Res Publica* 13 (2007): 441-458.

both because taxing is evil in itself and because the government should be neutral towards different products and lifestyles. However, if taxing is evil only because it is non-neutral, then the evil of taxing is parasitic on the evil of non-neutrality and so of no concern in its own right. Similarly, if benevolent killing or benevolent interference with liberty is wrong only because and to the extent that killing or interference with liberty is wrong, then the former two categories are of no concern in their own right.[6]

The thesis of this article is that the fact that a situation is an instance of paternalism does not add any moral content to that situation. In other words, this categorization is not morally relevant in its own right. We should not, therefore, assume that paternalism is prima facie wrong. Either it is not prima facie wrong or its wrongness is parasitic on some other moral category, in which case we would do better to invoke that property directly. My argument for this thesis is somewhat roundabout – it consists of a survey of five kinds of definitions of interference. I will focus on simple cases – cases where the actions in question are directed at adult, mature, well-informed people that are sober, calm and collected and under no duress or time pressure or other undue influence. The problem of impaired choice poses a difficult challenge for any account of paternalism, a problem I will leave for another occasion. My focus here is on the action type, on what distinguishes a typical case of interference. Interference is almost universally defined by invoking one or more of the following ideas: Liberal values such as liberty, autonomy or sovereignty[7]; a self-regarding sphere of life[8]; operationalizations of the liberal values in terms of choice, decision or agency[9]; legal prohibition or coercion more generally[10]; or wronging[11]. None of these action types provide a convincing account of interference.

Paternalism has a strong conceptual tie to authority, in the sense of rightful control over one's own life, or rather certain parts of it. Given a sphere of authority where a certain person should decide what is to be done, interference can be defined as unwelcome involvement in that sphere (assuming that unwelcome involvement simply means not letting the person decide what is to be done). Indeed, most accounts of interference seem to assume that a sphere of authority has been or can be specified. However it is not clear on what grounds such a sphere should be delimited. Self-determination, self-government, self-ownership – all of these notions must come with a specification of the self in question. Similarly, the notion of someone having authority must come with a specification of what is under this authority. The delimitation of

---

[6] Cf. Husak, 'Paternalism and Autonomy', p. 38: 'The conclusion that paternalistic interferences, qua interferences, are objectionable, is a good deal less interesting than the conclusion that paternalistic interferences, qua paternalism, are objectionable.'

[7] E.g. John Stuart Mill, *On Liberty*, in *On Liberty and Other Essays*, (Oxford: Oxford University Press 1991 [1859]); Dworkin, 'Paternalism'; Feinberg, *Harm to Self* (Oxford: Oxford University Press 1986).

[8] E.g. Mill, *On Liberty*; John Gray, *Mill on Liberty* (London: Routledge and Kegan Paul 1983); Feinberg, *Harm to Self*.

[9] E.g. David Archard, 'Paternalism defined', *Analysis* 50(1) (1990) 36-42; Gerald Dworkin, 'Some second thoughts', in Rolf Sartorius (ed.) *Paternalism* (Minneapolis: University of Minnesota Press 1983); Shiffrin, 'Paternalism, Unconscionability Doctrine, and Accommodation'.

[10] E.g. Dworkin, 'Some second thoughts'; Feinberg, *Harm to Self*.

[11] E.g. Bernard Gert & Charles M Culver, 'Paternalistic behavior', *Philosophy and Public Affairs* 6(1) (1976); Donald Van de Veer, *Paternalistic Interference* (Princeton: Princeton University Press 1986).

authority will be a re-occurring theme in the following survey of proposed definitions of interference.

Notions of authority may be of different force or moral status. Throughout the text, I will distinguish between two kinds of authority. On the one hand, there is what I will call the *sphere of all things considered authority*, or 'authority$_C$' for short. This is a sphere or area in which, all things considered, a certain person (or possibly some other entity) should decide what is to be done. On the other hand, there is what I will call a *sphere of fundamental authority*, or 'authority$_F$' for short. This is a sphere or an area in which a certain person has a strong prima facie right to decide what should be done, a right that is fundamental in the sense that it is independent of other considerations. This right need not be absolute, but could be qualified so as to either discount other considerations rather than trump them, or to yield to certain other values, such as long term autonomy.[12]

If interference is unwelcome involvement with authority$_C$, it is obviously all things considered wrong. A claim that some action is an interference then have the full moral content of all relevant considerations, whichever they may be. If so, however, paternalism as such is morally irrelevant by definition. We may distinguish between benevolent and other reasons for infringement of authority$_C$, but such distinctions cannot affect the proper limits of authority$_C$ since this concept already includes all relevant considerations. Similar though less obvious problems follow from identifying interference with wronging, the fifth approach to be surveyed.

If interference is unwelcome involvement with authority$_F$, a claim that an action is an interference has substantial moral content in its own right, irrespective of other considerations. However, since authority$_F$ does not rest on other considerations it stands in need of explanation. If paternalism is defined in terms of authority$_F$, then explaining the moral force of authority$_F$ should be the main task in fleshing out the definition. The first four accounts of interference to be surveyed depend on an idea of authority$_F$. However, as will be shown, this idea is usually invoked only implicitly and is never adequately explained.[13] This inadequacy and the lack of promising alternatives should lead us to reject paternalism as a morally relevant category and accept the minimal definition of paternalism as one value conflict among many.

DIRECT APPEAL TO LIBERAL VALUES

Paternalism is often defined in terms of traditional liberal notions such as liberty, autonomy or sovereignty. It is crucial to see that these notions can play three distinct roles in the context of paternalism. First, they can be predicated as honorary names of

[12] Authority$_C$ and authority$_F$ are in a sense end points on a scale. We could define intermediate kinds of authority that take into account some considerations but not others, for example normative but not empirical. Such attempts do not, I believe, avoid the critique in this article, but rather invite a more complex critique.

[13] This is no coincidence. I will not discuss the possibility of formulating an adequate account of authority$_F$, but I do not believe there is one. I tend to agree with William A Edmundson that a right to (and the value of) non-interference can only be understood in terms of protection against 'disproportionately severe and officious reactions'. (*An Introduction to Rights* [Cambridge: Cambridge University Press 2004], pp. 170-172) The appropriate proportions, I would say, are determined by the proper balance of values.

authority$_C$. Such names may be appreciated for carrying the collected wisdom of a long history of considering the proper limits of this sphere, although they also carry the collected confusion of that history. Second, liberal notions can refer to values that figure among the various considerations that determine the proper boundaries of authority$_C$. In this role, the liberal values are balanced against each other, and against other personal and non-personal values. This to me is the proper role of any value. Third, liberal values can refer to authority$_F$ as a value that trumps or diminishes other values. It is in this third role that liberal values can underpin anti-paternalist principles – principles saying that there are certain things that simply may not be done to a person for her good, or that may be so done only in certain circumstances.

John Stuart Mill famously stated that 'the sole end for which mankind are warranted, individually or collectively, in interfering with the liberty of action of any of their number, is self-protection.'[14] This general liberty principle implies that interference with liberty is not warranted by benevolence, or in other words: 'His own good, either physical or moral, is not a sufficient warrant [for the exercise of power over a person against his will].'[15] In other words, interference (as used here generically to denote actions that amount to paternalism when combined with benevolent reasons) is defined by Mill as interference with liberty (of action) and/or unwanted exercise of power. What, however, is the extent of a person's liberty? When are we exercising power over someone against her will? It has been argued that Mill's formulations are too narrowly focused on behaviour control.[16] This is not my worry here, and it seems to me that the notion of liberty can be extended to cover more than behaviour. Rather, the question is how the limits of liberty should be understood in the context of paternalism, regardless of whether liberty concerns only behaviour or also other things, such as bodily integrity and entitlements of various sorts. Mill emphasises that we may sometimes act for the good of other people, that indeed we may have a duty to do so. For example, we 'owe to each other help to distinguish the better from the worse, and encouragement to choose the former and avoid the latter.'[17] Furthermore: 'Considerations to aid [a person's] judgement, exhortations to strengthen his will, may be offered to him, or even obtruded on him'.[18] This is very reasonable – the good of others is often a fine and proper reason for action that may outweigh other concerns, such as a person's unwillingness to hear advice and warnings. What then may we *not* do for the good of others? More fundamentally, what decides this matter? Clearly, liberty features here as a limit on what may be done and not as one value among many. We may ask then if liberty in Mill's argument refers to something like authority$_C$ or something like authority$_F$.

Mill does little to explicate the notions of interference with liberty and the exercise of power over someone against her will. Rather, the exact content of his anti-paternalism is the end result of a series of considerations of empirical circumstances and of the potential of non-interference to promote the greater good in general and in

[14] Mill, *On Liberty*, p. 14.
[15] Ibid., p. 14.
[16] Gert & Culver, 'Paternalistic behavior'.
[17] Mill, *On Liberty*, p. 84.
[18] Ibid., p. 85.

particular to promote the value of individuality. Determining the limits of liberty seems to be a matter of balancing diverse considerations. Mill does not deny that people sometimes err in pursuing their best interests, or that, in theory, it is possible to promote a person's good through interference.[19] However: 'All errors which he is likely to commit against advice and warning, are far outweighed by the evil of allowing others to constrain him to what they deem his good.'[20] These aspects of Mill's argument indicate that his concern is to determine the proper boundaries of authority_C in light of the full range of relevant considerations. If this is the nature of the project, however, Mill's argument collapses into an unprincipled balancing of different considerations, and his principled formulations are demoted to rhetoric.

For the most part, however, Mill himself, as well as his interpreters, seems to think that his defence of liberty is a matter of principle. This is perhaps seen most clearly in borderline cases. In the much discussed argument against voluntary slavery, Mill does not say that the liberty value of continued legal autonomy outweighs the liberty value of having the present option of entering a slave contract. Rather, he suggests that allowing a person to enter such a contract would somehow be incoherent, since the voluntary slave 'defeats, in his own case, the very purpose which is the justification of allowing him to dispose of himself.'[21] Therefore, having the option of voluntary slavery is not part of liberty: 'It is not freedom, to be allowed to alienate his freedom.'[22] If liberty is authority_C, the conclusion makes perfect sense, since voluntary slavery is not (let us assume) something that should be allowed, all things considered. However, this is not the shape of the argument. Mill attempts to show that liberty cannot coherently be invoked to end liberty, *independently* of other considerations. Though David Archard[23] has offered some interesting arguments in support of this claim, Mill's position on voluntary slavery is more generally interpreted as a flagrant breach of principle.

What is pertinent to the present investigation is that both sides in this controversy can agree that voluntary slavery should not be allowed, all things considered.[24] The disagreement need not concern the proper limits of authority_C. Typically, it concerns the proper limits of liberty understood as an independent value. Furthermore, this value is not just one among many to be taken into consideration, for then it would be obvious that it is one good ting to be allowed to enter a voluntary slave

[19] Apart from the lack of an explicit denial of this possibility, which would support his anti-paternalist case immensely, Mill's position is further indicated by statements such as this: 'It is easy for one to imagine an ideal public, which leaves the freedom and choice of individuals in all uncertain matters undisturbed, and only requires them to abstain from modes of conduct which universal experience has condemned.' (p. 93)
[20] Ibid., p. 85.
[21] Ibid., p. 114.
[22] Ibid.
[23] 'Freedom not to be free: The case of the slavery contract in J.S. Mill's On Liberty', *The Philosophical Quarterly* 40 (1990) 453-65.
[24] Those that hold that liberty entails a right to enter voluntary slavery typically either hold that other considerations outweigh the importance of liberty in this case (e.g. Dworkin, 'Some second thoughts', p. 111), or that the right to enter a slave contract does not imply that liberty is infringed on by *abstaining to enforce* such contracts (e.g. Gray, pp. 93-97). Cf. Feinberg, *Harm to Self* pp. 71-81. Of course a few libertarians hold that voluntary slavery should be allowed (e.g. Nozick, *Anarchy, State, and Utopia*, Malden MA: Basic Books 1974, p. 331; Walter Block, 'Toward a Libertarian Theory of Inalienability', *Journal of Libertarian Studies* 17[2] [2003] 39-85).

contract, and another good thing to remain a free person. These things would simply be balanced against each other, none of them having any special claim to the name 'liberty'. Too much is supposed to hinge on the limits of liberty for this to be an accurate description of the controversy. What is supposedly at stake is the consistency of Mill's general anti-paternalism. This makes sense only if liberty is interpreted as a strong right to control certain parts of a one's life –i.e. as authority_F.

If liberty is authority_F rather than authority_C and so not to be derived from general moral and empirical considerations and circumstances, it is still an open question how to define its limits. As noted above, Mill very reasonably argues that some things are perfectly proper to do to a person for her good. What then would make true or false Mill's claim that prevention of voluntary alienation of liberty is one of those things? It is a reoccurring theme of Mill's argument that liberty is defined by what concerns or affects mainly a person herself and consenting others:

> But there is a sphere of action in which society, as distinguished from the individual, has, if any, only an indirect interest; comprehending all that portion of a person's life and conduct which affects only himself, or if it also affects others, only with their free, voluntary, and undeceived consent and participation. [...] This, then, is the appropriate region of human liberty.[25]

Mill holds that within this region of liberty, in that part of a person's conduct 'which merely concerns himself, his independence is, of right, absolute.'[26] The idea of the self-regarding thus seems to provide the proper limits of liberty as authority_F. However, as I will argue in the following section, this idea is in fact irrelevant to paternalism. If my argument is correct then Mill does not offer us an adequate definition of authority_F.

Joel Feinberg in many ways develops the Millian rejection of paternalism, holding that preventing voluntary self-harm is never a good reason for legal prohibition.[27] For Feinberg, the important notion is not so much liberty as autonomy and sovereignty. On Feinberg's view, autonomy may be many things, including a capacity, a condition, and an ideal.[28] However, when a person has a *right* to autonomy, she is *sovereign*. If a person is sovereign over some thing, her 'consent is both necessary and sufficient' for any dealings with that thing,[29] provided only that her 'decisions are genuinely voluntary'.[30] Feinberg's notion of sovereignty is an incoherent mix of authority_C and authority_F. He points out that the proper boundaries of the 'personal domain' depend on what liberty-limiting principles we accept, and advice that we move these boundaries to fit our

[25] Mill, *On Liberty*, p. 16.
[26] Ibid., p. 14.
[27] Feinberg, *Harm to Self*, chapter 17.
[28] Ibid., chapter 18.
[29] Ibid., p. 53.
[30] Ibid., p. 67. Voluntariness plays an important role in Feinberg's argument as an umbrella concept covering such diverse things as lack of duress, informedness, maturity, and lack of time pressure. Only when decisions are 'voluntary enough' are they properly seen as the decision *of the person* in question and so potentially under her sovereignty. For present purposes, we may assume that all relevant decisions are genuinely voluntary.

convictions.[31] This is perceiving of sovereignty as authority_C, a sphere of authority defined by various considerations. At the same time, Feinberg invokes personal sovereignty in his argument for drawing the lines of legitimate authority just where he wants them to be drawn. *Because* of personal sovereignty, we should reject paternalism, being one of several liberty-limiting principles. This is perceiving of sovereignty as authority_F, a moral consideration of special force that weighs heavily in determining the boundaries of authority_C.

On some level, Feinberg is aware that if sovereignty is to be as definitive and without exception as he wants it to be, and at the same time play a role in the moral reasoning defining authority_C, it needs an independent foundation. He invokes the analogy between personal and national sovereignty, as well as linguistic intuitions, to argue that sovereignty does not come in degrees, but is 'whole and undivided', 'an all or nothing concept', and so 'is respected in its entirety or not at all.'[32] It is not clear how this fact, if true, makes the notion of personal sovereignty any more potent. The allegedly binary character of sovereignty makes for more gruesome choices when values conflict, but provides no further grounds for favouring sovereignty over other values, nor does it even indicate that sovereignty is important.

In an attempt at providing sovereignty with some moral force, Feinberg, like Mill, falls back on the idea of the self-regarding, stating that 'we may locate within the personal domain all those decisions that are "self-regarding", that is which primarily and directly affect only the interests of the decision-maker.'[33] I will now turn to this idea.

THE SELF-REGARDING

Mill, Feinberg and others invoke the idea of the self-regarding to support their anti-paternalism. Could we define interference (again in the generic sense of actions that amount to paternalism when combined with benevolence) as unwelcome involvement with what is self-regarding? It is well-known from consequentialist critiques that it is difficult to draw the line between the self-regarding and the other-regarding. I will summarize this critique before I move on to the more novel argument from irrelevance. In essence, this argument is that in what way and to what extent a person is affecting others has no bearing on the moral status of promoting her good against her will.

First of all, a brief remark on terminology and scope: The predicates 'self-regarding' and 'other-regarding' are most often attributed to action, conduct or behaviour. However, some parts of a person's life affect others in ways that are largely independent of her behaviour. So, for example, that a person owns property affects others through our shared norms and institutions, without the person performing even the miniscule action of insisting on her property rights. Likewise, a person's right to information or to help in need affect others partly independently of her behaviour. Importantly, these are things which we may want to affect for a person's good. We may

[31] Ibid., p. 55.
[32] Ibid. pp. 47, 55, 94.
[33] Ibid., p. 56.

for example want to seize a person's property (some drug say) to save her from temptation, withhold distressing information from her to spare her anxiety, or refrain from helping her to induce her to build character. In order to avoid an overly narrow focus on behaviour, we should therefore consider the idea of the self- and the other-regarding independently of exactly what things can be sorted into these categories.

The self-regarding is a sphere of life that *affects directly* only a person herself and consenting others. The typical consequentialist critique can be interpreted as a demand that the idea of directly affecting be qualified in a number of ways before it can be taken seriously. Most obviously, it must exclude minor harms and distant harms. Marginally polluting the natural environment while otherwise staying clear of the other-regarding must be considered an indirect effect on others, as must causing the deaths of unknown people in poor countries by neglecting to send them money through international aid agencies. Unless such effects are disregarded as indirect, the sphere of the self-regarding shrinks to nothing.[34] More substantial harms must also be excluded as indirect if they are the inevitable consequence of fair competition under scarcity. Going about my own business I will affect others by occupying a certain spot in the park or a certain job position. If good opportunities are few, this may have significant effects on others.

A more subtle qualification is that of moral or aesthetic sensitivity. In the words of CL Ten, interpreting Mill, 'an action indirectly affects others, or the interests of others, if it affects them simply because they dislike it, or find it repugnant or immoral.'[35] Similarly, effects that come about only through the satisfaction or frustration of external preferences, preferences 'for the assignment of goods or opportunities to others' should probably be deemed indirect.[36] Even with these qualifications, there remain more complex situations where one person suffers because of another person harming herself, irrespective of moral sensitivity or external preferences. Children may suffer greatly because their parents do not take proper care of themselves.[37] Hard cases include those where there is a delay between the effects on the person herself and the ensuing effects on others. Does my being poor dinner company because I neglected to get enough sleep last night affect you directly or indirectly? How about my being a poor provider for my family because I did not put enough effort into my college studies?[38]

---

[34] These qualifications need not invoke the distinction between acts and omissions, and Mill and Feinberg are no friends of this distinction. Mill, p. 15: 'A person may cause evil to others not only by his actions but by his inaction'. Cf. Feinberg, *Harm to Others* (Oxford: Oxford University Press 1984), Chapter 4.

[35] CL Ten, *Mill on Liberty* (Oxford: Clarendon Press 1980), p. 14. Mill does nonetheless at one point (pp. 108-109) seem to allow the prohibition of public violations of good manners.

[36] Ten, *Mill on Liberty*, p. 30. Ten is drawing on (and referring to) Ronald Dworkin, 'Reverse discrimination', in *Taking Rights Seriously* (Cambridge: Harvard University Press 1977) pp. 223-38.

[37] Mill does not consider such negligence self-regarding, but rather argues that 'if, either from idleness or from any other avoidable cause, a man fails to perform his legal duties to others, as for instance to support his children, it is no tyranny to force him to fulfil that obligation, by compulsory labour, if no other means are available.' (p. 108)

[38] Ten seems to believe that the need to invoke the self-regarding disappears once we realize that anti-paternalism is not about protecting 'an area of conduct' from intervention, but rather about ruling out 'certain reasons for intervention' (p. 40). While this is indeed something we should realize, it does not make the concept of paternalism less dependant on the idea of authority$_F$. As noted in the introduction, whether paternalism is the performance of certain actions or the invocation of certain reasons for certain actions, the action component must be specified.

We may conclude that there are difficult boundary issues to be faced by any attempt to delimit the self-regarding. This is a serious problem since it is unclear what sort of considerations should decide these matters, especially if the self-regarding is supposed to delimit the self-standing moral concept of $\text{authority}_{\text{F}}$. However, let us now for the sake of argument grant that these boundary issues can be settled in a satisfactory way and move on to the matter of relevance. The liberal tradition has taught us the concept of a private sphere which 1) does not significantly affect (non-consenting) others and 2) others have no business getting involved in. These two alleged properties may seem related but are in fact quite distinct. 1 implies 2 only if we assume that we have business getting involved only in things which affect us. We may, however, care, and care deeply, about things which do not affect us. Things that do not affect us may also be of great importance. Among these things are the good of other people. The idea of the self-regarding concerns 1. To get to 2 we need a host of moral assumptions, including anti-paternalism. It seems circular then to invoke the self-regarding to define paternalism and anti-paternalism. Let us, however, investigate the idea of the self-regarding in more detail by positioning it in the context of the positive side of Mill's liberty principle – the harm principle.

The harm principle says that self-protection, or the prevention of harm to others, is a good or valid reason for limiting or interfering with liberty.[39] The effects of our lives on the lives of others are obviously of crucial concern when considering the proper delimitation of this principle. These effects may be clarified by distinguishing five properties of parts of our lives. First, there is the property of being *self-regarding*, of affecting others only indirectly. Second, there is the property of being *consented to* – of affecting only people who consent to being so affected. Third, there is the property of being *self-protection*, of affecting people as a (proportionate) means of preventing them from harming others. If harm is thought of as setback to interests or some such only weakly moral concept, parts of life that have any of these first three properties may still be harmful to others. However, in the context of specifying the proper domain of the harm principle, having any of these properties entails that a part of life does not cause *wrongful* setback to interests and therefore does not harm others in the relevant sense.[40] Parts of life that lack all these properties are not, however, necessarily harmful. There is the fourth property of *innocence*, of being truly harmless, independently of the self-regarding, consent and self-protection. Actions with this property may include the giving of gifts and advice, working in the public interest, for example with the design and construction of city infrastructure, and insisting on having one's legal rights protected. Owning some kinds of property and having rights more generally may also be considered examples of innocence. All and only those parts of life that have none of the four properties thus surveyed have the fifth property of being *harmful* in the relevant sense.

The point of this exploration of the harm principle is to make clear the proper context of the self-regarding. What parts of life affect others only indirectly? What effects

[39] Following Feinberg, e.g. *Harm to Self*, p. ix. Sometimes 'the harm principle' refers to what I call 'the liberty principle' – that is the principle quoted above, saying that *only* harm to others warrants interference.
[40] Cf. Feinberg, *Harm to Others*, chapter 3.

on others are consented to? What effects on others are proper means of self-protection? What effects on others are truly harmless? Specifying the harm principle, or its precise application, amounts to deciding the exact reach of the five properties. The idea behind the harm principle is that only in those parts of a person's life that harm others may their interests justify making her the target of unwelcome involvement.[41] Let us assume that this idea is correct. Now how does this relate to paternalism?

Mill and Feinberg (and others) seem to pick two of the properties discussed – being self-regarding and being consented to – and declare that parts of life with either of these properties are unfit targets for benevolent unwelcome involvement. The problem with this strategy is that there simply is no reason why the manner in which a person is or is not affecting others should decide when we may promote her good against her will. Why, for starters, would it be more reasonable to promote a person's good against her will through affecting parts of her life that are innocent or self-protecting, than through affecting parts of her life that are self-regarding or consented to? Is it more reasonable to use force to make a person promote her own interests in her layout of the new city hall than in her composition of a private letter? Is it more reasonable to prevent a person from taking unnecessary risks when she is defending her friend from an assailant than when she is playing soccer? Are we proper targets for benevolent unwelcome involvement as soon as we step outside of our private sphere of non-direct effects and consensual interaction, into the public sphere of innocent interactions and self-protection? I propose that no anti-paternalist has ever intended for her principle to be limited in this way, and that it is unreasonable that it should be.

None of the activities in the examples just given fall under the harm principle. Perhaps interference should be defined not in terms of the self-regarding and consented to exclusively, but rather in terms of all non-harmful parts of a person's life? This would protect not only writing private letters and playing soccer, but also planning city halls and defending one's friends. The same objection still holds, however – there is simply no reason why the limits of benevolent involvement with a person should be drawn according to how her life affects others. Why would it be more reasonable to promote a person's good against her will just because she is harming others? That it would not may be less obvious than for the cases of innocence and self-protection, since in cases of harm to others there are strong reasons for some kind of unwelcome involvement. We should not, however, assume that if one kind of unwelcome involvement is justified, so is any other kind. It could perhaps be argued that when a person harms others she forfeits the special protection formerly enjoyed and becomes a proper target for unwelcome benevolent involvement.[42] This seems somewhat arbitrary, but perhaps paternalism is only at issue with regard to upstanding members of the community. Such forfeiture does not, however, explain why the person was protected from unwelcome benevolent involvement in the first place. She is supposedly an upstanding member of the

[41] It should perhaps be noted that the principle does not require that all harmful interactions be prevented. Sometimes harm should be allowed as the lesser evil, and sometimes preventing harm is too costly, either in absolute terms or in relation to the cost of preventing other harms by means of the limited resources at hand.

[42] This possibility was pointed out to me by Richard Arneson.

community simply in force of her not harming others, taking us back to where we started.

An example might clarify the argument against drawing the line at harming: After you report an on-going burglary in your home, the police arrive at the scene and confront the burglar. In choosing their path of action, the police may obviously be directed by their desire or duty to prevent people from harming each other. Concern for you and your property will be a typical and commendable motive. Should you become violent, concern for the health of the burglar might also be a relevant motive. However, the question we must now ask is whether the police may, without paternalism, go beyond the protection of people from each other and exercise power over the burglar against her will *for her good*. May the police, for example, be harsher on the burglar than is necessary for protecting you and themselves, in order to promote *her* best interest? Such benevolent harshness seems an obvious case of paternalism. Indeed, it much resembles the idea of punishing offenders in their best (moral) interest, something that has been advocated by Herbert Morris under the appropriate heading 'A Paternalistic Theory of Punishment'.[43] If benevolent harsh treatment of burglars and benevolent punishment and other cases of benevolent unwelcome involvement with inflictors of harm involve some kind of paternalism, then paternalism cannot be defined in terms of the self-regarding, nor in terms of non-harmfulness more generally.

In sum, a private sphere cannot be established by appeal to its sheer privateness. It must be established by considering every objection to it on its own merits. If the patchwork of restrictions that define the self-regarding is enough to establish that indirect setbacks to the interests of non-consenting others do not warrant unwelcome involvement, then we must move on to consider reasons that concern direct setbacks to the interests of consenting others, the failure to promote the interests of others, 'free-floating evils', the protection and the promotion of the good of the person herself, and so on. If these other rationales do not warrant unwelcome involvement in the proposed private sphere, this must be explained. In the case of moralism, it may plausibly be argued that the prevention of free-floating evils is not a good reason for action in general, because there are no such evils. In the case of reasons that concern failure to promote the interests of others and reasons that concern the good of the person herself, this is not a plausible answer. A plausible reason to establish a sphere of authority that may perhaps be called private, is that the freedom to choose to do what does not promote the various goods listed is more important than are the net benefits of unwelcome involvement. This, however, is a case of balancing of values and not a case of principled anti-paternalism.[44]

---

[43] (1981) Reprinted in Rolf Sartorius (ed.) *Paternalism* (Minneapolis: Minnesota University Press 1983).

[44] A private sphere can of course be derived from some more general deontological framework where certain considerations are assumed to be lexically prior to other considerations. For example, the rights of self-ownership and ownership resulting from just acquisition and just transfer may be assumed to trump all other concerns, which are in that sense not significant (Robert Nozick, Anarchy, State, and Utopia [Malden MA: Basic Books], chapter 7). The private sphere can also be established by direct assumption, as seems to be the case when Arthur Ripstein ('Beyond the Harm Principle', *Philosophy and Public Affairs* 34[3] [2006] 216-246) declares that everyone has the right to control 'her own powers' and so unwelcome involvement with those powers is wrong regardless of its rationale. Ripstein argues against the harm principle for this

SUBSTITUTION OF JUDGEMENT AND DIMINISHING OF CHOICE

Definitions of interference in terms of liberal values depend on an account of authority$_F$ that cannot be provided by appeal to the self-regarding or non-harmfulness. Can this dependence be avoided by operationalizing the values into more concrete notions? Gerald Dworkin at one point followed Mill in defining paternalism in terms of 'interference with a person's liberty of action',[45] but later proposed an alternative account, apparently impressed with the objection that the focus on behaviour is too narrow. Presenting the alternative account, Dworkin in a few short paragraphs invokes a sequence of notions – 'to interfere with self-determination', 'violation of autonomy', 'usurpation of decision-making', 'denial of autonomy' and 'to treat others as means'.[46] These notions are all pregnant, but not very specific. A somewhat more specific definition is offered when Dworkin states: 'What we must ascertain in each case is whether the action in question constitutes an attempt to substitute one person's judgement for another's'.[47] This proposal too may be criticised for being too narrow, since it would seem that blocking some alternative actions (the most harmful ones) does not amount to substitution of judgement. If I hide the sleeping pills from my suicidal friend, or if the government prohibits heroin use, this does not amount to substituting one persons judgement for another's, yet would typically be regarded as paternalism (if benevolent).

For our present purposes, what is important is that Dworkin's proposal is too *wide* as it stands. Dworkin rightly rejects the overly lax idea that paternalism is simply acting benevolently against someone's wishes, in favour of the somewhat stricter account in terms of substitution of judgement. However, his account faces similar problems. You may well form judgements about how my property should best be used and my life lived. You may judge, for example, that I should spend more (less) time with you and give you more (less) expensive gifts. Substituting my judgement for yours in these matters, where you have no right to decide, cannot reasonably amount to interference. Your decisions about my life will be forced by my actions, yet this is morally unproblematic. What is needed to achieve the desired scope is some further criteria for identifying when forcing a decision amounts to substitution of one person's judgement for another's. Dworkin does not provide such criteria, and I see no obvious or straightforward way of generalizing to such criteria from the examples of paternalism and non-paternalism that he does provide: Dworkin tells us that a daughter refusing to acquire the education her father wants her to acquire, because she believes that her future success in the profession (that is also her father's) would ultimately embarrass her father, behaves *non-paternalistically*, while a husband hiding his sleeping pills from his wife, because he fears she might otherwise use them to commit suicide, as well as a tennis player refusing to

principle of sovereignty, which he claims will provide a better basis for anti-paternalism (p. 245). However, this principle is not distinctively anti-paternalist but quite general, and it seems to rely on a direct appeal to freedom as authority$_F$, with no attempt at defining its limits.

[45] Dworkin, 'Paternalism', p. 65.

[46] Dworkin, 'Some second thoughts', p. 107.

[47] Ibid.

play another game with her opponent, because she wants to avoid further upsetting her by winning, both behave *paternalistically*.[48] The father is refused the option of influencing his daughter's career plan, while the wife is refused the option of using her husband's sleeping pills and the opponent the option of playing another game. In all cases, these options are refused the person in question for her own good.

It would seem that Dworkin relies on an implicit account of some kind of authority such that interference is substitution of judgement *as it regards matters under a person's authority*. Apparently, having access to her husband's sleeping pills is within the wife's authority, and having her opponent choose sporting activities independently of her own future frustration is within the tennis player's authority, while having his daughter choose careers independently of his own future embarrassment is not within the father's authority. This is perhaps reasonable, but one would like to know what kind of authority is taken for granted and preferably be offered some general account of when something is within a person's authority so understood. Making the invocation of authority explicit may seem redundant, if the notion of *substitution* of judgement is taken to include as a presumption that the judgement concerns something that is under another's authority. Making this presumption explicit is, however, precisely the point. In order to decide whether some action is or entails a substitution of judgement, we must first decide who has authority over that which the action affects.

In order to avoid excessive narrowness (considering the fact that blocking some actions do not fully substitute judgement), substitution of judgement could be understood in the weaker sense of merely limiting a person's choice, without actually making judgements or decisions for her. David Archard has proposed that one condition of P behaving paternalistically towards Q is that 'P aims to bring it about that with respect to some state(s) of affairs which concerns Q's good Q's choice or opportunity to choose is denied or diminished.'[49] On a broad enough understanding of choice, this specification is arguably not too narrow. However, just like substitution of judgement, it is too wide, unless it relies on an implicit invocation of some kind of authority. Archard points out that his account can accommodate Dworkin's example of the sleeping pills. The husband's hiding of *his* sleeping pills from his suicidal wife diminishes *her* opportunity to choose to use them. However, Archard does not seem to notice that this case is only the tip of an iceberg. Most any use of one's property diminishes other people's opportunities to use it for themselves. However, it is arguably not paternalism for P to diminish Q's choice for Q's good in areas in which Q has no right to choose in the first place. Q's option of using P's property could concern Q's good, but under normal circumstances it cannot reasonably be interference if P brings it about that Q cannot use P's property.

[48] Pp. 106-7. The often cited example of the wife and the sleeping pills is unfortunate, since it hinges on our intuitions about property rights, while it is placed within the institution of marriage, which is often taken to dissolve individual property rights to some extent. We agree that hiding the pills does not violate a right or a moral rule, if we do, because the pills *belong* to the husband. We agree that this is nonetheless a case of interference, if we do, because certain belongings should be shared between spouses.
[49] 'Paternalism defined' , p. 36.

When it comes to justification, Archard says very briefly that paternalism 'ignores' the duty 'to respect the choices of others' and that the question is whether 'the duty to care for others take precedence' over the first duty. He does not say whether the duty to respect the choices of others extend to respecting their choice as regards anything, or if it is rather limited to respecting those choices that are under their authority_F (choices that are under their authority_C should be respected by definition – there can be no question about that). In the first, wide case, the value of any person's choice or influence over any state of affairs must simply be one of the considerations that determine the proper limits of authority_C. In the second, narrow case, Archard again assumes an implicit account of authority_F.

In an even more lax definition, Donald Van de Veer proposes that an action is an interference if the actor deliberately acts 'contrary to the operative preference, intention, or disposition of the subject' (or if she shapes or modifies these preferences in certain ways).[50] However, without your consent and against your preference, I may for example give something up that is mine to give – I may step out of your way, withdraw my candidacy, leave my spot in the public garden. Van de Veer himself provides a similar example - planting roses in my garden for my neighbour's sake.[51] His solution seems to be to accept this very wide definition and treat as problematic only interferences that are presumptively wrong, thus in fact committing himself the wrongness account, to be explored below.[52]

Admittedly, we *could* accept the substitution of judgement or the diminishing of choice account of interference without qualification. It could then be paternalism to act within one's right, for the good of another. Using my property would be an interference with you whenever this use was in opposition to your judgement or when it diminished your options. On the substitution of judgement account, clearing a road through my forest to ease your access to the lake would be paternalism if, according to your judgement, the road should have been cleared in some other fashion, or not cleared at all (perhaps due to environmental concerns). On the diminishing of choice account, clearing the road would be paternalism if it diminished your opportunities to, for example, walk in my forest without coming upon a road, or use the lack of a road as an excuse not to take your cousin fishing. Such a broad account of paternalism is not incoherent. Archard comes close to explicitly accepting this account in a later article: 'Paternalism can be the failure to perform a customary or expected action which need not be obligatory.'[53] The mere fact that people usually do not clear roads through forests like mine, or that they do not expect me to do so, would make my action interfering. These wide accounts of paternalism, however, do not set it out as a morally relevant category.

The connection between substitution of judgement and authority is made explicit in Seana Shiffrin's account of paternalism. Shiffrin, like Dworkin, defines paternalism in terms of substitution, though she argues that the substitution need not

[50] 'Paternalistic Interference', pp. 18-19.
[51] Ibid., p. 17.
[52] Ibid., p. 20-21.
[53] 'For our own good', *Australasian Journal of Philosophy* 72(3) (1994) 283-93, p. 288.

concern judgement specifically, but may alternatively concern agency, such as when one person promotes another's good against her will not because they disagree about what is best for her, but because the good promoter is (or believes herself to be) more able to act on their shared perception of what is best.[54] This is a very reasonable extension. More importantly for our present purposes, Shiffrin adds that in addition to substitution, paternalistic behaviour towards a person B must be 'aimed to have (or to avoid) an effect on B or her sphere of legitimate agency', and must be 'directed at B's own interests or matters that legitimately lie within B's control'. The notions of legitimate agency and legitimate control are never specified by Shiffrin. She explicitly admits that 'a full account of paternalism will depend on an account of what sorts of interests and matters legitimately lie within an agent's control' and she does not provide such an account.[55]

Like Dworkin, Shiffrin does provide a number of examples. One is similar to that of Dworkin's tennis player – Shiffrin claims that it is paternalism to refuse to help an acquaintance assemble some shelves (when one is under no obligation to do so) because one thinks that this will help her develop her own skills or her confidence in her skills.[56] It would seem then that on Shiffrin's account one has authority to get help from one's friends independently of their concerns for one's best interest. It is also paternalism, according to Shiffrin, to refuse to help a person get exploited by a third party, unless this is motivated by self-interest in a narrow sense.[57] It would seem that one has authority to get support from others independently of their evaluations of outcomes, except as regards their narrow self-interest. Unless two persons can have (conflicting) authority over the same thing, this very generous understanding of authority entails a corresponding restriction – one has no authority to decide according to one's own values when to help others, even when that which one would choose is not only harmless, but beneficial. This is an uncommon and peculiar delimitation of authority, whether understood as authority_F or authority_C.[58]

At this point, something needs to be said about attitudes. On Shiffrin's account, a further condition on paternalistic behaviour is that it is undertaken on the grounds that the agent 'regards her judgement or agency to be (or as likely to be), in some respect,

---

[54] Shiffrin, 'Paternalism, Unconscionability Doctrine, and Accommodation', p. 215. Dworkin would presumably not call this paternalism.

[55] Ibid., p. 218.

[56] Ibid., p. 213. The example is not as instructive as it could have been since Shiffrin only considers the two alternatives of either honestly and successfully persuading the friend to assemble the shelves herself, which she deems non-paternalistic, and baldly and without explanation refusing to help, which she deems paternalistic. She does not consider the obvious intermediate alternative of refusing to help and explaining why, while failing to convince the friend of the merits of this choice. It is therefore not quite clear if it is the lack of forthrightness and honesty involved in baldly refusing that makes this alternative paternalistic according to Shiffrin, or if it is enough to simply not help, for the reasons in question.

[57] Ibid., p. 227.

[58] De Marneffe ('Avoiding paternalism', pp. 77-79) argues against Shiffrin's account that any decision to influence someone may be seen as substituting one's judgement for hers, regardless of whether it is aimed at her good or the common good. Substitution of judgement to protect others from harm is not problematic and therefore, de Marneffe argues, neither is substitution of judgement to protect the person herself. Husak presents a similar argument ('Paternalism and Autonomy', pp. 41-42). My argument, in contrast, concerns the limits of authority and does not depend on the claim that there is no morally relevant distinction to be made between private and common good in this context.

superior' to that of the person acted towards.[59] Similarly, on Archard's account, P behaves paternalistically towards Q only if 'P discounts Q's belief that P's behaviour does not promote Q's good.'[60] Clearly, if paternalism is to have any moral content, it must be in some sense unwelcome. Responding to someone's request for help is not paternalism.[61] However, Shiffrin and Archard go further. Indeed, it is a central part of Shiffrin's account that paternalism entails an attitude of disrespect toward a person's judgement or agency. This attitude 'is central to accounting for why paternalism delivers a special sort of insult to competent, autonomous agents'.[62] Shiffrin writes of this attitude as if it was a motive, as if sheer disrespect could move us to act.

Attitudes are distinct from motives. It is one thing to promote your good against your will, it is another thing to disrespect your judgement or discount your view of the good. Involvement is unwelcome, when it is, for a reason. A person may resist involvement because she believes it to be bad for her, because it makes her feel insignificant, because she simply dislikes this particular kind of involvement, or for any other reason. Undertaking or approving of unwelcome involvement fails to give these reasons priority, it 'discounts' or 'disrespects' them in the sense that it does not attribute to them the high relative impact that the person herself attributes to them. Such lack of agreement on the relative impact of reasons, however, hardly amounts to a patronizing or otherwise bad attitude. The cool and correct judgement that I know better than you what will best promote your interests in some context need not involve an insulting attitude of disrespect, nor a discounting of your views on the matter. I may simply have more experience or insight, and know it.[63] Moreover, I may on other occasions act spontaneously, out of pure compassion, or alarm, without stopping to consider how you judge the situation, what you deem to be your good, or how you would have acted had I passively stood by.

Attitudes of superiority and condescendence, of disrespect of others' judgement and agency and view of the good, are not necessarily tied to benevolent interference, but may be shown in all kinds of situations and affect all kinds of decisions and actions. I may for example favour others over you because I (correctly or incorrectly) find your judgement inferior or your will weak. I may similarly disregard your views on your good or on any other matter because I do not trust your judgement. I propose that while these attitudes are interesting in their own right and obviously related to paternalism, they do not define the concept. Paternalism is based in action and in reasons for action, such that a person can be judged to act paternalistically in force of her motives for or the possible justifications of her actions, regardless of her attitudes.

Leaving attitudes to one side then, the trouble with Dworkin's, Archard's and Shiffrin's accounts of paternalism is that they are either fatally wide, or depend on an

[59] Shiffrin, 'Paternalism, Unconscionability Doctrine, and Accommodation', p. 218.
[60] Archard, 'Paternalism defined', p. 36. This condition seems to entail that it is not paternalism to coerce a person, against her will, to maximize what is good for her according to her own judgement.
[61] Unless, perhaps, the request for help is not voluntary.
[62] Shiffrin, 'Paternalism, Unconscionability Doctrine, and Accommodation', p. 220.
[63] Another example may be self-paternalism, which arguably most often does not involve an attitude of superiority or disrespect, see Husak, 'Paternalism and Autonomy', pp. 43-45.

implicit, or in Shiffrin's case explicit but not explicated, idea of authority. Without delimitation to a sphere of authority, the accounts do not define a class of actions that are morally problematic when combined with benevolent reasons. That an action benefits a person and that it is not what she would have decided had she been in control are two common properties of actions and their frequent co-instantiation in one action raises no special concerns. In particular, it seems outrageous to oppose paternalism so defined. Why would it be morally wrong to do something just because it benefits you and is not what you would have decided be done? It is doubtful whether your judgement should have any moral weight unless my action interferes with you in some further sense. Assume, in line with standard liberal theory, that some resources are mine to spend, regardless of what others think (as long as I do not harm them). If I may spend what is mine simply because I feel like it, why may I not spend it because this would be (very) good for you? How could we justify such a restriction on people's motives for spending what is theirs to spend?

If, on the other hand, the accounts presuppose an account of authority_C, such that substitution of judgement (or agency) and diminishing of choice should be understood to take place only within this sphere, then nothing hinges on when something is paternalistic, since this concept takes for granted rather than influences the verdict of who should be in control of what. Indeed, Dworkin and Archard seem more concerned to investigate paternalism in light of linguistic intuitions, than to investigate its moral aspects. In his later article, Archard comes very close to abandoning the notion of paternalism as a distinct moral problem.[64] This is of course exactly what we should do. Dworkin is more prone to stick with the traditional position that paternalism is always prima facie wrong, but diminishes the importance of this wrong to almost nothing.[65] In conclusion, to establish paternalism as a morally relevant category, we need an account of authority_F or something close to it. Lacking such an account, the proposed definitions of interference in terms of operationalizations of liberal values are incomplete at best.

COERCION

Neither the liberal values, nor the concept of the self-regarding, nor operationalizations of the liberal values in terms of choice, decision-making, judgement or agency, give us a satisfactory account of the limits of authority_F, and so do not provide a definition of paternalism as a morally relevant category in its own right. Can coercion do any better? Can it either define authority_F, or help us do without it? I propose that it cannot.

Benevolent coercion is admittedly a paradigmatic case of paternalism. Before going on to advance the account discussed in the previous section, Dworkin defends his earlier, Millian specification of interference as interference with liberty of action, saying it

[64] David Archard, 'For our own good'.

[65] Dworkin, 'Some second thoughts', p. 110: 'In the final analysis, I think we are justified in requiring sailors to take along life-preservers because it minimizes risk of harm to them at the cost of a trivial interference with their freedom.'

is reasonable given the interest to investigate 'the proper limits of state coercion'.[66] On 'standard views of liberty', according to Dworkin, if something is not coercion, 'or force', it is not a restriction of liberty.[67] It is true that on many accounts, liberty is distinguished from ability, so that the lack of possibility or opportunity to do something is a lack of liberty to do so only if it is (intentionally) caused by a person (or other agent). However, it is uncommon to claim that all lack of liberty is a result of coercion. Closing a door is not normally considered coercive, even though a person is thereby not at liberty to enter. If we do nonetheless equate coercion with interference with liberty, the former concept is stretched beyond its normal use and so the latter is doing all the conceptual work and so we are back to defining interference by direct appeal to liberal values.

Despite its popularity in examples and applications, I know of no explicit definition of paternalism in terms of coercion, beyond Dworkin's half-hearted defence just mentioned. This is perhaps because upon closer scrutiny the concept either, on a more normative explication, reduces to one of the other strategies in this survey, or, on a more conceptual explication, involves specific behaviour such as forcing and threatening, which makes it unsuitable for defining interference. I will nonetheless briefly consider some approaches to coercion and their potential to distinguish paternalism as a morally relevant category.

On some accounts, coercion is all but equivalent with substitution of judgement. The idea is that coercion is enforcing a certain outcome or decision irrespective of protests or preferences to the contrary. In the *Stanford Encyclopedia* article on 'Coercion', Scott Anderson quotes JR Lucas:

> Force, then, we say, is being used against a man, if in his private experience or in his environment either something is being done which he does not want to be done but which he is unable to prevent in spite of all his efforts, or he is being prevented, in spite of all his efforts, from doing something which he wants to do, and which he otherwise could have done by himself alone. A man is being *coerced* when either force is being used against him or his behaviour is being determined by the threat of force.[68]

According to Anderson, this account captures the dominating conception of coercion in pre-modern and modern political theory before Nozick's seminal article on coercion appeared in 1969.[69] As with substitution of judgement and diminishing of choice, forcing someone on this account may or may not be morally problematic depending on what is being forced. You may be in your full right to use force to control what is rightfully yours to control – only in matters over which others should have some control is using force

[66] Ibid., p. 105. In the earlier, Millian definition Dworkin in fact used both 'interference with a person's liberty of action' and 'the person being coerced', assuming the two notions to be equivalent. ('Paternalism', p. 65)
[67] Ibid., p. 106.
[68] *The Stanford Encyclopedia of Philosophy (Spring 2006 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2006/entries/coercion/>. Original quote from *The Principles of Politics* (Oxford: Oxford University Press 1966), p. 57.
[69] 'Coercion' (1969), reprinted in *Socratic Puzzles* (Cambridge MA: Harvard University Press 1997) 15-44.

morally problematic. Assuming that we may do as we please in certain areas that we rightfully control, we may also use force in those areas benevolently if that is what we want, without raising special moral concerns. For example, assume that you have the right to make someone leave your property by force. Now if you do this not for your own sake but for the sake of the intruder (perhaps because you do not want her to be attacked by your dogs), this is not morally problematic. Defining interference in terms of this account of coercion is therefore fatally wide or presupposes an account of authority.

This by now familiar problem of wideness will apply to any account of coercion in terms of force or pressure.[70] For example, Feinberg's account of coercion in large part concerns diminishing of choice and modification of the 'price tags' of different options by application of 'coercive pressure'.[71] In addition, such accounts may face problems of narrowness. In their critique of Dworkin's old definition, Bernhard Gert and Charles Culver offer counter-examples to limiting paternalism to acts that constitute 'an attempt to control the behaviour of the person': Merciful killing, transfusion of blood to an unconscious Jehovah's Witness, and benevolent withholding of information all seem to be cases of paternalism and actions to which the liberal anti-paternalist would object.[72] These examples cannot be accommodated by an account of coercion in terms of pressure or force applied to behaviour, though perhaps by the account suggested by Lucas' formulation 'something is being done which he does not want to be done'.

Lucas' traditional account defined coercion as the use of force *or* the threat of force. Today, following Nozick, coercion is most often thought to be confined to threats or, more neutrally, proposals.[73] Coercion confined to threats may supply a definition of interference that escapes the wideness objection by not depending on an account of authority. It need not be within my authority to threaten to do things that it is within my authority to do. Perhaps then I am never within my rights to threaten you and so whether or not a threatening is benevolent could in principle be morally relevant. However, whether a part or full definition of coercion, threats seem ill suited to define paternalism as a morally relevant category. Why would benevolent involvement through threats be particularly morally problematic? How does threatening differ from manipulation (fraud) or property violation (theft) in its relation to benevolence? Is it paternalism to make you stop smoking by threatening to steal your scarf, but not by actually stealing your pipe? Is it paternalism to stop your suicide by threatening to bury

---

[70] William A. Edmundson considers and rejects 'pressure theory' as an account of coercion in *Three Anarchical Fallacies*, Cambridge: Cambridge University Press 1998, pp. 97-98. Though this rejection is conceptually sound, social pressure may certainly compete with coercion as regards usefulness in explaining human interactions.

[71] Feinberg, *Harm to Self*, chapter 23. Feinberg's discussion of coercion is intended to specify voluntariness.

[72] 'Paternalistic Behavior', p. 46-49.

[73] Nozick's account draws on HLH Hart and AM Honoré (*Causation in the Law*, Oxford: Clarendon 1959) and on Hart (*The Concept of Law*, Oxford: Clarendon 1961). Joseph Raz proposes a threat-based modification of Nozick's account in *The Morality of Freedom* (Oxford: Oxford University Press 1986), pp. 148-9. For Raz, coercion essentially entails an invasion of autonomy and a justification or excuse for the coerced person. Much of contemporary discussion concerns 1) whether such normative properties of coercion are essential to the concept and 2) how threats may be specified in terms of the threatened consequence causing a deviation from some expected (predicted, required, preferred) baseline. How these issues are resolved do not affect my argument.

you on unholy ground, but not by deceitfully convincing you that other people would do so? This seems completely arbitrary. This arbitrariness is not lessened by including as coercive proposals some offers, as has been suggested on some accounts.[74]

Feinberg develops his anti-paternalism within the context of his greater project of investigating *The Moral Limits of the Criminal Law*.[75] The criminal law, Feinberg assumes, is coercive in that it is 'meant to control behaviour by threat.'[76] The general restriction to criminal law is motivated in part by the fact that legal punishment potentially entails an 'immense destructive impact' on the individual.[77] As we have seen, threats and behaviour control in general are not suitable to define interference. Further restricting interference to the criminal law does not make them any more suitable. Drawing the line at criminalization is hopelessly arbitrary since what is criminalized and what criminalization entails differ amongst legal systems. Criminal punishment has no inherent, morally relevant properties that distinguish it from other sanctions.[78] There is no morally relevant difference between on the one hand decrees that are upheld by the threat of small fines administrated by the police, such as laws requiring motorists to wear safety belts or bikers to wear protective helmets, and on the other hand fees that are administrated by other government agencies or contracted entrepreneurs, such as parking tickets or fees for using mass transit without a ticket (which are not legal punishments in many countries).[79]

The appeal of Feinberg's restriction and of coercion accounts more generally may lie in the fact that coercion is supposed to be a severe type of intrusion or limiting of liberty. The most extreme form of coercion is arguably not legal punishment but compulsion – forcing another person to do a specific thing, or rather forcing her to be in a certain situation with no possibility of doing anything about it.[80] Compulsion is perhaps never within anyone's rights, as opposed to enforcing outcomes or preventing actions which concern things that one rightfully controls. Furthermore, while some instances of other kinds of actions, such as manipulation and material sabotage, are certainly more

---

[74] E.g. Feinberg, *Harm to Self*, chapter 24; David Zimmerman, 'Coercive wage offers', *Philosophy and Public Affairs* 10(2) (1981) 121-145; Daniel Lyons, 'Welcome Threats and Coercive Offers', *Philosophy* 50 (1975) 425-436.

[75] The four-volume work of which *Harm to Others* is the first and *Harm to Self* is the third.

[76] *Harm to Others*, p. 22. The (traditional liberal) presumption that law is coercive is convincingly challenged by William A Edmundson in *Three Anarchical Fallacies* (Cambridge: Cambridge University Press 1998).

[77] Feinberg, *Harm to Others*, pp. 3-4.

[78] One could perhaps invoke the idea that legal prohibition always entails moral condemnation, though this would rise the question what is particularly problematic about benevolent moral condemnation. Feinberg for one does not invoke this idea. For a discussion see William A Edmundson, 'Comments on Richard Arneson's "Joel Feinberg and the Justification of Hard Paternalism"', *Legal Theory* 11 (2005) 285-291.

[79] The moral basis of Feinberg's anti-paternalism rests with the concepts of autonomy and sovereignty rather than with (legal) coercion. It is not clear what role the restriction to criminal law plays in his opposition to paternalism. Feinberg states that different liberty-limiting principles are needed for different forms of state policy, and that his anti-paternalism does not hold for regulatory and taxing policies (even though they too are coercive). (*Harm to Others*, pp. 23-24; *Harm to Self*, p. 134) There is a tension between this restriction and the alleged absoluteness of autonomy and sovereignty.

[80] Nozick mentions that Christian Bay has claimed that 'all infliction of violence constitutes coercion' but finds this over-inclusive since simply beating up or killing a person is not coercion, though implicitly threatening to do so may be. (p. 20) Michael Bayles distinguishes between 'occurent' and 'dispositional' coercion, where the former is physical compulsion. ('A concept of coercion' in *Nomos XIV: Coercion* [Chicago: Aldine 1972] 16-29, p. 17)

destructive than some instances of compulsion, we may perhaps delimit the concept so that all cases of compulsion are very destructive, or harmful, or damaging to autonomy, or some such thing. Might these properties of compulsion, possibly shared with some other specifications of coercion not considered here, make benevolent compulsion (etc.) a morally relevant category? Even if they would, this would be a very narrow definition of interference and so of paternalism, which would exclude most cases commonly discussed under this heading. More to the point, however, benevolent compulsion is not a morally relevant category.

A serious intrusion entails a large cost or loss and so there is a strong reason against that intrusion. Let us assume that our reasons against compulsion are so great that they are never outweighed by even the strongest benevolent reasons (such as survival or preservation of long term autonomy). In other words, the value of non-compulsion is always greater than the value of any other good for the person. If so, benevolent compulsion is indeed prima facie wrong (prima facie since other consideration such as harms to third parties may still make it justified). That is, benevolent reasons are always insufficient to justify compulsion. However, it is not the fact that benevolent and insufficient reasons for compulsion are benevolent that makes benevolent compulsion unjustified, but the fact that they are insufficient. There are many other potential rationales for compulsion, such as prudential reasons (call prudential interference 'prudentialism'), reasons to benefit others ('benefitism'), reasons to entertain oneself or others ('entertainism'), and reasons to diminish 'free-floating evils' (moralism). If benevolent compulsion is prima facie wrong, then, arguably, so is compulsion for any of these reasons. Similarly, other narrowly defined values give rise to other lists of prima facie wrongs. One example may be the destruction of one culture for the expansion and material gain of another – call this colonialism. If colonialism is prima facie wrong this is not because of some special relationship between the value of the non-destruction of a culture and the value of expansion and material gain, but simply because anyone's reasons against destroying a culture are always stronger than her reasons for achieving expansion and material gain.

WRONGING

We have reached the fifth and last account of interference to be considered. Gert and Culver have proposed that an interference with a person is an action believed by the agent to be beneficial but a violation of a moral rule with regard to that person, and where the agent believes herself to be qualified and justified to act on the person's behalf irrespective of her consent.[81] The belief qualifications makes this account very subjectivist – a benevolent wronging is not paternalism unless believed to be wrong. Let us disregard this qualification, since paternalism can arguably consist in any combination

[81] And the agent believes that the person knows in general what is for her good, something we can assume is true if the person is adult, mature, well-informed, sober, calm and collected and under no undue influence, as assumed in the introduction. Gert & Culver, 'Paternalistic behavior', pp. 49-50.

of actual and believed interference.[82] The condition that the paternalist be acting on the person's behalf sounds a lot like substitution of judgement. However, Gert and Culver seem to understand by this condition only that the paternalist must believe herself to be a more competent judge than the person interfered with, in force of some general distinction, such as that between professional and layman, between normal and retarded, or between sober and drunk, the purpose seemingly being to exclude cases where the agent is obviously less competent, such as 'a small child'.[83] I fail to see why this restriction is needed and it seems anyway rather unimportant and so I shall disregard it too. With these modifications, Gert and Culver's proposal comes close to identifying interference with wronging.

For another proponent, Van de Veer, as noted above, defines what he calls 'interference' very widely but proposes that benevolent interference requires special justification only when it involves a 'presumptive wrong'.[84] Actions are presumptively wrong 'because they invasively interfere by transgressing independently specifiable moral principles'.[85]

One advantage of a wronging account is that many different kinds of actions can be counted as interference. The wronging need not be an attempt to control behaviour, but can be a case of 'killing, causing pain, disabling, depriving of pleasure, deception, or breaking a promise.'[86] Another advantage, more pertinent for our present purposes, is that there seems to be no need for a sphere of authority if we can invoke directly the more general distinction between right and wrong. We have seen above that we are sometimes within our rights to affect people in ways which limit their liberty broadly construed, in ways which reduce their choice or options, and maybe even in ways which coerce them. With an appropriate account of wronging, however, we are never, by definition, within our rights to wrong someone.

To claim that it is morally problematic to wrong someone for certain reasons is not necessarily circular – one kind of wronging can be part of another. However, we do need an account of the kind of wronging in question, since it cannot be all things considered wronging, lest paternalism be morally irrelevant. There seems to be two possible solutions – either wronging is understood as some sort of prima facie wronging, or wronging is understood as an *almost* all things considered wronging – all things, that is, except the import of paternalism itself.

Pursuing the first solution, Gert and Culver take the violation of a moral rule to be performing an action that is prima facie wrong, or wrong 'unless one has an adequate

---

[82] If paternalism is morally relevant at all, it should be paternalism in some interesting sense to approve of or sanction paternalistic interference with a person because it benefits her, regardless of the motives of the interfering person. See Grill, 'The Normative Core of Paternalism'.

[83] Ibid., pp. 50-51.

[84] *Paternalistic Interference*, pp. 19-21, 88. Van de Veer excludes wrongs to third parties but explicitly makes room for presumptive wrongs to the agent herself, entailing that such wrongs can be justified by appeal to the current or hypothetical informed and non-impaired consent *of the person acted towards* (such appeal forms the core of Van de Veer's principle of justified paternalism – or 'Autonomy-Respecting Paternalism [p. 89]). This is probably a slip.

[85] Ibid., p. 21.

[86] Ibid., p. 51.

justification'.[87] On this account paternalism can have moral impact – a prima facie wronging may be more wrong, or amount to an additional wrong, if it is motivated or (allegedly) justified by benevolence. A major problem with this account is that it entails that an action can be an interference even if the prima facie wrong it amounts to is counterbalanced by another moral rule, or by another instance of the same moral rule. Such a counterbalancing consideration can be based on obligations to the same person or on other obligations. In the latter case, it may be that my action prima facie wrongs you but is justified by third party interests, such as justice or the preservation of the natural environment. If I then act for your good, this is paternalism. In the former case, I may face a situation where all available options prima facie wrong you, though one is less wrong than the others or at least morally preferable in some sense. If, for example, I promised you before you left town not to speak to the authorities, I may have to either betray your confidence and let the authorities know of your innocence, or keep silent and cause you to be innocently (and unwillingly) convicted in your absence. Now, if I act on the preferable option for your good, I am submitting you to paternalism. Paternalism thus understood does not seem morally relevant. Why should it matter morally, when I do something that is right, for your good, whether what I do is right because there are no conflicting obligations, or is right because conflicting obligations are outweighed? The conflicting obligations have moral importance in the weighing of course, and may entail residual obligations even after they are outweighed, but why should the right action have the additional tainting of being classified as an interference, assuming this classification has moral relevance?

The second solution avoids this problem by counting as wrongings only those actions that are wrong almost all things considered. All things, that is, except the impact of paternalism itself. Only after paternalism has been considered do we arrive at the final judgement of right and wrong. In other words, paternalism takes as its input a list of all those wrongs that are independent of it, and produces as its output a complete list of wrongs.

A possible problem for the wronging account appears in different but similar forms in both variations. On both accounts, paternalism is rendered impotent in a sense. On the almost all things considered account, the moral impact of paternalism cannot change the status of an action from right almost all things considered to all things considered wrong, since only those actions that are wrong almost all things considered can be paternalistic. In other words, an action which is wrong by a very small margin independently of paternalism can be wrong by a larger margin because of paternalism, but an action which is right by a very small margin independently of paternalism cannot be wrong by a very small margin because of paternalism. This seems slightly incoherent, or at least an unfortunate restriction.

On the prima facie account, the impotence is of a different kind – the moral impact of paternalism cannot make something prima facie wrong. Since interference is benevolent prima facie wronging, paternalism is parasitic on the pre-existing wrong. So

[87] Ibid.

for example, unless it is prima facie wrong to tax people (to raise revenue for national defence say), it cannot be paternalistic to design the tax system so as to encourage a healthy or morally high-standing lifestyle. This kind of impotence may simply be accepted.[88]

Finally, the two variations share a common weakness. Since, on the wronging account, paternalism only adds (negative) moral content to actions that are morally wrong independently, it presupposes a more or less complete moral theory. It is hard to give a final verdict on the account before such a theory has been developed. This also creates indeterminacy in concrete cases. For example, you may ask me as a friend to aid you in some shady business. Benevolence might lead me to aid you or to refrain, depending on what I think is your best interest. Now, if I refrain and this would be to wrong you (since you are my friend and I can help you at little cost to myself) I submit you to paternalism, and if I aid you and this would be to wrong you (since I facilitate your shady business) I submit you to paternalism. Consequently, it may be that I submit you to paternalism if I aid you, or if I refrain, or, if the wrongs are prima facie or if there are moral dilemmas, whatever I do (since you are my friend but it is a shady businesses). This shows that what is paternalism on the wronging account depends entirely on the background moral theory in a way that makes it hard to apply the concept to concrete cases.

Furthermore, while the wronging account allows that paternalism have moral impact in certain limited ways, the moral status of paternalism cannot help decide what kind of moral theory is the right one more generally. Is morality in general Kantian or utilitarian? On this issue, paternalism is quiet. All it says is that whatever turns out to be (prima facie) wrong, whether it is failure to maximize happiness or treating people as bare means, to act wrongly out of benevolence has certain moral properties. This is not incoherent, but one would think that a position on the moral status of paternalism should affect and be affected by other moral positions. On the wronging account, however, paternalism takes on a rather peculiar role, appearing as a sort of extra consideration at the end of moral evaluation.

CONCLUSION

Many difficult issues surround the attempt to define paternalism. Among them: Is it the interference with a person that is paternalistic or is it the invocation of her good as a reason for interference? Must the interference be actual? Must it be intended? Must the benefit be actual and/or intended? Must the interference be targeted directly at the person benefited? Must the person interfered with be competent? Must she be mature, informed, rational? Must the paternalist's view of the good for the person conflict with her own views? Must the paternalist display certain attitudes? However these question are answered, a principled account of paternalism must say something about interference – about what kind of actions are paternalistic when performed for the good of a certain

[88] In fact, Mill for one accepts sin taxes since the state needs revenue, and so it is 'the duty of the State to consider, in the imposition of taxes, what commodities the consumers can best spare' (p. 112).

person. If the account is intended to concern something more than our linguistic practice, the specification of interference must point to a type of action that, in combination with benevolence, is morally relevant in its own right.

My survey of the most common strategies for providing such a specification of interference has found them wanting. The liberal values of liberty, autonomy and sovereignty tend to oscillate between pointing to a sphere of all things considered authority that cannot be invoked to define a problem that should say something about the delimitation of that very sphere, and a sphere of fundamental authority that has not been adequately explained. Mill and Feinberg invoke the idea of the self-regarding to provide some more precise content to the liberal values, but whether or not a part of a person's life directly affects non-consenting others, or whether or not it is harmless, has no bearing on whether or not it is problematic to affect that part of her life for her good. The liberal values can be operationalized to become more substantial and specific. The operationalizations offered, however, whether in terms of substitution of judgement or agency, or in terms of diminishing of choice, make interference fatally wide or depend for their specification on an idea of authority that, again, is not adequately explained. Coercion in the form of behaviour control faces the same problems. Coercion in the form of threats is conceptually unsuitable to define interference. Compulsion, whether legal or non-legal, is often or always very intrusive and therefore perhaps seldom or never justified by benevolence. This, however, does not mark benevolent compulsion as a special moral category, any more than aesthetic coercion or other insufficient rationales for coercion. Defining interference in terms of wronging may avoid the need for an account of authority, but at the price of making the concept morally irrelevant. Actions that are presumptively wrong, but justified by other considerations, are not more problematic to perform benevolently than other justified actions. Invoking instead an idea of almost all things considered wronging makes paternalism a peculiar add-on that cannot change the moral status of an action from almost all things considered right to all things considered wrong.

Two strategies remain, both of which accept that paternalistic interference cannot be defined in terms of some other moral concept. First, interference may be considered a primitive moral concept in its own right, such that we can perhaps judge from case to case if something is an interference or not, yet cannot explain these judgements in terms of other moral concepts. There is, on this account, a primitive property of actions that consists in them being morally problematic when coupled with a benevolent motive or justification. Second, interference can be defined in descriptive terms, in practice by providing a long list of actions or kinds of actions that amount to interference, including perhaps such things as 'withholding of information about a person's health', 'taxing a person more for unhealthy than for healthy consumption', 'requiring that machinery (such as cars) be unpleasant to use unless the provided safety equipment (such as seat belts) is taken advantage of', etcetera. Perhaps the items on the list must be much more specific and free of implicit moral assumptions to avoid the problems associated with the accounts surveyed. The strength of both these strategies is their independence from (other) moral concepts. This independence, however, is also

their weakness, for it is doubtful whether either a primitive concept of interference or a descriptive list can carry the moral weight that allegedly makes paternalism morally problematic, without underpinning by (other) moral concepts such as liberty or choice.

I have not, of course, proved that any account of interference will fail or have serious deficiencies. There may be other strategies, and the strategies surveyed may be combined in different ways. I have not provided a very detailed critique of any one approach, but rather hurried through several, grouping them together under headings that I have deemed appropriate, often extracting what I take to be the core aspects of a strategy while omitting or only briefly discussing other aspects. My purpose has been to expose the difficulties involved in the very project of defining interference, arguably the most essential component of paternalism. In the process, I hope to have provided at least patches of detailed critique of the different strategies.

Philosophical concepts are often hard to define and specify. Should we not just live with the fact that interference is one of those concepts? We should not, for we do not need the concept. We do not need to define paternalism as a morally relevant category. The tendency to do so is a result of the traditional liberal scepticism of society's meddling in the affairs of individuals. There are good reasons for such scepticism, including state corruption and incompetence, the risk of oppression and of unnecessary convergence and impoverishment of culture and individual expression, and of course the basic value of autonomy and individual liberty. However, we do not need the concept of paternalism to defend liberty and individuality. We can simply acknowledge that such values as health, happiness and achievement are important and should be protected and promoted, while holding that self-direction is even more important and so most often outweighs opposing considerations.

Value conflicts are common, creating a need to explicate the conflicting values and their relationship and relative importance. Such investigation is hard work, allowing no easy short-cuts and no appeal to complex principles. Should we sacrifice some equality for the sake of greater opportunities for the well off? How do we strike a balance between stability and innovation? Should we restrict individual eccentricity for the sake of social cohesion? Should we promote popular and vulgar or obscure and elitist art? Should we preserve the natural environment at the cost of reduced convenience? Should we promote public health at some cost to liberty? Such questions are best answered not by appeal to principle but by thorough understanding of the values involved.

ACKNOWLEDGEMENTS

Thanks to Richard Arneson for discussion on these intricacies and for convincing me that I was not entirely on the wrong track. Thanks to Niklas Möller for constructive comments and to Lars Lindblom for detailed and constructive comments on at least two different drafts.

Forthcoming in *Public Health Ethics* 2(2)

Liberalism, Altruism and Group Consent

Kalle Grill

ABSTRACT: This article first describes a dilemma for liberalism: On the one hand restricting their own options is an important means for groups of people to shape their lives. On the other hand, group members are typically divided over whether or not to accept option-restricting solutions or policies. Should we restrict the options of all members of a group even though some consent and some do not? This dilemma is particularly relevant to public health policy, which typically target groups of people with no possibility for individuals to opt out. The article then goes on to propose and discuss a series of aggregation rules for individual into group consent. Consideration of a number of scenarios shows that such rules cannot be formulated only in terms of fractions of consenters and non-consenters, but must incorporate their motives and howmuch they stand to win or lose. This raises further questions, including what is the appropriate impact of altruistic consenters and non-consenters, what should be the impact of costs and benefits and whether these should be understood as gross or net. All these issues are dealt with in a liberal, anti-paternalistic spirit, in order to explore whether group consent can contribute to the justification of optionrestricting public health policy.

INTRODUCTION

According to standard liberal political theory, an action or a policy that restricts the options of a competent person stands in need of justification. One source of such justification is the consent of the person whose options are restricted. Public health measures often restrict the options of competent people. It would seem an important task for liberal political theory to investigate whether such measures can be justified by the consent of those affected. In contrast to medical care, public health measures normally target groups, with no possibility for group members to opt out. Most often, some members consent or would consent, while other members do or would not. A central part of the task, therefore, is to explore how various distributions of consenters and non-consenters within a group can justify restricting the options of the whole group. Such exploration is the aim of this article.

My conclusion is that aggregation rules for individual into group consent must consider the motives of (non-)consenters – typically self-interested or altruistic, as well as the costs and benefits to them. Such rules will therefore be rather complex. Attention to costs and benefits does not imply outright consequentialism and certainly not cost-benefit analysis in simplistic monetary terms. There are aggregation rules that consider both the fraction of consenters versus non-consenters, with different motives, and the distribution of costs and benefits. I propose that such rules are the closest we can get to a classically liberal justification of option-restricting policies by group consent.

This conclusion is hedged with assumptions. The central assumption we may call the *group consent assumption*, being that groups can morally be treated *as if* they had consented, even if some members do not consent. As we will see, this assumption is shared by many liberals, though not by strict libertarians who would not accept non-voluntary restriction for any benefit. I assume that liberalism implies individualism and so that talk of group consent is only metaphorical. Group consent is therefore used here as a normative notion, with no claim to metaphysical or linguistic accuracy. I will also assume that justification of a policy by consent does not require that consent can be given coherently or rationally to a series of policies.[i] I will further assume that consent can be aggregated independently of practical or semi-normative issues of delimitation – deciding who is a member of the relevant group – and coordination – enabling or facilitating group discussion.

An alternative strategy for dealing with aggregation of consent is to impose rules for the delimitation of groups, in terms of how severely individuals are affected by a policy. In contrast, I assume that affects on members may be of any degree of severity and come in any distribution. This assumption is warranted by two circumstances. First, from a liberal perspective, any option-restriction must be justified and so there is no rationale for excluding individuals because the effect on them is deemed insufficiently severe. Indeed, for the purposes of this discussion I will assume that any imposition of a cost amounts to a restriction.[ii] Second, from a practical perspective, it is normally impossible to delimit groups according to how individuals are affected, other than in the most rudimentary sense, such as where they live or work. At the end of the day, policy makers will most often face situations where some people in a given group consent and some do not. My question is what advice the liberal should give a policy maker in such a situation.

My method is explorative and constructive in the sense that I start from a very simple aggregation rule and work myself towards ever more complex rules in order to incorporate aspects that are shown to be important by argument and by the consideration of a series of scenarios. In the next section I describe the dilemma posed to liberal theory by collective self-regulation and in the third the futility of the literature on paternalism in this area. In the fourth section I propose some aggregation rules based on fractions of consenters and non-consenters and their motives. In the fifth section I go on to discuss aggregation rules that consider costs and benefits and the issues they raise. The sixth section concludes.

COLLECTIVE SELF-REGULATION

I take it for granted that restricting a competent person's options for her good without her consent is paternalistic and therefore illiberal.[iii] However, a person may *want* to have her options restricted, since this may bring great benefits and be an important means of shaping her life. As an individual, I sometimes aim specifically to restrict my own options. I may place the alarm clock some distance from the bed in order to restrict my option of turning it off and going back to sleep (without first standing up). I may

promise to meet you at the gym or invest in a gym membership in order to restrict my option of skipping exercise (without cost). I may ask you to stop offering me cigarettes, or to hold on to my car keys and not return them before I sober up, for obvious reasons. In general, I make promises and plans and investments in order to direct my future self by restricting my options. Restricting options is a kind of self-direction or self-regulation that makes use of the world and not only the mind.

Sometimes I am more active in creating situations where my options are restricted, sometimes less. Other people may offer to restrict my options, leaving me to accept or refuse their offers. Gyms offer me memberships, friends offer me to make joint gym plans. When I am more on the offer-taking side, we may most naturally speak of consent. Odysseus asked his sailors to tie him to the mast when they approached the sirens' island. If the sailors rather than Odysseus himself would have been the more informed and proactive, they may have offered to tie him to the mast and he may have consented. While it is sometimes important who takes the initiative, we may speak of consent regardless of whether a person actively creates the situation or passively, but informedly and intentionally, accepts it (cf. Feinberg 1986, chapter 22).

As shown by the examples, other people are often essential in enabling us to restrict our options. Depending on the options, we may need the aid of our relatives, our friends, or colleagues, or our community. Importantly, public policies can set up systems to restrict options. The examples above correspond to various forms of public health policy – prohibition and punishment (reproach for breaking a promise to be at the gym), subsidies (lower cost of exercise after buying gym membership), technical design (friend holding car key), and infrastructure design (alarm clock far from bed). When consented to, these policies may be seen as forms of collective self-regulation.

Public health policy can help groups regulate their primarily self-regarding behaviour, but also their other-regarding behaviour. Such regulation can solve prisoners' dilemmas and other coordination problems. There are many options that, while I would prefer that I have them rather than not, I would much prefer that none have them rather than all. Such options may concern direct harm to others, but may also concern the use of common assets. I might like having the option of dumping waste in the city park, but prefer that no one has this option rather than all. As for the prisoners, if they have a chance to restrict their dominating option of confessing, they will each deny and so receive a less severe sentence. Regulation of other-regarding behaviour that does not directly harm others may, when consented to, be seen as another form of collective self-regulation.

Restricting unhealthy and dangerous options will normally promote public health. This may be seen as a value in itself. This article, however, concerns justification by consent. As shown by the examples, restricting options may be an integral part of shaping one's life. Even disregarding the value of health, therefore, public health policy presents liberal theory with a dilemma where the value of enabling people to shape their lives according to their preferences conflicts with the disvalue of restricting people's options without their consent. Though policies can sometimes be adjusted to cover only consenters, this is often impossible or prohibitively expensive.

PATERNALISM AND GROUPS

In the literature on paternalism and anti-paternalism, many person or group cases are seldom discussed. When they are, the discussion is overly simplistic. Important anti-paternalists such as (the young) Richard Arneson (1980), Gerald Dworkin (1983) and Joel Feinberg (1986) adopt the group consent assumption without much discussion.[iv] This is surprising considering the tension between this assumption and the anti-paternalist core position that benefits to a person do not justify limiting her liberty, for example by restricting her options. While group cases are importantly different from single person cases, the group consent assumption implies that, at least interpersonally, losses in terms of restricted options can be justified by gains in, for example, health.

Arneson, Dworkin and Feinberg take very similar positions on group consent. Their view, we may call it the standard view, is to assume that, when a group is divided among consenters and non-consenters, the rationale behind an option-restricting policy targeting that group is either to benefit the consenters, in which case it is non-paternalistic, or to benefit the non-consenters, in which case it is paternalistic.

Arneson (1980) discusses group cases in connection with anti-duelling laws. He notes that people may prefer not to be confronted with duelling situations even though, if challenged, they prefer to preserve their honour by accepting, rather than avoid harm. If all agree, prohibiting duelling is an unproblematic case of collective self-regulation. However, as Arneson admits, there will always be some dissenters. Thus – the standard view: Even if some potential duellers are against the policy,

> if it is this pattern of desires [not to be confronted with dueling situations] *that generates reasons* for forbidding dueling, then the antidueling law (even if it is unfair or unjust) is nonpaternalistic. (Emphasis added, pp. 471-2)

Dworkin (1983) adopts the standard view in a discussion of fluoridation of water, a common public health measure which is typically resisted by a minority:

> [T]he restriction on the minority is *not motivated* by paternalistic considerations, but by the interests of a majority who wish to promote their own welfare. Hence, these are not paternalistic decisions (emphasis added, p. 110).

Feinberg (1986) has some minor issues with Arneson's account but adopts a very similar position:

> When most of the people subject to a coercive rule approve of the rule, and is legislated [etc.] *for their sakes*, and not for the purpose of imposing safety or prudence on the unwilling minority ('against their will'), then the rationale of the rule is *not* paternalistic. [...] Depending on the collective good involved, the costs and benefits, and the comparative sizes of the majority and minority, the statute may be fair or unfair, wise or unwise, but in either case, it will not be 'paternalistic.' (Emphasis in original, p. 20)

These three accounts are almost identical. If the rationale for an option-restricting policy is to benefit the consenters, then it is not paternalistic. Arneson does not, like Dworkin and Feinberg, explicitly state that the consenters must be in the majority, but he certainly assumes that they are.

The quoted passages are from the 1980s. However, the standard view is alive and well. In a recent article, J.D. Trout (2005) claims that an 'intervention that is *based on* third-party effects is not paternalistic.' (Emphasis added, p. 412) Trout's example is fluoridation of the local water supply. Given that such a policy is less expensive and more effective than the distribution of fluoride pills and that it is in some people's best interest, Trout concludes that the imposition of this policy on you against your will 'is for their sake and not solely for yours' and so is non-paternalistic. (p. 413) In other words, the policy is non-paternalistic even if some people 'want to defect'. (p. 413).

There are three related and serious problems with the standard view with its focus on the rationale (reason, motive) for a policy. First, it is not clear how the motive of the policy-maker affects the moral status of the policy. We must distinguish between moral evaluation of a policy-maker's decision and moral evaluation of the policy itself. This distinction is warranted because we might accept or support a policy because of its effects, or because of the legitimacy of the procedure that produced it, or because it is consented to by those affected, and all this independently of the motives of the policy-maker. A policy may be motivated by whatever obscure reason, such as furthering the policy-maker's career or making good on a bet. These motives might affect our moral assessment of the policy-maker's character and our attribution of praise and blame. However, in deciding whether *a policy* is justified, we are interested not in motives or psychological reasons, but in justificatory reasons.[v]

Second, while policy-makers might certainly view policy-making as a means to achieving certain pre-defined goals, they arguably should not. They should not act on their motives independently of the preferences of those affected, but rather consider the preferences (and perhaps interests) of all those affected and allow them their due impact on the formulation of policy. The good policy-maker considering whether to enact an option-restricting policy should ask: What is the due impact of the consenters and the non-consenters, respectively? On the standard view, since the rationale of a policy is either to benefit the consenters or to benefit the non-consenters, this question is not even intelligible. The standard view is only applicable to policies for which there is a single, pre-defined rationale.[vi]

Third, and most saliently, the standard view takes no notice of the fact that non-consenters have their options restricted against their will. Arneson and Feinberg point out that non-paternalistic polices may be objectionable on grounds of justice or fairness. They do not mention liberty. In the context, calling a policy non-paternalistic and not mentioning other potential conflicts with liberty is to adopt the group consent assumption. It is also to accept that societies with majorities bent on zealous self-regulation may impose strict health regimes on all citizens. It seems that Arneson, Dworkin and Feinberg do not consider the restriction of the options of the minority to

be in itself a moral obstacle to enactment of policy once the majority has consented. This is too hasty. As will become clear, the group consent assumption can be specified in a number of different ways, which are more or less liberal, more or less in tune with the spirit of anti-paternalism. Norman Daniels is right to point out that a claim that workplace safety regulation, or any other protective measure, enacted in response to majority will 'is not paternalistic ignores [...] the libertarian insistence that the autonomy of the minority is a fundamental liberty, a right, not a privilege so easily suspended at majority whim.' (2008, p. 197) Depending on our terminology, such measures may or may not be paternalistic, but regardless of terminology they are problematic from a liberal point of view.

It may seem misguided to look for answers in the paternalism literature when aggregation of consent is more reminiscent of aggregation of votes or preferences and so akin to social choice and democratic theory. However, group consent should not be confused with democratic decision making. Majority vote by a legitimate parliament may justify restricting people's options. This, however, is a separate and controversial claim. The liberal would typically hold that policies can be illiberal even if they are sanctioned by a democratic government (or that the government is truly democratic only if it abstains from such policies). The justification provided by consent is more substantial than that possibly provided by democratic decision making (which explains why some theories see consent as the foundation of democratic legitimacy). This is obvious if we look to medical ethics and the common position that treatment is justified only with the individual patient's informed consent (and so could not be sanctioned by democratic decision-making among patients).

In single person cases, consent is generally presumed to fully justify restriction of options. In other words, restricting the options of someone who consents is not to limit her liberty and so there is no prima facie wrong that needs justifying. On a strong interpretation of the group consent assumption, the same is true for group cases – that the group can be treated as if it had collectively consented means that the restriction of the options of its members is in no way morally problematic. This, however, may be too strong. Weaker interpretations are possible. I will not commit to a position on the exact moral impact of group consent but rather investigate when groups can plausibly be said to consent to a policy, under the assumption that such consent has substantial moral impact, if not enough to completely justify the restriction of options.

AGGREGATION RULES FOR FRACTIONS

In order to determine when a group can be treated as if it consents to an option-restricting policy, we clearly need a more fine-tuned theory than what can be found in the paternalism literature, and one focused not so much on the policy-maker as on group members. In this section and the next, I will consider a series of increasingly complex rules for aggregation of individual into group consent. Perhaps the most obvious factor to consider is the fraction of consenters. As noted, Feinberg, Dworkin and Arneson assume, explicitly or implicitly, that a group consents only if a *majority* of its members

consent. However, we must remember that group consent is not a matter of democratic decision-making but of more substantial moral justification. The issue is one of balancing liberal interests in self-regulation against liberal interests in non-restriction of options. The more libertarian our liberalism, the higher a fraction should be demanded. Consider, therefore, this simple rule:

The fraction consent rule: A group consents to an option-restricting policy if a large enough fraction of group members consent to the policy.

'Large enough' can be specified once and for all or can be allowed to vary with context. One problem with this rule arises from the possibility of altruistic consent.[vii] Consider the following scenario:

> **Altruism.** A group consists of two types of people – As and Bs. There is a policy that can be applied to the whole group or not at all. The policy restricts an unhealthy option (that the As hardly ever choose but that the Bs choose frequently). The Bs do not consent to the policy because they think it goes against their best interest. The As consent to the policy because they think it is in the best interest of the Bs.

From a liberal perspective, it would seem that in consenting, the As join rank with the paternalistic policy-maker in forcing a restriction on the Bs against their will. It seems counter to the spirit of justification by consent that the altruistic consent of the As should justify restricting the options of the Bs. Furthermore, it seems irrelevant what is the exact number or fraction of As. That a million rather than a thousand altruistic consenters accept a restriction that they have little or no personal interest in, for the sake of a few non-consenters, does not make it more reasonable to treat the group as if it had collectively consented.

This is not to say that altruistic consent counts for nothing. It is arguably less morally problematic to restrict the options of altruistic consenters than to restrict the options of non-consenters. In fact, this is exactly what I will soon argue. However, the point of collective self-regulation, from a liberal perspective, is that people should be free to restrict their own options in order to shape their own lives, not that they should be free to limit the freedom of others for their good against their will.

Though Arneson, Dworkin and Feinberg do not explicitly consider the motives of consenters, their insistence that a non-paternalistic policy be enacted 'for the sake of' (etc.) the majority suggests excluding justification based on altruistic consent. Here is a rule that does:

The fraction self-interested consent rule: A group consents to an option-restricting policy if a large enough fraction of group members consent to the policy out of self-interest.

Altruistic consent gives rise to another kind of problem for this rule. Excluding altruists from the fraction that counts towards group consent means including them in the remainder. In other words, every extra altruist counts against consent. With a constant number of self-interested consenters, and a constant number of non-consenters, increasing the number of altruistic consenters will change the status of the group from consenting to non-consenting, and vice versa for decreasing numbers of altruists.

Moreover, every extra altruist counts against consent to the same extent as every extra non-consenter. I propose that this position entails an unacceptable disregard for the disvalue of restricting a person's options against her will. The rule is inconsistent with the strong intuition that it is less morally problematic to restrict the options of altruistic consenters, who do after all consent, than to restrict the options of non-consenters against their will. To see this more clearly, assume that the required fraction is 70% and consider this scenario:

> **Gambling.** A local prohibition of gambling restricts the options of a group consisting of 24 members – 20 who are addicted gamblers and who consent out of self-interest, two who would hardly ever gamble anyway and who consent for altruistic reasons, and two who love to gamble and who therefore do not consent.

According to the *fraction self-interested consent rule*, the group consents to this policy (83% self-interested consenters). Now if the community grows with five people who would hardly ever gamble anyway and who therefore accept the prohibition for altruistic reasons, the group no longer consents to the policy (69% self-interested consenters). This might already seem counter-intuitive – why would the liberty of these new members to gamble be so important, given that they are not that interested in gambling and that they consent to the prohibition?[viii] What is more disturbing, however, is that if the community grows with four people who love to gamble and so do not accept the prohibition, the group still consents to the policy (71% self-interested consenters). While five additional altruistic consenters would end group consent, four additional non-consenters would not. Those of us who find this counter-intuitive should prefer this rule:

The fraction self-interested consent vs. non-consent rule: A group consents to an option-restricting policy if the fraction of members that consent to the policy out of self-interest is large enough compared to the fraction that do not consent.

Another form of altruism gives rise to similar complications. Group members might feel that our typical altruistic consenters, as well as self-interested consenters, trade off their freedom too lightly. They may therefore refuse to consent for liberty-preserving altruistic reasons, possibly against their own self-interest. The issues raised by this possibility are analogous to those of standard health-promoting altruism. In response, we may propose this modified rule:

The fraction self-interested consent vs. self-interested non-consent rule: A group consents to an option-restricting policy if the fraction of members that consent to the policy out of self-interest is large enough compared to the fraction that self-interestedly do not consent.

This rule takes altruists out of the equation altogether (along with group members who consent or not for yet other reasons, or are indifferent). Altruists neither contribute to nor subtract from group consent. This may be reasonable. However, it implies, for example, that a group consents if it consists of a million altruistic consenters, two self-interested consenters, and one self-interested non-consenter, while it does not consent if it consists of a million altruistic consenters, one self-interested consenter, and two self-interested non-consenters. Depending on one's convictions or intuitions, this might seem to either undervalue the free options of altruists, or to undervalue the altruists' free consent. To satisfy such intuitions, more complex rules could be designed that let altruistic consenters contribute to group consent, only discounted by some factor less than one, or that let altruistic consenters count against group consent, similarly discounted. In the same discounted fashion, altruistic non-consenters could count against group consent.

There are further complications. Just like policy-makers, consenters may have more than one motive. People may consent to a policy that restricts their options partly because they see that this will promote their health, and partly because they think that it will promote the health of others. Indeed, such motives are typical. In order to accommodate mixed rationales for consent, we could distribute the consent or non-consent of each member over the categories of the altruistic and the self-interested. With this scheme in place, we could further allow that preferences be distributed over both consent and non-consent, in order to accommodate hesitation and people who are conflicted concerning altruism and self-interest. Alternatively, and more rigidly, we could attribute to each member both a self-interested preference and an altruistic preference. These two kinds of preferences may then count equally or differently in the balancing of consent against non-consent. All this means that there are several alternative ways to accommodate altruism. Here are two rules that allow altruistic non-consent to count against group consent and altruistic consent to count either for (first rule) or against (second rule) group consent, though discounted, and that employ the former, less rigid strategy for accommodating mixed motives:

The fraction self-interested plus discounted altruistic consent vs. self-interested plus discounted altruistic non-consent rule: A group consents to an option-restricting policy if the fraction of self-interested consents plus the discounted fraction of altruistic consents is large enough compared to the fraction of self-interested non-consents plus the discounted fraction of altruistic non-consents.

The fraction self-interested consent vs. self-interested plus discounted altruistic consent and non-consent rule: A group consents to an option-restricting policy if the

fraction of self-interested consents is large enough compared to the fraction of self-interested non-consents plus the discounted fraction of altruistic consents and non-consents.

Consents and non-consents should here be understood not as individual consents but as distributed consents as just explained. This is as far as I will go with aggregation rules based on numbers or fractions of consenters and non-consenters, with different motives. We may conclude that in order to accommodate varied but common preferences, such rules must be rather complex. In the following section, I will argue that aggregation rules must consider not only fractions and motives, but also costs and benefits.

AGGREGATION RULES FOR COSTS AND BENEFITS

Aggregation rules based on fractions do not consider the cost or the benefits to different members. From a liberal perspective, this may be considered a virtue. However, sometimes the benefits for each consenter are great and the cost to each non-consenter trivial. Sometimes it is the other way around. I propose that once we accept the group consent assumption and so allow the interests of some members to override the interests of others, it is unreasonable not to consider the relative strength of these interests.[ix]

Consider this scenario:

> **Spartan Regime.** A group of warriors train hard in the mornings and evenings but tend to spend the warm afternoons lying around in the shady courtyard outside the barracks. A decree will prohibit loitering in the courtyards, in order to promote training. 81% of the warriors are good Spartans and though they would not train more but rather spend afternoons walking the fields, they appreciate the spirit of the decree. When the captain asks them they say they would welcome the decree. The remaining 19%, however, are not so good Spartans. They consider it a great honour to be warriors, but they find the training very burdensome. In fact, they could not stand it were it not for the relaxed afternoons in the courtyard, when they can share their troubles and give each other support. The decree would destroy the fragile social context that has evolved around the courtyard. When the captain asks them, therefore, they say they would not welcome the decree, for these reasons.

Assume that the good Spartans would appreciate the spirit of the decree in the sense that they would like being regulated by it themselves, not that they would want the not so good Spartans to stop loitering. This allows me to put altruism to one side for the moment. According to the *fraction self-interested consent vs. self-interested non-consent rule* with an 80% requirement, the group consents to the decree. I propose that this is unreasonable. There are of course several senses in which the group welcomes the decree, for example the sense that the majority welcomes it. However, in the context of consent, this is not a relevant sense. When the general asks the captain how the warriors feel about the

proposed decree, the captain should not say that they welcome it. The option-restricting effects on the non-consenting minority are not balanced out by the consent of the qualified majority. We are not warranted to treat the group as if it had collectively consented. This is so because the trivial (or non-existing) benefits to the majority are too small compared to the great costs to the minority.

Now consider this contrasting scenario:

> **Meanwhile in Athens.** A group of warriors train hard in the mornings and evenings and tend to spend the warm afternoons walking the fields. A decree will command the construction of a shady courtyard outside the barracks, in order to promote socializing and culture. The courtyard will make it more difficult to get to the fields. 81% of the warriors are good Athenians that are very proud to be warriors but are on the brink of despair because they so miss the cultured discussions of the civilian lifestyle. The proposed courtyard would mean the world to them and this is what they tell the captain when she asks. The remaining 19%, however, are immigrants from Sparta. They would not use the courtyard anyway and though they would not mind the longer path to the fields they do not appreciate the spirit of the decree, and so when the captain asks they say they would not welcome it.

This group may perhaps be said to consent to the decree, in accordance with the *fraction self-interested consent vs. self-interested non-consent rule* with an 80% requirement. When the Athenian general asks his captain how the warriors feel about the proposed decree, the captain could without fault say that they welcome it. If there is no time to describe the situation in further detail, that is indeed what she should say. Though the group is divided, it may on the whole be treated as if it had collectively consented. This is so because the benefits to the majority are so great compared to the trivial (or non-existing) costs to the minority.

It could be argued that it is simply misleading to talk of group consent in these cases. However, the issue is not whether group consent is a coherent notion when a group is divided. We are discussing aggregation of individual into group consent under the assumption that it is meaningful to do so under some circumstances, for example when a policy-maker, or general, must decide one way or other and perhaps wants to consider the consequences for the group in terms of enabling self-regulation and avoiding non-consented to restriction of options. Spartan Regime is not so unlike the prohibition of smoking in pubs, with the significant difference that smoking directly harms third parties. The question of balancing the consent of the good Spartans against the non-consent of the not so good Spartans is analogous to the question of how to balance the (let us assume) consent of the light smoking majority who wants to quit against the non-consent of the heavy smoking minority who have their whole lifestyle structured around smoking in the pub and do not want to quit. Meanwhile in Athens is analogous to common public health measures such as product safety regulation, sin taxes

and subsidies, where these are urgently welcomed by a qualified majority but opposed as a matter of principle by a minority of libertarians.

The *fraction self-interested consent vs. self-interested non-consent rule* with an 80% requirement implies that both the Spartans and the Athenians consent. The requirement could of course be any fraction and the scenarios reformulated accordingly. What is problematic is that the rule does not distinguish between the cases. Another way to bring out this problem is to lower the fraction of consenters in Athens to 79%. Now this group does not consent, while the Spartan group does. This is unreasonable. I conclude that rules for aggregating individual into group consent must pay some attention to costs and benefits.

There are as far as I can see two ways to introduce such consideration. One is to discard fractions and focus entirely on costs and benefits. Here is a rule that does (and excludes altruists):

The cost-benefit self-interested consent vs. self-interested non-consent rule: A group consents to an option-restricting policy if the benefits to those members who consent to the policy out of self-interest are greater than the costs to those members who self-interestedly do not consent.

Spartan Regime and Meanwhile in Athens show that costs and benefits must be considered, not that they must be decisive. A more conservative way to introduce considerations of costs and benefits is to keep the focus on numbers or fractions and let them be adjusted by the size of costs and benefits:

The fraction cost-benefit self-interested consent vs. self-interested non-consent rule: A group consents to an option-restricting policy if the benefit-adjusted number of members that consent to the policy out of self-interest is large enough compared to the cost-adjusted number of members that self-interestedly do not consent

Different versions are possible. The adjustment according to cost and benefit need not be strictly proportional.[x] The adjustment can be made for each individual or by median or mean for the respective subgroup.

Talk of costs and benefits does not imply a commitment to some particular theory of the good. The scenarios show that, in some sense, the costs of the option-restricting policy in Sparta are larger than the benefits, while in Athens the benefits of the option-restricting policy are larger than the costs. In this sense there are costs and benefits of policy alternatives and to this extent they are comparable. I will leave it an open question whether these costs and benefits are dependent on the objective value of health and liberty, or whether they are rather dependent on subjective preference. I will also leave it an open question to what extent these kinds of costs and benefits are comparable in general. As a matter of practical necessity, they must often be compared when making policy choices, or at least policy choices must be made as if they had been compared.[xi]

Whether we should prefer the *cost-benefit self-interested consent vs. self-interested non-consent rule* or the *fraction cost-benefit self-interested consent vs. self-interested non-consent rule* depends in large part on how we deal with two main issues that face both rules – should costs and benefits be understood as gross or net, and (how) should the rules be modified to consider the cost and/or benefits to (part) altruists. I will deal with these issues each in turn. I should state at the outset that I will not take a definite stand on these issues, nor on which of the rules should be preferred.

NET OR GROSS

We have been concerned exclusively with benefits to consenters and costs to non-consenters. However, a policy that we self-interestedly consent to because of its benefits need not be free of costs. I may welcome a prohibition on gambling while I recognize that it will prevent not only my excessive Friday night gambling sprees (the gross benefit), but also my innocent and pleasant Sunday afternoon poker. In considering the benefit to me of stopping the gambling sprees, should we subtract the cost of stopping the Sunday poker (to get the net benefit)? I propose that we have a basic intuition in favour of net cost or benefit. Assume that the effect for A is a great benefit and an almost as great cost, while the effect for B is a small benefit and an even smaller cost. Assume that the net benefit is equally large for A and B. It seems arbitrary and uncalled for that the effect on A should count for more than the effect on B.

However, net cost or benefit may be thought problematic in that it implies that the cost to self-interested non-consenters can be negative – i.e. they may benefit from a policy they do not consent to. For example, people who do not consent to fluoridation of tap water may benefit more from improved dental health than they lose in restricted options, though they themselves do not think so. Indeed, (paternalistically inclined) policy-makers will often believe that this is the case. Conversely, the net benefit to self-interested consenters can be negative – i.e. they may not benefit from a policy they consent to, but rather face a net cost. For example, people who consent to subsidies for gym memberships in order to make themselves exercise may fail to do so while still paying for the subsidy. Whether or not costs or benefits are in fact negative in a certain case will depend on empirical circumstances (and on what is the correct theory of the value of health and of the disvalue of restriction of options).

Under the *cost-benefit self-interested consent vs. self-interested non-consent rule*, net cost or benefit imply further that it does not matter whether or not members consent. This is so because we count both the costs and the benefits to both consenters and non-consenters and all these effects count equally. The rule can therefore be simplified:

The cost-benefit self-interested rule: A group consents to an option-restricting policy if it entails a net benefit to its self-interested members.

If we take the further step of counting also costs and benefits to altruists, as I will later argue we should if we settle for net cost or benefit, the rule reduces to standard consequentialist cost-benefit analysis:

The cost-benefit rule: A group consents to an option-restricting policy if it entails a net benefit to group members.

These rules may be reasonable. If we accept one of them, consent can still have an indirect effect in that the cost of having ones options restricted is greater, ceteris paribus, if one does not consent to such restriction. It is noteworthy that aggregation of individual into group consent will take us to cost-benefit analysis under these assumptions.

The possibility of negative costs and benefits is excluded if we opt for gross rather than net (gross costs and benefits to different members are of course aggregated into a net for the group – we must distinguish between the individual net and the group net). The lowest gross cost or benefit is simply zero, which in the case of the *fraction cost-benefit self-interested consent vs. self-interested non-consent rule* may be taken to leave the number of (non-)consenters intact but without positive adjustment. More generally, opting for gross costs and benefits does not allow that benefits that are not consented to or costs that are consented to count for or against group consent. In other words, while effects on individual interests can enhance the impact of an individual's (non-)consent, it cannot diminish it. This may be a means of preserving a strong form of respect for individual choice even while allowing costs and benefits some impact – an impact that is sufficient to explain our intuitions in Spartan Regime and Meanwhile in Athens, where there are no or only trivial benefits to non-consenters and costs to consenters.

However, these arguments for gross do not undermine our basic intuition in favour of net cost or benefit. Furthermore, this intuition can be strengthened by considering scenarios where the benefits to non-consenters are great, or where the costs to consenters are great, and where these great costs and benefits are recognized as such by the (non-)consenters themselves. Consider:

> **Smoking.** A public policy targeting a group of heavy smokers will levy high taxes on cigarettes (and use the surplus elsewhere). A majority consent to the policy because they know it will help them to marginally decrease their smoking. However, they recognize that the financial cost to them will be substantial. A minority do not consent to the policy because they know that it will not help them decrease their smoking and they recognize the financial cost. (If you think that the majority is being irrational to opt for a small benefit at substantial cost, assume that they think they deserve to be punished for their imprudent and morally irresponsible lifestyle.)
>
> **Smoking 2.** Like Smoking but the consenters are only light smokers and so the cost to them is smaller. Nonetheless, the health benefits from a marginal decrease in smoking will be equal to those of the consenters in Smoking.
>
> **Smoking 3.** Like Smoking but the non-consenters know that the policy will help them decrease their smoking substantially. (They still do not consent because they are against being taxed.)

If we count only gross benefits to consenters and gross costs to non-consenters, the three scenarios are equivalent. However, the net costs and benefits are substantially different. If the fraction of consenters required for the policy to be justified in each scenario varies, this indicates that we should understand costs and benefits as net.

We must not be fooled by possible intuitions to the effect that the lower cost in Smoking and the higher benefit in Smoking 2 should have *some* impact on the justification of policy. The question is which costs and benefits should have an impact on justification *by group consent*. Remember that the basic rationale for the group consent assumption is that consenters should be free to shape their lives according to their wishes. We might ask, therefore, whether it is more important for the consenters in Smoking 2 to restrict their options than for the consenters in Smoking. I propose that it is.

As argued in the introduction, wanted and beneficial restrictions are important in shaping our lives. I propose that restrictions that are more wanted and more beneficial are more important. When I consider what is beneficial to myself in shaping my life I think in terms of net – I include the entailed costs of beneficial and wanted outcomes. If I decide to get married or to pursue a career in philosophy or to run for office, I consider both pros and cons, both costs and benefits. A rule that protects my life shaping and that of my fellow group members should distinguish between great benefits at low cost and great benefits at great cost.

We may then ask whether it is less important for the non-consenters in Smoking 3 to avoid restriction than for the non-consenters in Smoking. Again, I propose that it is, for similar reasons. Unwanted and non-beneficial restrictions are important to avoid. Restrictions that are more wanted (less unwanted) and more beneficial are less important to avoid. A rule that protects my options and those of my fellow group members should distinguish between great costs for no benefit and great costs for substantial benefit.

In sum, I find the case for net cost or benefit stronger than the case for gross. However, there are good arguments on both sides and so I am happy to leave the matter undecided.

ALTRUISM

As noted, the possibility of altruistic consent requires that we specify the motives of (non-)consenters. The cost-benefit rules I have formulated so far in this section exclude altruistic (non-)consent. This avoids the possibility that altruistic consenters indirectly force benefits on non-consenters against their will or that altruistic non-consenters indirectly prevent consenters from shaping their lives according to their preferences. However, as noted, it may be too drastic to simply disregard altruistic (non-)consent. The fraction and number based rules can all be modified to grant altruism full or discounted impact. This is true also for the *fraction cost-benefit self-interested consent vs. self-interested non-consent rule*. In the case of the *cost-benefit self-interested consent vs. self-interested non-consent rule*, fractions do not matter and so the question is simply whether costs and benefits to altruists count and if so whether in full or discounted.

Consider this scenario:

Gambling 2. A local prohibition of gambling restricts the options of a group of addicted gamblers. The members can be divided into six subgroups. The first subgroup knows the policy will help them out of their addiction – a great net benefit – and consent for that reason. The second subgroup knows that the policy will destroy their most cherished hobby – a great net cost – and for that reason do not consent. The third subgroup knows the policy will help them out of their addiction, but have no concern for themselves. On the other hand, they also know that the policy will help the first (and the fifth) subgroup out of their addiction and consent for that reason. The fourth subgroup knows the policy will destroy their most cherished hobby, but have no concern for themselves. They also know, however, that the policy will destroy the cherished hobby of the second (and the sixth) subgroup, and for that reason they do not consent. The fifth subgroup, like the third, knows the policy will help them out of their addiction, but have no concern for themselves. Unlike the third subgroup, they do not consent, in order to protect the cherished hobby of the second (and the sixth) subgroup. The sixth subgroup, like the fourth, knows the policy will destroy their most cherished hobby, but have no concern for themselves. They consent, however, in order to help the first (and the fifth) subgroup out of their addiction.

Which of these costs and benefits should have an impact on group consent? The four altruist subgroups can be divided along two dimensions – whether they stand to win or lose from the enactment of the policy, and whether they sympathize with those who stand to win or with those who stand to lose. Their sympathies in turn determine whether they consent or not. However, I propose that the singularly most relevant aspect, cutting across the two dimensions, is whether or not the (non-)consents of the altruists are in conflict with their own interests (this aspect could have been exemplified with two altruistic subgroups – all four are included for comprehensiveness, illustrating, as noted above, that there may be altruists both among consenters and non-consenters).

We are arguably most likely to accept as relevant the costs and benefits to those altruists (the third and fourth subgroup) who consent or not consistently with their own interest. Their noble or self-denying character should not count against them and there is no conflict between what they prefer and what is in their interest. We are probably more reluctant to accept as relevant the costs and benefits to those altruists (the fifth and sixth subgroup) who consent or not in conflict with their own interest. To allow that these costs and benefits affect group consent is to allow individual interests to diminish or count against individual consent.

The issue is essentially the same as whether or not costs to self-interested consenters and benefits to self-interested non-consenters should ever lessen the impact of their (non-)consent. As noted, that they should not is a good argument for understanding costs and benefits as gross rather than net. Consistency therefore requires that we either settle for gross costs and benefits and disregard costs and benefits to altruists that consent or not in conflict with their own interest, or that we settle for net

cost or benefit and consider the cost or benefit to all altruists. It is this reasoning that implies that the *self-interested cost-benefit rule* reduces to *the cost-benefit rule*, as indicated above.

After deciding which types of altruist cost and benefits should count in aggregating group consent, the next pressing matter is whether these costs and benefits should count as equal to those to self-interested members or whether their impact should rather be adjusted by some factor. Many possibilities suggest themselves. However, this is as far as I will go with this issue.

In sum, there are strong reasons to include at least some altruist (non-)consenters in the numbers under the *fraction cost-benefit self-interested consent vs. self-interested non-consent rule*, and among the 'containers' of costs and benefits under the *cost-benefit self-interested consent vs. self-interested non-consent rule*. If we should include altruists that consent or not in conflict with their own interest, and whether or not (different kinds of) altruists should count equally with self-interested members, these questions I leave undecided.

CONCLUSION

Option-restricting public health policies that are welcomed by some of those affected but not by others raise intriguing problems for liberal political theory. One approach to these problems is to investigate under what conditions groups that are divided can be treated as if they had collectively consented. This issue has been too hastily dismissed by several anti-paternalists, even though the core values underlying anti-paternalism are clearly relevant to how it should be handled. Once we recognize the need to consider costs and benefits, we must ask whether they can count against individual consent. In the spirit of anti-paternalism, we should perhaps say no. This entails saying no to negative costs and benefits, and no to discounting altruistic consent because it is in conflict with self-interest. On the other hand, once the importance of costs and benefits is acknowledged, strong intuitions drive us to consider not only gross, but net cost or benefit. If we do, consenters who stand to benefit more will count for more than those who benefit less, non-consenters who stand to loose more will count for more than other non-consenters, and altruistic consenters that consent or not in conflict with their own interests will count for less than those who consent or not consistently with their own interest. It is not obvious how we should react to these results. They should therefore be investigated in further scenarios and preferably also in application to real cases.

Regardless of how these issues are finally resolved, aggregation rules for individual into group consent must consider both costs and benefits to group members, and their motives for (non-)consent. Under net cost and benefit, the *cost-benefit self-interested consent vs. self-interested non-consent rule* reduces rather straightforwardly to the *cost-benefit rule*, saying that a group consents if its members can on the whole expect a net benefit. Coupled with an understanding of cost and benefit as gross, however, this rule offers an interesting compromise between outright consequentialism and more principled liberalism. The *fraction cost-benefit self-interested consent vs. self-interested non-consent rule* offers another such compromise and can be combined with either gross or net costs and benefits. This rule is very general and can be specified in any number of ways. If the

general approach has some merit, more detailed rules should therefore be formulated. Both rules should be adjusted to allow altruistic non-consenters who do not benefit to count against consent. Both rules may or may not be adjusted to allow other altruists to have an impact on group consent.

Stepping back from the details, we may conclude that a theory of justification of option-restricting policies by group consent is possible but will necessarily be quite complex. These complexities, and the introduction of costs and benefits, may tempt the liberal to reject the group consent assumption and claim in a Nozickian manner that restricting the options of competent people against their will is simply impermissible, except when doing so is the only way to avoid catastrophe (Nozick 1974). However, depending of course on how loosely catastrophe is defined, the price would be great in terms of limits on how individuals can use the power of collective self-regulation to shape their own lives. The opposite reaction is perhaps more sensible – to reject group consent as a source of justification for public health policy and look to other sources.

---

[i] I will therefore not consider the growing literature on preference aggregation and, more recently, judgement aggregation, which deals predominately with such issues. Aggregation of consent is distinct from aggregation of preferences, judgement or welfare. Aggregation rules may always be vulnerable to Condorcet's paradox (the voting paradox) and similar paradoxes. Such vulnerability does not imply that aggregation has no moral impact. Majority vote may make a government legitimate, even if there is a possible majority among voters that would prefer another government over the present one and another majority that would prefer a third one over the second one and yet another majority that would prefer the present one over the third one. Likewise, group consent may justify option-restricting policies, even in the face of similar cyclicality.

[ii] What counts as an imposition of a cost is always relative to a baseline. My investigation must therefore be understood against the background of some general theory of justice which defines such a baseline.

[iii] Provided that she is sufficiently capable and informed and so her choices sufficiently voluntary.

[iv] I count Dworkin as an anti-paternalist because he holds that paternalism is prima facie wrong, even though he is happy to make exceptions. Regardless of the appropriateness of this label, his account of paternalism and group cases is typical of liberal philosophers.

[v] Arneson's 'generates reasons' seems at first to refer to justificationary reasons, but he immediately goes on to sort cases according to 'the motivation of the lawmakers' (p. 472).

[vi] Admittedly, claims that option-restricting policies are paternalistic only if their rationale is to benefit non-consenters could be reinterpreted, in the spirit of Husak (2003) and Grill (2007), to mean that it is paternalistic to allow that benefits to non-consenters count in favour of option-restricting policies. However, if anti-paternalism is limited to excluding such reasons, it has nothing to say about the dilemma policy-makers face when they must choose whether to enable people to shape their lives according to their wishes or to avoid restricting people's options without their consent.

[vii] Similar problems arise for consent based on any kind of external preference (a preference regarding the outcome for others). The case of altruism (directed at other group members) is especially interesting because it is common and apparently benevolent and legitimate.

[viii] It may be argued that these non-gamblers should not count as members of the group. However, as noted in the introduction, their options are still restricted and it is practically very difficult to delimit groups according to individual interest. It is of course equally difficult to ascertain individual interest for the purpose of aggregating consent. However, consent aggregation is an ideal to be approximated, an abstract moral rule that must be adjusted in the face of practical constraints. If we understand group delimitation the same way, and if we allow that delimitation is not digital but come in degrees, then what I say about consent aggregation can be translated into delimitation talk, and the two strategies are in that sense equivalent.

[ix] Similar intuitions seem to have led Brighouse and Fleurbaey (2008) to argue for power in proportion to stakes in any decision-making process. Their discussion is more general and placed in the context of democratic theory. They dismiss the problem of altruism on grounds I fail to see.

[x] This opens the door to mimicking various theories of distributive justice. For example, prioritarian versions are possible. However, we must remember that what we are investigating is aggregation of

consent and not considerations of justice, which may in themselves of course justify restricting options, or define what is a restriction in the first place.

[xi] Those who are still wary of talk of costs and benefits might prefer to formulate aggregation rules in terms of reasons. That some restrictions of options are more severe than others and that some benefits are larger than others might be taken to imply that some reasons against restricting options are stronger than others and that some reasons for creating benefits are stronger than others. Switching terminology to reason talk may seem to automatically give us the common currency of strength with which to compare different considerations. In fact, however, this gain is superficial, since we still have to determine the function to the strength of reasons from the severity of restrictions of options and from the size of benefits, respectively. I will stay with the language of costs and benefits rather than that of reasons, with the assumption that everything I say can be straightforwardly translated into reason talk.

References

Arneson, R. (1980). 'Mill versus Paternalism'. *Ethics* 90: 470-489.

Brighouse, H., and Fleurbaey, M. (2008). 'Democracy and Proportionality' *Journal of Political Philosophy* Advance Access published July 10, 2008, doi: 10.1111/j.1467-9760.2008.00316.x.

Daniels, N. (2008). *Just Health: Meeting Health Needs Fairly*. Cambridge: Cambridge University Press.

Dworkin, G. (1983). 'Some Second Thoughts'. In Rolf Sartorius (ed.), *Paternalism*. Minneapolis: University of Minnesota Press, pp. 105-11.

Feinberg, J. (1986). *Harm to Self*. Oxford: Oxford University Press.

Grill, K. (2007). 'The Normative Core of Paternalism'. *Res Publica* 13: 441-458.

Husak, D. (2003). 'Legal Paternalism'. In *The Oxford Handbook of Practical Ethics*. Oxford: Oxford University Press.

Nozick, R. (1974). *Anarchy, State, and Utopia*. Malden MA: Basic Books.

Trout, J.D. (2005). 'Paternalism and Cognitive Bias'. *Law and Philosophy* 24: 393-434.

Acknowledgements

An early version of this paper was presented at the Public Health Ethics panel of the 2008 Workshops in Political Theory, in Manchester. Thanks to the attendees, in particular Jurgen De Wispelaere and James Wilson, for constructive proposals. Thanks also to the Ethics Group at the Division of Philosophy at the Royal Institute of Technology for comments. Thanks most of all to Niklas Möller and Lars Lindblom for detailed and thorough comments that led to substantial changes.

References

Arneson, R. (1980). 'Mill versus Paternalism'. *Ethics* 90: 470-489.

Daniels, N. (2008). *Just Health: Meeting Health Needs Fairly*. Cambridge: Cambridge University Press.

Dworkin, G. (1983). 'Some Second Thoughts'. In Rolf Sartorius (ed.), *Paternalism*. Minneapolis: University of Minnesota Press, pp. 105-11.

Feinberg, J. (1986). *Harm to Self*. Oxford: Oxford University Press.

Grill, K. 2007. 'The Normative Core of Paternalism'. *Res Publica* 13: 441-458.

Husak, D. (2003). 'Legal Paternalism'. In *The Oxford Handbook of Practical Ethics*. Oxford: Oxford University Press.

Nozick, R. (1974). *Anarchy, State, and Utopia*. Malden MA: Basic Books.

Trout, J.D. (2005). 'Paternalism and Cognitive Bias'. *Law and Philosophy* 24: 393-434.

Forthcoming in Dawson, A (ed.) *The Philosophy of Public Health* (Ashgate)

Anti-paternalism and Public Health Policy: The Case of Product Safety Regulation[1]

Kalle Grill

The UK General Product Safety Regulations 2005 states that products may not be brought to market if they present more than "the minimum risk compatible with the product's use, considered to be acceptable and consistent with a high level of protection for the safety and health of persons."[2] This recent regulation grants enforcement authorities the power to have products withdrawn from the market and recalled from buyers, and extended powers to halt the process of bringing a product to market.[3] In other words, decisions on acceptable risks from consumer products are to a considerable extent placed with government authorities, decisions that would otherwise be made by individual consumers. The UK regulation is based on a 2001 European Union directive; similar regulations apply throughout the Union, as well as in some other countries.

The risks involved in using consumer products are often risks to the user herself, rather than to third parties. Product safety regulation therefore typically involves paternalism. This article aims to distinguish the paternalistic content of product safety regulation, and in so doing, providing a more general framework for distinguishing the paternalistic content of any public health policy. Distinguishing the paternalistic content of policy is important if we want to evaluate the widespread resistance to paternalism codified in liberal anti-paternalist principles.

Most accounts of paternalism can be accommodated by the general definition *interference with a person, against her will, for her good*.[4] In the following, each of these three components will be discussed and applied to the case of product safety regulation, without commitment to a certain specification of any of the components. Concerning interference, I will consider five aspects of product safety regulation that make it an interfering policy, aspects that are relevant for public health policy more generally. Concerning will, I shall focus on the complexities arising from the fact that policies affect many persons and so can be welcomed or accepted for quite different reasons. Concerning good, I will consider the important and often misunderstood role of reasons in understanding paternalism, again with special attention to many person cases. This three-fold interpretation will then be used to discuss the normative status of paternalism

[1] This text was completed during a four month visit to University of California San Diego, financed by the Swedish Foundation for International Cooperation in Research and Higher Education. Thanks to Richard Arneson for our many discussions on paternalism and anti-paternalism during that visit, to Niklas Möller for comments on an early draft, and to the editor of this book for comments on a later draft.

[2] *The General Product Safety Regulations 2005*, Statutory Instruments 2005 no. 1803, p. 4.

[3] Peter Cartwright, 'Enforcement, risk and discretion: the case of dangerous consumer products', *Legal studies* 26(4) (2006): 524-43.

[4] Or an interference with several people against their will for their good.

and anti-paternalism in public health policy. Throughout, references to the philosophical debate are largely confined to the footnotes.

INTERFERENCE

For there to be paternalism, there must be some kind of interference on the part of the paternalist. Involvement with a person, merely effecting her in some way, is not enough. Interference may generally be thought of in terms of restriction or limitation of liberty. It is, however, a matter of controversy how the line should be drawn between innocuous involvement and interference. Disagreement may arise both concerning which concrete actions qualify as interference and concerning how interference should be specified in general terms. There are (at least) five reasons for holding that product safety regulation amounts to interference.[5]

First, regulation restricts options.[6] Less safe products may quite possibly have more or other functions and designs. Prohibiting the market exchange of less safe products thus restricts the options of individual consumers in a non-trivial sense. Importantly, the restriction of options may matter also for a person who would not have taken advantage of those options had they been available. Freedom is arguably about having more options available than those you actually choose to realize.[7]

Second, as a special case of restriction of options, regulation imposes a cost on the individual consumer. It is in general more expensive to produce safer products. If nothing else, the process of ensuring that the product is safe and that it accords with relevant regulation entails a cost. Consumers are in effect forced to spend money on safety features. If they were allowed to choose from a wider range of products, they should be able to find less expensive products that would serve the same purpose as more expensive, safer products. In the long run, the aggregated cost of safety may be quite high.

Third, because of the above traits of product safety regulation, it may go against the preferences of individual consumers.[8] People may value the opportunity to buy less safe products because they are less expensive, or because they have more or other functions, or simply because of a preference for simple, old-fashioned, or 'raw' products.

---

[5] I will assume throughout that those affected by the policy are sufficiently mature, informed, competent, and so on, to qualify as potential targets of illegitimate paternalism, according to liberal principles. The exact specification of these factors is an important and difficult matter for any anti-paternalist principle. The most ambitious attempt to accommodate this difficulty is arguably Joel Feinberg's in *Harm to Self*, Oxford University Press 1986, especially chapter 20.

[6] Restriction of options has been taken to constitute interference by e.g. David Archard ('Paternalism defined', *Analysis* 50(1) (1990): 36–42, p. 36), proposing as one condition of paternalism that a person 'P aims to bring it about that with respect to some state(s) of affairs which concerns [another person] Q's good Q's choice or opportunity to choose is denied or diminished.'

[7] Isaiah Berlin (Five essays on liberty: Introduction, in *Liberty*, Oxford University Press 2002 (1969), p. 41) remarked that '[t]he extent of a man's negative freedom is, as it were, a function of what doors, and how many, are open to him; upon what prospects they open; and how open they are'. It is, of course, not obvious that the net effect of regulation will be a loss of liberty so understood.

[8] Donald Van de Veer (*Paternalistic intervention*, Princeton University Press 1986, pp. 18–19) proposes that an action is an interference if the agent deliberately acts 'contrary to the operative preference, intention, or disposition of the subject' (or if she shapes or modifies these preferences in certain ways).

People may also prefer to have options available that they do not in fact want to take advantage of.

Fourth, the purchase and use of consumer products is a typically private affair. Whether individuals use safe or less safe products usually has no direct effects on other people, or on society at large. There are of course indirect effects of people having accidents and subsequently becoming a burden on their loved ones and more generally on the health care system, while contributing less to society. Such effects may ensue, however, from all kinds of actions, however private. If there is such a thing as an area of personal sovereignty or a region of liberty, as anti-paternalists typically claim, the purchase of consumer products seems a good candidate for inclusion under this domain.[9]

Fifth, product safety regulation is backed up by criminal sanctions. It thus qualifies as interference also on those narrow accounts of paternalism that are restricted to the criminal law.[10] That the law punishes the seller rather than the buyer might make this a case of 'impure' paternalism, or a 'two party case' – the direct interference is with one party and the concern is with the health of the other party. On the other hand, since the buyer is an active and willing party to a mutual agreement, she too may be interfered with by the threat of sanctions to the seller. Whether or not the interference is also with the buyer, these kinds of sanctions are typically and reasonably held to potentially involve paternalism.[11]

There are many ways to specify interference and every specification entails a different version of anti-paternalism. We may conclude, however, that there are several good reasons to count product safety regulation as interference. These reasons are quite general and may obviously apply also in other areas of public health policy. That product safety regulation is interfering does not of course mean that it is unjustified all things considered, nor that it necessarily involves paternalism.

WILL

That the effects of a policy goes against the preferences of a person subject to that policy is one reason to count the policy as an interference with that person, as noted above. However, preferences, or will, may also be considered an independent component of paternalism. If a policy constitutes an interference with a person on other grounds than going against her preferences, she may welcome the policy, fully aware of its interfering properties. We may want the government to ensure that there are no unsafe products

[9] Feinberg's (chapter 19) account of paternalism rests heavily on the concept of personal sovereignty and the distinction between self-and other-regarding decisions; John Stuart Mill (*On Liberty*, in *On Liberty and Other Essays*, Oxford University Press 1991 (1859), p. 16) explains his anti-paternalist principle of liberty by pointing to 'the appropriate region of human liberty' as being 'that portion of a person's life and conduct which affects only himself, or if it also affects others, only with their free, voluntary, and undeceived consent and participation.'

[10] Feinberg explicitly restricts the domain of his anti-paternalist principle to criminal prohibition. There is a discrepancy between this narrow occupation with the criminal law and the fact that Feinberg's main argument against paternalism is based on the broad concept of personal sovereignty, see Richard Arneson, 'Joel Feinberg and the Justification of Hard Anti-paternalism', *Legal Theory* 11 (2005): 259–284, pp. 262–3.

[11] Gerald Dworkin, 'Paternalism', *Monist* 56 (1972): 64–84, p. 68; Feinberg, chapter 22, e.g. p. 172.

available on the market, even if this restricts our options, because we do not think the risk of buying an unsafe product worth the possible benefits. The risks may include harm to oneself as well as the risk of harming others with the product (with possible liability). A policy that is welcomed on these grounds arguably does not involve paternalism. Importantly, it does not involve something that is opposed by liberal anti-paternalist principles.[12]

There are several choices to be made concerning the specification of the will component of paternalism. We may say that an interfering policy is unwelcome either when it goes against a person's expressed opinion, or when it is against her will or judgement, whether expressed or not, or whenever it does not have her expressed approval. We may also add conditions demanding that the approval or disapproval be more or less informed and competent.[13] Acknowledging that different interpretations of will lead to different versions of anti-paternalism, we may for our present purposes assume that a policy is normally welcomed by a person if she either explicitly approves of it, or would approve of it if the matter was brought to her attention.

As is now and then pointed out, public health policy differs from medical health contexts in that it affects large numbers of people, often in a non-discriminatory way.[14] This means that a policy may be welcomed for a number of different reasons. We should distinguish between welcoming the interfering effects of a policy on oneself, and accepting these effects as a necessary evil that is outweighed by the greater good of having the policy apply to all. The former case involves no more paternalism than an interference with one person that is welcomed by that person. Concerning the latter case, we should further distinguish between welcoming a policy because of the good effects for ourselves from interference with everybody else, and welcoming it because of the good effects on others from interference with them. As an example of the former, we may accept out of *self-interest* that the government ensures that we, as well as our neighbours, drive safely or keep our lawns tidy. As an example of the latter, we may accept out of *benevolence* that the government prevents us, as well as those more prone to addiction, from using heroine or tobacco. Similarly, we may accept interfering product safety regulation either because we do not want others to use dangerous products that may harm *us*, or we may accept it out of concern that they may harm *themselves*. A policy that interferes with us but that we welcome as a means to ensure compliance with a scheme that promotes our self-interest does, arguably, not involve paternalism for us.[15] On the other hand, it is undoubtedly paternalism to support a policy because it prevents other people from harming themselves. The hard question is if a policy may subject *me* to paternalism, if I accept it because of the good it will do others. It seems we can go either way. On the one hand, the policy may count as involving paternalism for me because it

[12] Mill (p. 14) typically opposes benevolent interference with a person only when it is 'against his will'.
[13] Again rising questions of what thresholds to accept, see footnote 5.
[14] This difference has been emphasised in recent calls for a public health ethics distinguished from traditional bioethics, see e.g. Ronald Bayer and Amy L. Fairchild, 'The Genesis of Public Health Ethics', *Bioethics* 18(6) (2004): 473–492.
[15] This is in line with Dworkin's (p. 69) argument that without paternalism, 'individuals [...] may need the use of compulsion to give effect to their collective judgement of their own interest by guaranteeing each individual compliance by the others.'

interferes with me and I do not find that the interference is made worthwhile by any benefit to me. On the other hand, the policy may count as involving no paternalism, because I do nevertheless welcome it.

It seems likely that on most specifications of the will component, most people would welcome product safety regulation on the level common in the European Union. For them, the regulation will not involve paternalism. However, there are certainly some people who do not welcome regulation, and with whom the regulation is an interference, for some or all of the reasons pointed out in the previous section. Product safety regulation, therefore, amounts to unwelcome interference with some people, but not with others.

GOOD

If a policy amounts to unwelcome interference, it may involve paternalism. For there to be paternalism, however, the interference must be in some sense for the good of those interfered with.[16] We tend to think of paternalism as residing in actions, including complex state actions – policies. The good component then functions as a condition on which actions qualify as paternalistic. This is anyway how paternalism is defined in the philosophical literature. Unwelcome interferences are typically said to be paternalistic if they are *motivated* by the good of the person interfered with, or, less commonly, if they are *justified* by the good of this person.[17] However, actions are often motivated, as well as justified, by several different reasons. While this complexity is sometimes acknowledged, the solutions are unsatisfactory. Actions are typically counted as paternalistic when their rationale is solely[18], or mainly[19], the good of the person interfered with, or to the extent[20] that this is their rationale. These standard interpretations are ill suited for distinguishing those morally problematic aspects of paternalism that liberal anti-paternalists are concerned with. The essence of paternalism is the invocation (or acceptance) of the good of a person as a reason for unwelcome interference with her, regardless of the relative strength of this reason as compared to other reasons for the same interference.[21]

When we call a policy paternalistic, this should be interpreted merely as a convenient way to say that it *involves* paternalism, in the sense that the policy is interfering, unwelcome, and that the good of some people who are subject to this unwelcome interference is invoked as a reason for the policy. If we want to distinguish the paternalistic content of a situation more precisely, we must accept that policies are not

---

16 Seana Shiffrin ('Paternalism, Unconscionability Doctrine, and Accommodation', *Philosophy and Public Affairs* 29(3): 205–50, pp. 215–17) takes an uncommon stand on this issue and argues that acting out of disrespect for a person's judgement or agency is paternalistic regardless of whether or not it is for the good of the person.

17 Or possibly both, as proposed by Peter De Marneffe, 'Avoiding Paternalism', *Philosophy and Public Affairs* 34(1) (2006): 68–94, pp. 73–74.

18 E.g. Gerald Dworkin, 'Paternalism', *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Winter 2002 Edition), URL = <http://plato.stanford.edu/archives/win2002/entries/paternalism/>.

19 E.g. Archard, pp. 38–39.

20 E.g. John Kleinig, *Paternalism* (Manchester: Manchester University Press 1983), p. 12.

21 For a more detailed argument for this interpretation of paternalism, see Kalle Grill, 'The Normative Core of Paternalism', *Res Publica* 13(4) (2007): 441-458.

paternalistic as such, but only in combination with certain reasons. The most obvious and sensible reason for introducing product safety regulation is to protect people from the risk of harm from unsafe products. However, there could be other reasons. A corrupt politician might push for regulations because they favour certain manufacturers. A more altruistic and far-sighted politician may propose regulations because they would stimulate technological innovation, spilling over into other areas. The invocation of the these reasons for the policy may be more or less appropriate, but is not paternalistic, regardless of whether or not avoidance of harm reasons are invoked for the same policy. Similarly, what is paternalistic about involuntary psychiatric treatment is invoking the good of the patient as a reason for treatment, rather than the treatment as such; what is paternalistic about drug criminalization is the invocation of the good of (potential) drug users as a reason for criminalization and punishment, not the criminalization itself; and so on for other policies. We should insist on interpreting paternalism in terms of the invocation of reasons because this, unlike standard action-focused accounts, distinguishes precisely that aspect of policy-making and implementation that is resisted by anti-paternalism.

There is an important complication to be noted in many person cases. Even if a policy interferes with a person and is unwelcome, this may not be enough to make the invocation of her good as a reason for that policy paternalistic. This is because a policy that affects many may promote the good of each by interfering with the others. This is typically the case for policies that we do not think of as involving paternalism, such as laws against theft, assault and murder. These laws restrict the options available to all and may plausibly be unwelcome for some people. There are thus people that are protected by these laws but for whom these laws amount to an unwelcome interference. These people are not, however, protected through the unwelcome interference *with them*, but rather through interference with others, who are not allowed to harm them. Similarly, product safety regulation may to some extent protect people from risks of harm through the interference with other people, as noted above. I may oppose regulation and regulation may be interfering for me, yet when I benefit from the fact that my neighbour is not allowed to buy a dangerous lawnmower that could explode next to my garden table, this benefit occurs as a result of interference with her and not with me. Invoking this benefit to me as a reason for the policy is therefore not paternalism. As we saw in the previous section, restricting the options of others to harm me may in some cases be an interference also with me, as when others are prohibited from selling me dangerous or unhealthy products. The point is that regardless of how we specify interference, we must count as paternalistic only the invocation of a person's good for an action that achieves that good through interference *with her*.

THE MORAL STATUS OF PATERNALISM

A great advantage of interpreting paternalism in terms of the invocation of reasons is that we may distinguish different kinds of reasons and consider for each kind whether its invocation is paternalistic or not. Based on such an analysis we may then approach the important normative question of how to evaluate different forms of paternalism. The

two main kinds of reasons to consider are arguably psychologically motivating reasons, or motives, and justificatory reasons. Motives cause and explain the actions they are motives for, while justificatory reasons justify, or contribute to the justification of, actions they are reasons for. We could also, however, focus our attention on officially stated reasons, or reasons invoked in some other context.

Resistance to paternalism may take somewhat different forms depending on what kind of reasons are invoked paternalistically. In terms of justification, anti-paternalism may most obviously be interpreted as a restriction on what reasons should count when we evaluate actions. The liberal anti-paternalist is not necessarily opposed to policies interfering with you. Whether an unwelcome interference is acceptable or right overall may depend on several considerations. What the anti-paternalist claims is that your good is not one of these considerations. We may say that this kind of reason is *invalid* as a reason for this kind of policy. In terms of motives, there may similarly be several reasons why a policy-maker may want to interfere with you against your will. According to anti-paternalism, your good should not be among those reasons. We may say that this motive is *inappropriate* for this kind of policy. As we have seen, having an inappropriate motive does not exclude the possibility of being motivated also by other reasons, which in themselves may be impeccable to the anti-paternalist and which may make the interference in one way commendable.[22]

Given that there are potentially many kinds of reasons, including different interpretations of what counts as a motive and a justification, there is room for a large number of mixed positions on the moral status of paternalism. However, the two end point strategies are perhaps the most coherent ones. These are general anti-paternalism and the full rejection of anti-paternalism. General anti-paternalism holds that paternalism is never acceptable, neither in motive nor in evaluation, nor in any other kind of reason. In evaluating the desirability of product safety regulation, to determine whether or not it should be introduced, or continued, anti-paternalism directs us to disregard good that will come about through unwelcome interference. The idea of disregarding the interests of some affected people is straightforward, and common in public policy evaluation. We commonly disregard the interests of non-citizens, non-residents and future generations. Similarly, we could disregard the interests in health and safety of those people for which regulation would be an unwelcome interference.

Anti-paternalism is a typically non-consequentialist position. Non-consequentialism holds that, when a moral right or duty is at stake, other considerations are excluded or become irrelevant. Only within the side constraints set by rights and duties may we consider a broader set of more or less worthy aims.[23] Some degree of anti-paternalism therefore forms a natural part of any system of rights or duties where there is no duty to

[22] There could be moderate anti-paternalist positions that do not completely disregard certain reasons, but rather discount them in some fashion. This seems for example to be the position of Louis Groarke in 'Paternalism and Egregious Harm', *Public Affairs Quarterly* 16(3) (2002): 203–230. Discounting a reason can not be equivalent to simply attributing to the reason a lesser strength in relation to other reasons, since the moral status of paternalism does not tell us anything about the relative strength of different reasons.

[23] See e.g. Frances Kamm, 'Rights' in *The Oxford Handbook of Jurisprudence and Philosophy of Law*, ed. Jules Coleman and Scott Shapiro, Oxford University Press 2002.

protect or benefit people through unwelcome interference with them, and no right of people to be so protected or benefited. If people have a moral right to buy and sell unsafe products, and there is no conflicting right to be protected against the dangers of such products, that settles the matter against regulation. *No* other considerations than rights and duties are valid, and the avoidance of harm is simply one of these other considerations. This, however, is not a specifically anti-paternalist position. Anti-paternalism can be incorporated into a non-consequentialist theory in full through the right not to be interfered with against one's will for one's good, or in other words the right not to have one's good count as a reason for unwelcome interference. Such a right is general and holds for all unwelcome interference, regardless of whether or not there is a more substantial right to do or have something.

By telling us to disregard certain reasons in making all things considered judgements, anti-paternalism introduces a level of normative consideration that is prior to the common comparison and weighing of reasons. It may be argued that this framework makes moral judgement unnecessarily complicated, or that it unjustifiably attributes to some reasons a special trumping quality.[24] Moral rights and duties may be invoked as values, but their relative importance must always be measured against the importance of other considerations. To reject anti-paternalism is to hold that all reasons should be admitted into the process of comparing and weighing reasons. No commitment is thereby made concerning the relative importance of different kinds of reasons. The rejection of anti-paternalism is perfectly consistent with strong opposition to policies involving paternalism. Resistance to product safety regulation may take the form of insisting on the value of self-determination or autonomy, and on the greater importance of these considerations relative to the minimizing of risks and promotion of health. The five reasons for holding product safety regulation to be interfering, considered above, can count as reasons against regulation, without trumping or making invalid what reasons there are *for* regulation. If we reject anti-paternalism, liberal values can simply be assigned whatever relative importance we think they deserve, save perhaps infinite importance (which would in effect amount to anti-paternalism).

The rejection of anti-paternalism is furthermore consistent with the use of anti-paternalism-like rules of thumb. Rules of thumb that regulate what reasons to consider may for example arise through the expectations we attach to certain social roles. We should arguably put our private interests aside when we act as representatives for some organisation or agency, even if these interests are normally appropriate motives for action.[25] This moral demand is merely instrumental, however, and does not mean that our private interests lose their normative importance. Rather, the private interests of all are best promoted if we sometimes disregard our own interests. Perhaps, similarly, policy makers should sometimes put aside the interests of people facing unwelcome interference, because this is expedient. It may be that our interests in freedom from interference in some area is so great, and our other interests so small or difficult to

[24] Feinberg (p. 26) explicitly calls autonomy a 'moral trump card'.
[25] This and other reasons for disregarding reasons are discussed by Thomas Scanlon in *What We Owe to Each Other*, Harvard University Press 1998, p. 51–52.

ascertain, that the risk of error would be too great to make the effort to consider all affected interests worthwhile, or that it would simply be a waste of resources. Anti-paternalist rules of thumb will only be motivated, however, in areas where it is both wasteful to even consider and estimate all affected interests, and where this is not obvious without a rule of thumb. In view of our great interest in health, and the vast resources available for making and implementing policy in modern welfare states, such areas may be hard to find in the public health context.

CONCLUSION

Paternalism is the invocation of the good of a person as a reason for unwelcome interference with her. To the extent that product safety regulation amounts to unwelcome interference with some people, invoking the health of these people as a reason for such regulation is paternalism. Anti-paternalism requires that we disregard these reasons. Rejecting anti-paternalism means considering all relevant reasons for and against regulation, without first discarding some as invalid or inappropriate. While anti-paternalism is wide-spread and inherent in the liberal tradition, liberal values need not trump other values in order to be attributed great importance.

Is the General Product Safety Regulations 2005 a good or justified policy? This would seem to depend on its effects in terms of public health, the restriction and expansion of options, the frustration and satisfaction of preferences, and other relevant values. We should not accept the classification of a public health policy as 'paternalistic' to tell against it, without further argument. The paternalistic content of the situation must be distinguished and the moral status of paternalism must be decided in light of what this content is. Hopefully, this contribution has provided some analytical tools for making such distinctions and decisions.

Forthcoming in Dawson, A., Donckers, H. & Maes, L. (eds.) *Ethics of Health Promotion* (Springer)

Responsibility, Paternalism and Alcohol Interlocks

Kalle Grill & Jessica Nihlén Fahlquist

ABSTRACT: Drink driving causes great suffering and material destruction. The alcohol interlock promises to eradicate this problem by technological design. Traditional counter-measures to drink driving such as policing and punishment and information campaigns have proven insufficient. Extensive policing is expensive and arguably intrusive. Severe punishment may be disproportionate to the risks created in most single cases. If the interlock becomes inexpensive and convenient enough, and if there are no convincing moral objections to the device, it may prove the only feasible as well as the only justifiable solution to the problem of drink driving. Taking this to heart, the former Swedish government, supported by the National Road Administration and a 2006 final report of the Alcohol Interlock Commission, proposed that interlocks should be required as standard equipment in all cars. This article assesses two possible moral objections to a policy of mandatory interlocks: 1) That it displaces the responsibility of individual drivers, and 2) that it constitutes a paternalistic interference with drivers. The first objection is found unconvincing, while the second has only limited bite and may be neutralized if paternalism is accepted for the sake of greater net liberty. If technological development can make mandatory interlocks cost-efficient, the proposed policy seems a commendable public health measure.

INTRODUCTION

Drink driving is a grave public health problem, a top contributing factor behind the 1.26 million annual deaths in traffic worldwide (WHO, 2004). The alcohol interlock is a novel technology that promises to eradicate this problem. Interlocks have been used extensively in the US and Canada as a requirement for people convicted of repeated drink driving offences. Such requirements are slowly becoming more widespread and calls are sometimes heard for a wider use of interlocks (e.g. Wald, 2006). Voluntary programs for offenders have been carried out or instigated in Sweden, France, Belgium, Finland and Australia (Svensson Smith, Nilsson, Schönning & Sjöström, 2006, pp. 84-85).

The former Swedish government announced that alcohol interlocks would be part of the standard equipment in all new cars registered in Sweden by the year 2012. At the time of writing the new government has withdrawn their initial commitment to this policy, and the public debate is ongoing. A report on the technical, economical, and legal aspects of mandatory interlocks was presented by a special commission in the summer of 2006 (Svensson Smith, Nilsson & Schönning, 2006). The Swedish reform must be approved by the European Union before it can come into effect. Already, however, interlocks are becoming more and more common in government and commercial

vehicles, and are increasingly offered as an alternative to revoked driver's license for offenders. The car manufacturer Volvo recently announced that they will offer integrated interlocks as an optional feature for some of their sedan models. While the future of a general requirement is uncertain, interlocks are unquestionably becoming an integral part of Swedish traffic safety policy.

The main focus of this article is on two possible objections to mandatory interlocks – that such a requirement inappropriately places the responsibility for sober driving with system designers rather than with drivers, and that the policy is a paternalistic interference with voluntary risk-taking. We take these issues to be the most complex moral issues to be faced by proponents of mandatory interlocks. The two objections are closely related and should therefore benefit from shared treatment. In order to evaluate the objections, we investigate the concepts of responsibility and paternalism as they apply to the case at hand. In the main, we find the objections unconvincing and so tentatively commend the Swedish policy.

While the discussion is of general relevance, we base our inquiry mainly on Swedish data. Sweden has among the least traffic accidents per capita and the least instances of drink drivers among highly developed nations. The problem of drink driving is greater in other countries and should be a grave concern in practically all countries with heavy reliance on the car for transportation. A possible solution to this problem should be of general interest. As we shall see, the technical solution offered by the interlock may be the only justifiable as well as the only feasible way to seriously diminish drink driving.

Since drink driving is a controversial issue and since alcohol interlocks are a novel technology, we will discuss both the problem and its possible solution in some detail before moving on to the core matters of responsibility and paternalism. The second section of this article briefly describes the extent of the problem and considers the efficiency and moral status of traditional responses – mainly policing and punishment. The third section is devoted to describing the interlock, its potential to stop drink driving and the more tangible costs involved. In the fourth section, we discuss social and individual responsibility and whether and how they can co-exist in the case of mandatory interlocks. In the fifth section we discuss whether and how liberty-limiting policies involve paternalism and how rejecting or accepting paternalism affects the moral status of mandatory interlocks.

DRINK DRIVING

Estimating the impact of alcohol on traffic accidents is a complex problem, due in part to great variations in police practice and to the susceptibility of autopsy studies to error due to lower blood alcohol concentration (BAC) at the time of death than at the time of accident. As a general indication, autopsy studies in some European countries show that between 20 and 50 percent of drivers killed in accidents are intoxicated (Austrian Road Safety Board, 2003, pp. 14-20). The Swedish Commission on Alcohol Interlocks (henceforth the 'Interlock Commission') estimates that in 2004 about 108 people were

killed and 1450 severely injured in Sweden in accidents caused by drink drivers (being a large fraction of alcohol-related accidents more generally). This corresponds to 22.5% of all people killed in traffic accidents. The material cost of accidents caused by drink drivers (BAC above .2 g/l) in 2004 is estimated to about 1.5 billion Swedish krona (~€170 million) (Svensson Smith et. al., 2006, pp. 73-75). This cost includes net loss of productive contribution (estimated at 800.000 Swedish krona/death) but not the cost of law enforcement, nor costs arising in the justice and penal system. Arguably, the human cost is much higher. In the US, the National Highway Traffic Safety Administration estimates that over the last years about 17.000 people have been killed yearly in alcohol-related accidents (where at least one person involved had a BAC above .1 g/l), amounting to 40% of the total number of people killed in traffic (National Highway Traffic Safety Administration, 2004, p. 32). The total material cost of these accidents is estimated to be about 51 billion dollars (~€38 billion) for the year 2000 (Blincoe, Seay, Zaloshnja, Miller, Romano, Luchter & Spicer, 2000, p. 40).

In Sweden both the law and the general population consider drink driving a serious crime. Driving with a BAC above 1 g/l entails a minimum of one month in prison and a maximum of two years (eight years if someone is killed), in addition to revoked licence with no right to apply for a new licence for 12 to 36 months. Driving with a BAC below 1 g/l but above .2 g/l entails fines or prison for up to six months, plus suspended licence for up to 12 months. Though many prison sentences are suspended or transformed to community service (Agge, Folkesson & Sjöström, 2002), these punishments (or measures) are rather severe compared to Sweden's comparatively mild treatment of offenders generally. Even so, 90% of the population are of the opinion that punishments should be harsher and calls for harsher treatment are often heard in the public debate. A small majority of the population is also of the opinion that the legally accepted BAC should be lowered from the already very low .2 g/l to zero (Swedish National Road Administration, June 2006, pp. 4-14). This in spite of the fact that the lowering of the concentration from .5 g/l to .2 g/l has had no measurable effect on behaviour (Austrian Road Safety Board, 2003, p. 83).

In contrast, philosopher Douglas Husak (1994) has argued against regarding drink driving a serious offence. Husak points out that most cases of drink driving are not mere foolishness with no social utility. Rather, people drive intoxicated for much the same reasons they drive sober – mainly to get places. There is neither malicious intent nor extreme recklessness (Ibid., pp. 58-60). Husak argues that, risk-wise, drink driving is not all that different from other kinds of driving. Though intoxication makes driving more dangerous, so does sleepiness, stress and distracting activities such as talking on the phone, eating, shaving, reading or applying make-up. None of these other risk-enhancing factors are punishable as such, but only if they result in risky driving, which is and should be a crime in itself. This discrepancy would perhaps be motivated if intoxication was much more likely to cause accidents than was other factors. However, Husak cites studies showing that a typical driver with a BAC of 1 g/l is between three and seven times more likely to cause an accident than the typical sober driver (Ibid., p. 64). That magnitude is not enough, Husak argues, for distinguishing a quite accepted activity such

as sober driving from an activity punishable by imprisonment. Husak's numbers are in tune with the classical Borkenstein study of actual crash frequencies at various BACs (Borkenstein, Crowther, Shumate, Ziel & Zylman, 1964, p. 165) as well as the similar but more recent study by Blomberg, Peck, Moskowitz, Burns & Fiorentino (2005, p. xviii), which both assign a multiple of six to seven for the probability of causing an accident at BAC 1 g/l. Husak argues further that since the probability of being killed on a five mile drive is only one in ten million, even a tenfold increase of this probability must be negligible.

Husak's main argument hinges on two comparisons. First, there is the comparison between drink driving and sober, non-impaired driving. This comparison does not necessarily support the argument. A sixfold or tenfold increase of a small probability of grave negative consequences may well be unacceptable and punishable. Moreover, to the extent that the risks of sober driving are on a par with those of drink driving, that may be an argument against the acceptability of sober driving rather than for the acceptability of drink driving. Compared to other modes of transportation, the risks of sober driving are substantial. Indeed: 'The difference in risk between driving while intoxicated and driving while sober is less than the difference in risk between driving while sober and taking public transportation.' (Husak, 1994, p. 63) It may be argued that sober driving is legal, in spite of the risks involved, because it is socially accepted, rather than the other way around. In fact, Husak himself explores this side of the issue in another article (2004).

Second, there is the comparison between drink driving and impaired driving of other kinds. This comparison does support Husak's argument. Speeding is a contributing factor in a comparable number of lethal accidents (about 13.000 yearly in the US). However, while it is prohibited, and punished on occasion, neither the social stigma nor the legal consequences are nearly as harsh as for drink driving. It may of course be argued that rather than relaxing our stance on drink driving, we should start punishing other kinds of impaired driving (more harshly). As long as such measures are not taken, however, punishing moderate drink driving much more harshly than other kinds of risky driving is at least morally problematic. Even extreme drink driving, where the risks are much higher than for non-impaired driving, may have equivalents in other kinds of behaviour (such as extreme speeding). At high BACs there is also the additional difficulty that drink drivers are to a disproportional extent alcoholics and so possibly less responsible for their actions.

In sum, we find that Husak's argument shows that punishing drink drivers with imprisonment or severe fines is at least morally problematic. Independently of this moral problem, there is also a practical problem. Policing and punishment simply have not solved the problem, as shown by the numbers surveyed above. The deterrence effect of legal prohibition is most tangibly determined by two factors - the *severity* of punishment and the *probability* of punishment. In the case of drink driving, it seems uncertain whether the severity of punishment has any impact, possibly because the probability of detection is too low for potential convicts to consider punishment a real possibility (for a discussion of other possible explanations, see Houston & Richardson, 2004). Though

even this is disputed, it does seem likely that a higher probability of punishment would contribute to the deterrence effect (Benson, Mast & Rasmussen, 1999; Austrian Road Safety Board, 2003, p. 81-84). Increasing the likelihood of punishment would of course require increased policing, which is expensive. More convictions would also entail higher costs for administration and prisons. (It would seem rational to at least decrease punishments and legal administration and use the savings to increase policing.)

The deterrence effect of prohibition is also to a large extent dependent on social norms which help shape the subjective probability of detection and punishment. Norms also have a direct influence on behaviour independently of prohibition. Apart from policing and punishing, the main strategy for reducing drink driving has, quite properly, been information campaigns of different kinds. While such campaigns seem to have an effect, especially when used in combination with other measures such as increased policing (Elder, Ruth, Shults, Sleet, Nichols, Thomson & Rajab, W., 2004), they have proved insufficient in solving the problem. In part, this shortcoming is due to the fact that those persons that are most likely to drive with high blood alcohol concentrations are relatively unaffected by measures based on deterrence and persuasion (Beirness DJ., Simpson, Mayhew & Wilson, 1994; Coben & Larkin, 1998).

There are of course other possible responses to drink driving beyond affecting norms and policing and punishment. Alcoholism and the consumption of alcohol may be targeted generally. Bar and restaurant personnel may be trained not to serve people that are likely to drive and are approaching a certain degree of intoxication. To reduce recidivism specifically, licenses may be revoked, though many who have their licences revoked as a result of drink driving keep driving, without a license (Austrian Road Safety Board, 2003, pp. 87-88). Convicted drink drivers may be offered treatment for alcoholism, though this is expensive. Cars may also be impounded or licence plates confiscated, though these measures may affect others than the driver. More proactively, doctors may be required to report alcoholics and driver's licences may be revoked preventively, though doctors are reluctant to do so since it undermines trust and is considered a breach of confidence (Bjerre, Heed & Kers, 2004).

Some of these measures can be fine-tuned. The availability of treatment programs for alcoholics involving alcohol interlocks rather than revoked licences would most likely increase doctors' inclination to report alcoholics (in Sweden around 70 times according to Bjerre et al., 2004, p. 1818). Recidivism is rather efficiently prevented by requiring alcohol interlocks for convicted offenders. However, several studies have shown that once the interlock is removed, drivers tend to resume their old patterns of drinking and driving (see e.g. Raub, Lucke & Wark, 2003), even if this tendency can be weakened with more comprehensive and more exclusive programs, which include regular medical check-ups and which expel participants that don't meet the requirements (Bjerre, 2005). In sum, reducing recidivism as well as proactive prevention is most efficient when interlocks are used.

THE INTERLOCK

If, for safety reasons, a machine should not be used in a certain way, it is wise to incorporate some feature preventing such use into the design of the machine. If cars should not be driven by people over a certain BAC, it would be wise to simply prevent such use by technical design. The alcohol interlock promises to provide such a safety feature. This device measures the driver's BAC before the car starts, for example through an exhalation sample. The interlock is connected to the car's ignition and if the measured concentration is above the maximum set, the car won't start. With this device installed in all cars, drink driving could be virtually eradicated.

The interlock is presently in a phase of rapid technological development. As the development progresses, detection becomes more and more accurate, and circumvention becomes harder. There will always be ways for the smart and skilled to circumvent safety features, but as long as the misuse is not widespread, this is not a serious problem. It is becoming increasingly difficult to by-pass an interlock breathalyzer by having anything other than a human blow air into it. Preventing sober persons from blowing for an intoxicated driver is harder. One possibility is additional tests during driving, with failed tests leading to gradual shutdown. If this is deemed unsafe, failed test may be registered, reported and later prosecuted (safety would probably be optimized by gradual shut down at high BACs and merely reporting failed tests at lower concentrations). With a system of registration, reporting and the threat of prosecution, interlocks could also come with an override feature to be used in emergencies, without encouraging misuse of that feature.

Electronic driving licenses would ease the prevention of circumvention. With such licences, it could be registered who started a car at any given moment. Technically, the licence could be required to stay in the car during driving, making starting a car with a borrowed licence more difficult. If only certain persons are required to use interlocks, this information could be stored in the licence and accessed by the car. If certain persons are exempt from a general requirement (because they cannot breathe normally for example), this information can likewise be stored and accessed (Austrian Road Safety Board, 2003, pp. 95-96).

While a policy of mandatory interlocks would be expensive, there are a number of reasons to believe that it would be worthwhile in the long run. The technological development of interlocks means that the cost of production is steadily decreasing. The higher volumes that would be needed with a general requirement would likely lead to economics of scale that would further reduce costs. The Interlock Commission estimates that based on the available technology the future cost of having interlocks integrated into the basic design of all cars would be SEK 3.000 (~€330) per car yearly. About a third of this cost is due to the inconvenience of use (the interlock has to warm up which may sometimes take a full minute or more). The total cost for having interlocks in all Swedish cars would amount to SEK 14 billion (~€1.6 billion) yearly (Svensson Smith et. al., 2006, pp. 91-92). The Swedish National Road Administration (NRA) in an official statement (2006:22883, p. 9) regards these estimates as too conservative given the expected technological development. Still, as noted above, the annual material cost of accidents

caused by drink drivers is only SEK 1.5 billion (~€170 million). On the other hand, the human cost should certainly be given some weight, whether it be higher, lower or equal to the Interlock Commission's estimate of SEK 5.5 billion (~€600 million). Human costs should include anxiety and distress caused by the risk of accident as well as by actual accidents. In addition, the commission proposes that interlocks will have positive effects on public health more generally, mainly from earlier detection of alcoholics and lowered consumption of alcohol (Svensson Smith et. al., 2006, p. 92). Finally, one might question whether the status quo is the proper baseline – is the current reliance on cars for personal transportation beneficial as such so that any obstacle to it should be valued at its full monetary cost?

If a general requirement of interlocks in all cars should be deemed too costly, there are various options for making interlocks mandatory only for certain groups of drivers or cars. Convicted drink driving offenders is one obvious category. Young people is another possibility. Focusing on cars, possible categories include government vehicles, commercial vehicles, taxis, buses, and trucks. Again, the costs of either a general or a more limited requirement will depend on technological developments that are hard to predict. It is quite possible that in the not so distant future we will have interlocks requiring less or no air, less warming time and less service. If so, the cost would decrease dramatically. It would seem that some policy of mandatory interlocks is very much a practical possibility. While we will have reason to return to the cost aspect, most of the moral arguments below are made against the background assumption that mandatory interlocks, for some or all cars, is cost-efficient in the wide sense that we consider the death and suffering prevented worth the net material cost.

Interlocks may or may not include a logging function, registering failed tests. As interlocks have historically often been used in experimental programs, collection of data has been crucial. It may be thought, however, that such registration threatens privacy. Why should the government or anyone else know how many times I have tried to start my car after drinking if this is not in itself a crime? This may be a valid concern and it should be noted that logging and collection of data is not a necessary feature of mandatory interlocks. Interlocks may be designed not to store information of failed attempts. As noted above, failed attempts *during driving* may have to be reported in order to deter circumvention. Such reports, however, concern the criminal offence of drink driving, as well as circumvention. Respect for privacy can hardly require that these crimes not be reported. *If* information about failed attempts is stored it could be used by employers as well as by health care providers to identify people in early stages of alcoholism. Even without logging, however, people who repeatedly fail to start their cars due to high BAC may themselves realize that their alcohol habits are not healthy.

RESPONSIBILITY

According to the current Scandinavian traffic safety paradigm, the ambitious goal is that no one be killed or seriously injured in road traffic (Swedish Government, 1996/97). An important means to this end is placing responsibility for preventing traffic accidents

partly on system designers. "If road users fail to abide by the rules – for example due to lack of knowledge, acceptance or ability – or if personal injuries occur, the system designers must take additional measures to prevent people from dying or being seriously injured." (our translation, Traffic Responsibility Commission, 2000, p. 69) System designers include public and private organizations involved in the design and maintenance of roads, vehicles and transportation services, as well as those involved in the design and implementation of rules and regulations, education, surveillance, rescue work, care and rehabilitation (Swedish Government, 1996/97, p. 17). This paradigm has been criticised for eroding individual responsibility (Ekelund, 1999). The possible erosion or displacement of responsibility is the first moral objection to mandatory interlocks.

Discussions about the balance between individual and societal responsibility wage back and forth in several areas, including unhealthy diets and drug use more generally (when not driving). It is important to realize that responsibility for a given event or problem is not a zero-sum game. Making the police responsible for fighting crime does not mean that people become less responsible for the crimes they commit. In certain cases, however, shared responsibility could mean less responsibility for each party. To evaluate the claim or worry that mandatory interlocks erode individual responsibility, therefore, we need to thoroughly analyze the case at hand.

Historically, responsibility for traffic accidents has been ascribed to the driver or drivers involved. The typical response to an accident is to investigate who among those involved is to blame. Interestingly, the narrow focus on individual responsibility can be contrasted with the current trend in 'human factor' research, which tends to investigate aviation rather than road traffic. The same focus on individual responsibility used to be prevalent in aviation, i.e. blaming individual pilots for accidents, but recent research has shifted interest towards the context in which decisions are made and actions carried out (Decker, 2002). This is very appropriate – both aviation and road traffic take place in complex systems and consequently accidents in these systems tend to have complex and multiple causes.

When system designers step in to take responsibility for the context in which decisions are made, they may be filling an empty space rather than usurping individual responsibility. The responsibility ascribed to system designers is of the forward-looking kind, aimed at preventing future accidents rather than distributing blame for past accidents. Forward-looking responsibility does include an element of potential blame for future accidents, if efforts at prevention turn out to be insufficient. However, it is essentially a responsibility to get certain things done, rather than to take blame. This should be distinguished from backward-looking responsibility, which is essentially focused on distinguishing the immediate causes of an accident and on the blameworthiness of those immediately involved (Fahlquist, 2006; Dworkin, 1981). In the public debate, both kinds of responsibility ascriptions are common, though not well distinguished from each other. Importantly, the two kinds of responsibility can co-exist without the one diminishing the other. Indeed, the same kind of responsibility can be ascribed to more than one agent without necessarily diminishing responsibility. Under the Scandinavian paradigm, drivers are still expected to do their part in preventing

accidents by driving responsibly and following traffic rules. This is a forward-looking responsibility, since failure to drive safely can incur blame even when no accident occurs.

A public policy focusing on assigning backward-looking responsibility to individual road users could, for instance, emphasise incarceration. The main worry from that perspective is who is to blame for any given accident. If, on the other hand, focus is on forward-looking responsibility, alcohol interlocks is a natural way of managing the problem of drink driving – system designers are responsible to put an affordable and effective systemic solution in place and individual drivers are left with no choice but to take their forward-looking responsibility for not driving after drinking.

Are there reasons to believe that ascribing forward-looking responsibility for accident prevention to system designers will in fact make drivers feel less responsible for their driving and so less cautious? Technical systems that are very sophisticated and where almost all safety hazards are guarded by automatic systems can erode the operator's feeling of responsibility. This has been observed in airplanes, where familiarity with safety devices has led to inattention and complacency (Perrow, 1999, pp. 152-54). However, these effects result from safety devices that take over a certain task from the pilot or driver and that work continuously through the whole journey, such as a collision avoidance system. The interlock, on the other hand, merely establishes whether the driver is sober before she can start the engine. This test has no direct effect on the driving experience. It does not at all guarantee that the driver is a good one or that the safety of the driver and of other road users is automatically protected. There are many other safety features and conveniences in cars that do make drivers more passive, such as automatic transmission, cruise control and automatic breaking systems. The interlock, on the other hand, only prevents people above a certain degree of intoxication from driving and is itself passive during the journey.

Could it be that despite these considerations, people will come to think of the interlock as a general test for being fit to drive, such that they will discount the risks of driving tired, stressed or under the influence of other drugs than alcohol? This may of course be possible, all sorts of misconceptions can spread, but there seems to be no direct reason to expect such a development. It is explicit and obvious that the (standard) interlock measures BAC and nothing else. Could people come to think that activities that are not protected by interlocks are safe to perform after drinking? Again, this seems farfetched. It is obvious that many activities are risky to perform after drinking and it is common knowledge that drink driving is a serious problem. Attending to this problem should not induce people to lose their everyday experience and knowledge of the impairment that comes with intoxication. In sum, the case for claiming that interlocks erode individual responsibility seems very weak.

Drink driving is a shared, social problem not only in light of its grave aggregate consequences, but also in the sense that social norms sometimes indirectly encourage drink driving. Alcohol is a natural ingredient in social life for many people. In most European and in many other cities, public transportation is extensive and runs at night time. It is then possible for most people to engage in social life, drink alcohol, and avoid driving. However, in rural areas as well as many cities in the US and elsewhere, there is

no convenient and affordable alternative to driving, especially at night. Social norms then require one to show up at a bar or restaurant or friend's place, to drink alcohol, and then to get home in some fashion. Responsible people try to assign a designated driver or otherwise plan their getting home without driving after drinking, but this is cumbersome and it is not surprising in the circumstances that people often drive intoxicated. Especially so since every single instance of drink driving with a moderate BAC is not that dangerous, despite the severe aggregate outcome. Individuals make their own choices about how to spend their nights, but these are made against the background of social expectations, city planning, nightlife culture, laws and regulations, and technology. Should mandatory interlocks become a fact, social life would simply have to adjust to the technical circumstances. It seems likely that this would encourage ways of socializing without alcohol, extensive public transportation, and local pubs and other ways of meeting more locally.

The problem of drink driving, and of impaired driving more generally, is a problem where many individuals fail to be responsible enough, with grave aggregate consequences, but where punishment of these individuals is very costly and possibly morally unjustified. The best way to solve such a problem is to change the background circumstances. Directly influencing social norms and increasing the (subjective) probability of detection are two ways to combat the problem, but they are insufficient. Drivers will continue to make mistakes and break the rules. Profound change will only come by conscious design of the system within which individual decisions and mistakes are made. Today, the technological design of cars provides drivers with opportunities which are illegal and dangerous, such as driving very fast and driving after drinking. The danger is not only to the driver, but to other road users as well. While driving after drinking is not to be dismissed as totally lacking utility, the right to drive after drinking is arguably rather trivial and defeated by other road users' rights to safety. The government should strive to eliminate opportunities that are harmful, dangerous, and relatively unimportant. Eliminating the opportunity to drive after drinking by making interlocks mandatory, if worth the material costs, seems a perfect example of sound public health policy.

PATERNALISM

The Interlock Commission explicitly states that the purpose of the interlock is 'mainly' to protect other road-users from harm (Svensson Smith et. al., 2006, p. 2). The Swedish National Road Administration takes the same position (Swedish NRA, 2006:22883, p. 4). However, death and injury to drink drivers themselves forms a large portion of the total cost of drink driving. Both government entities base their recommendations to implement mandatory interlocks on total cost estimates. In an important sense, therefore, the desire to avoid self-inflicted harm comprise a large part of the rationale for the policy. This raises the spectre of paternalism - limiting the liberty of drink drivers for their own good. The charge of paternalism is the second moral objection to mandatory interlocks. The Interlock Commission and the Swedish NRA understandably attempt to avoid a

complex moral problem by referring to harms to others as the main rationale. However, this moral problem should not be avoided, but rather recognized and analyzed.

Mandatory interlocks are potentially paternalistic because they limit liberty.[1] People would unquestionably be freer if they did not have to succumb to a BAC test before driving. However, liberty-limiting policies are not necessarily paternalistic. All criminal laws are liberty-limiting in that people would be freer if they did not have to avoid the prohibited activity, be it murder, theft or forgery. Policies are only paternalistic in so far as they are supported by certain reasons.[2] There are in principle three kinds of reasons that may potentially justify mandatory interlocks – direct protection of others from harm, avoidance of indirect costs to others from accidents, and direct protection of drivers themselves. We will, in turn, discuss these kinds of reasons and whether or not invoking them for limiting liberty is paternalistic.

The Interlock Commission and the Swedish NRA state that the main reason for mandatory interlocks is direct protection of others from harm. Limiting liberty for this reason is clearly non-paternalistic. A liberal justice system allows liberty to be exercised only within boundaries set by concern for others. Drink driving imposes significant risks on others for no comparable benefit and so the first rationale for mandatory interlocks should be morally relatively unproblematic. Objections may possibly be raised from a libertarian point of view, from which it is always a grave matter who bears the cost of reducing risks. It is reasonably clear that libertarians can support prohibiting drink driving, since it generates widespread fear and anxiety (Nozick 1974, pp. 73-84). It is not clear, however, whether this prohibition may be ensured by technical means, and if so, who should bear the cost. It is true that the costs of mandatory interlocks, both in terms of the monetary cost of installation and service of the interlock and in terms of the inconvenience of testing, is shared by all, regardless of whether or not they would have driven after drinking themselves, and regardless of whether or not they would have been victims of drink driving. For more moderate liberals, the fair distribution of costs is intertwined with broader questions of fair distribution, and it is taken for granted that society should protect its members by general safety measures, even if these impose some costs on the collective. This is very reasonable. Some drivers are very skilled and cautious and never cause an accident. Nonetheless, these drivers have to share the cost of roadside safety barriers and speed cameras. The same drivers could probably be allowed to drive through red lights when they deemed it safe to do so – still they are inconvenienced by traffic laws shaped to suit the general population. It is an open question whether in any one case the costs are worth the benefits. It is generally not considered a form of paternalism, however, to force all to share the costs of protecting all or some from the mistakes or misbehaving of the few.

The second kind of reason for mandatory interlocks is that they prevent the incurring of costs on others, cost that are not in the form of direct harms. These indirect costs include the psychological cost of knowing that people kill themselves driving after drinking, and occasionally seeing it happen. However, the largest indirect cost is arguably the material cost to society from drink drivers causing death and injury to themselves, with subsequent need for medical attention and diminished productive contribution to

society. As noted, these costs form a large portion of total costs (we are not aware of any estimates as to how large exactly). Is it paternalistic to count the avoidance of these costs as a reason for mandatory interlocks? If it is, and if paternalism is unacceptable, these costs should simply be disregarded when considering the costs and benefits of the policy, making it much less cost-efficient.

There are good arguments on both sides of this issue. On the one hand, other people than the drivers themselves are as a matter of fact made to bear much of the material costs of drink drivers harming themselves. On the other hand, it may be argued that parts of these costs are voluntarily assumed by society, which need not provide free health care to drink drivers and which may charge these drivers for other costs of the accident, such as the cost of clearing up the road and costs resulting from delays in traffic. Furthermore, it may be argued that the net loss of productive contribution to society is something that society has no right to expect or demand from an individual, who may at any time chose to end her life, or to live in ways that provide no net contribution, if she can do so without infringing on the rights of others. If this individualistic argument is correct, the costs incurred by drink drivers harming themselves and disrupting traffic are costs that they should themselves bear and so accepting the avoidance of these costs as a reason for limiting their liberty is indeed paternalistic.

Against the individualistic argument it may be argued that most of us want to live in a humane society that provides (emergency) health care to all and that we are within our rights to create such a society. If so, at least those costs of accidents and care that can not be paid for by those directly responsible are quite properly costs to the collective. It may also be argued that distinguishing between on the one hand the sober or insured, who deserve health care, and on the other hand the intoxicated and uninsured, undeserving of health care, would incur administrative costs, possibly as large or larger than those of providing care also for the undeserving. If so, it seems that mandatory interlocks do in the end prevent the incurring of costs on others.

Yet again, the pain one feels when others bleed to death in the street is perhaps an other-regarding pain, caused by one's own sensitivity and so an improper ground for limiting liberty. Furthermore, the administrative costs may possibly be charged to the undeserving, so that drink drivers would not only bear their own costs, but would also pay for the administration that decides whether or not they are deserving, similar to the way in which people are sometimes made to bear court costs when they lose a civil court case.

It seems that the final judgement as to whether the second kind of reason for limiting liberty is paternalistic or not depends on whether or not one favours a welfare state with free emergency health care and an ambition to avoid unnecessary suffering regardless of its cause. This background assumption should be recognized. If we reject paternalism and still accept that indirect costs provide grounds for mandatory interlocks we should admit that we take the welfare state or a humane society for granted, or provide some other explanation for why these costs are relevant.

The third kind of reason concerns to the direct protection of drivers themselves. We may think that saving people from being killed or injured through their own drink driving is a good reason for mandatory interlocks, independently of the resulting material cost to society. In terms of cost-benefit analysis, this attitude entails putting the 'human cost', the loss of quality of life, to drink drivers themselves on the scales. This clearly amounts to paternalism in the sense of limiting the liberty of drivers for their own good. However, at this point we should distinguish between 'hard' and 'soft' paternalism, where the former is the limiting of people's voluntary choices, while the latter is the limiting of choices that are substantially involuntary, or not voluntary enough (to warrant protection from interference) (see e.g. Feinberg, 1986, chapter 19).

Intoxication is a standard case of impairment not only in the context of driving a car, but also in the context of making rational decisions. The decision to drive after drinking is to some extent impaired and so less than perfectly voluntary. At some degree of intoxication, the decision is substantially involuntary – not voluntary enough that benevolent usurpation of that decision qualifies as hard paternalism. Moreover, drink drivers who are alcoholics may not only be acting involuntarily when they chose to drive, but also when they get themselves intoxicated in the first place. This point bears also on the indirect costs discussed above – even on an individualistic account it may not be paternalistic to avoid indirect costs by limiting liberty, if they are brought about involuntarily. Varying estimates indicate that about 50 per cent of drivers killed after drinking are alcoholics (Brinkmann, Beike, Köhler, Heinecke & Bajanowski, 2002; Swedish NRA, June 2002, p. 8). Still, not all drink drivers are alcoholics and not all alcoholics always act substantially involuntarily. Presumably, some drink drivers are acting voluntarily (enough). Unless (hard) paternalism is accepted, the costs incurred by these drivers would have to be disregarded, again making a policy of mandatory interlocks less cost-efficient. Exactly which costs should be disregarded depends on where the line is drawn, in this particular context, between voluntary and not voluntary enough.

Importantly, the fact that there exists a paternalistic rationale for mandatory interlocks in no way affects the reasonableness of other rationales. A paternalistic rationale is not something that stains a policy so that its mere existence makes the policy less justified than it would otherwise have been. The moral status of paternalism determines whether or not the protection of the very people whose liberty is limited (and who act voluntarily enough) should be accepted as a contributory reason for a given policy (Grill 2007; Husak 2003). If it should not, other reasons for that policy remain in full force.

As already noted, the Interlock Commission and the Swedish NRA have no clear position on the issue of paternalism. They point out that they support mandatory interlocks 'mainly' for other reasons, while they include the costs of death and injury to drink drivers themselves in their calculations, without commenting on the possible inconsistency. This is perhaps the standard procedure in public policy matters – the least controversial reasons are the ones officially cited, while costs are taken into consideration regardless of whether or not they are self-inflicted (and voluntary). Such a procedure

implicitly entails comprehensive acceptance of paternalism – in the actual policy decision the avoidance of voluntary self-harm is assumed to be as valid an aim as the avoidance of involuntary self-harm or harm to others.

It is far from clear that paternalism as understood here should be rejected. It is common to hold that paternalism involves some sort of bad or wrong, but this may simply be because it entails a cost in terms of liberty or autonomy. The principled anti-paternalism that holds that this cost cannot be outweighed by any benefit is uncommon and arguably unreasonable. Naturally, if paternalistic reasons are accepted as valid, they must still be balanced against other reasons, such as reasons that concern the value of liberty. Barring a principled anti-paternalism, the liberty cost of mandatory interlocks should most obviously be compared to the corresponding liberty gain. It is not obvious that interlocks entail a greater limitation or interference with liberty than do policing and punishment. On any one occasion, being forced by the police to undergo a random exhalation test is surely more inconvenient and intrusive than being forced by the technical design of the car to do the same thing. Random police tests are less intrusive only to the extent that they are less frequent. Of course, the less frequent they are, the less efficient they are. If comprehensive, efficient policing would be acceptable, so would interlocks. Would it? We propose that in the case of drink driving, as well as any other activity that should be prevented because of its potential destructiveness on any single occasion (and not just because of the accumulative effect of activities of that type), extensive policing is in principle acceptable, as long as it is not too costly or too inconveniencing. To the extent that interlocks can become cost-efficient and non-inconveniencing then, they are acceptable, and less intrusive than policing. As for imprisonment, it is of course the most severe interference when it is actually carried out. Again, the small probability of actually being punished may make a policy of policing and punishment less interfering, but to that extent also less efficient. In comparison with the amount of liberty taken away by imprisonment or even by heavy fines and/or revocation of one's driver's licence, the inconvenience of the interlock and the loss of the freedom to drive intoxicated seem rather trifling.

To sum up, it makes little sense to hold that a policy of mandatory interlocks would be paternalistic as such, since it is supported by strong non-paternalistic reasons. The fact that it may also be supported by paternalistic reasons does not change this fact. The question is, rather, whether paternalistic reasons should be allowed to bear on the issue. Such reasons are assumed to be valid in official investigations of the costs of drink driving. This seems to us very reasonable, as long as the costs of limiting liberty are not forgotten, but properly weighed against other, perhaps more tangible costs. A look at the liberty costs of policing and punishment indicates that these costs are comparable to the liberty costs of mandatory interlocks. If, contrary to our tentative position, paternalistic reasons should be disregarded when deciding whether or not to implement mandatory interlocks, the first step should be to look closer at which costs of drink driving are costs to drink drivers themselves. If soft paternalism should be acceptable, but not hard paternalism, a further important issue is to what extent drink drivers are acting

voluntarily, especially in view of the fact that many, in particular at higher BACs, are alcoholics.

CONCLUSION

Drink driving is a societal problem of great proportions. Punishing drink drivers has proven an insufficient measure and it may be questioned if harsh punishment is morally justifiable. The interlock offers a technological solution to the problem. The costs are at present too high to make a policy of mandatory interlocks in all cars cost-efficient in the short run. However, technological development might change this estimate, especially if stimulated by large orders. Should a comprehensive program still be too expensive, various limited programs are possible.

We propose that the responsibility for dealing with drink driving is to a large extent the forward-looking responsibility of system designers, including politicians. Individuals should take responsibility for their choices, but choices are always made in a context and this context can be changed by system design. It is quite consistent to hold system designers responsible for the circumstances in which individual choice is made, while at the same time holding individuals responsible for the choices they make in these circumstances. Furthermore, there seems to be no cause for worrying that greater social responsibility for system design will erode the individual feeling of responsibility for driving in the case of mandatory interlocks. Unless such a cause can be identified, sound public health policy favours social responsibility in this case.

We propose that paternalistic reasons may well be acceptable as long as the cost in terms of limiting liberty is recognized and considered. By default, the government should promote public health when it is cost-efficient to do so and when doing so does not involve a net loss of liberty or other important values. Should paternalistic reasons nonetheless be rejected as invalid, and certain costs therefore excluded from cost-benefit analysis, great care should be taken to distinguish exactly what these costs are. Regardless of whether these costs are included or not, there are strong reasons for society to combat drink driving, as it presents an obvious risk of harm to others. Given technological development, the interlock may soon be the only justifiable as well as the only feasible way to seriously diminish drink driving.

---

[1] Joel Feinberg (1986, p. ix) explicitly defines paternalism in terms of 'limiting liberty'. Other definitions speak instead of ”interference with liberty of action” (Mill, 1991 [1859], p. 14), ”violation of autonomy” (Dworkin, 1983, p. 107), or use similar expressions referring to a diminishing or disrespect of some liberal value.

[2] Most discussion of paternalism takes for granted that what is paternalistic is an action, law, institution or policy. Whether or not a policy is paternalistic then depends in part on what reasons motivates or justifies the policy. In opposition to this standard account, we assume here that what is paternalistic is the invocation of certain reasons for a policy etcetera. For a defence of this account, see Grill (2007).

REFERENCES

Agge, M., Folkesson, C. & Sjöström, L.O. 2002. *Vem bryr sig? – rattfylleriets omfattning och konsekvenser* (Who cares? – the extent and the consequences of drunk driving). Stockholm: Motorförarnas Helnykterhetsförbund.

Austrian Road Safety Board. 2003. *Preventive measures to prevent driving while under the influence of alcohol/drugs – Literature study for the Swedish National Road Administration*. Vienna.

Beirness D.J., Simpson H.M., Mayhew D.R., Wilson R.J. 1994.Trends in drinking driving fatalities in Canada. *Canadian Journal of Public Health* 85: 19-22.

Benson, B.L., Mast, B.D. & Rasmussen, D.W. 1999. Deterring drunk driving Fatalities: an economics of crime perspective. *International Review of Law and Economics* 19: 205-225.

Bjerre, B. 2005. Primary and secondary prevention of drunk driving by the use of alcolock device and program: Swedish experiences. *Accident Analysis and Prevention* 37: 1145-1152.

Bjerre, B., Heed, B., & Kers, S. 2004. Bara 1 av 1000 alkoholberoende anmäls enligt Körkortslagen (Only 1 out of 1000 alcohol dependants are reported in accordence with drivers licence law). *Läkartidningen* 101: 1814-1819.

Blincoe, L., Seay, A., Zaloshnja, E., Miller, T., Romano, E., Luchter, S., Spicer, R. 2000. *The economic impact of motor vehicle crashes 2000*. Washington D.C.: National Highway Traffic Safety Administration, U.S. Department of Transportation.

Blomberg, R.D., Peck, R.C., Moskowitz, H., Burns, M. & Fiorentino, D. 2005. *Crash Risk of Alcohol Involved Driving: A Case-Control Study*. Stamford CT: Dunlap and Associates Inc.

Borkenstein, R.F., Crowther, R.F., Shumate, R.P., Ziel, W.B. & Zylman, R. 1964. *The Role of the Drinking Driver in Traffic Accidents*. Indiana University.

Brinkmann, B., Beike, J., Köhler, H., Heinecke, A., & Bajanowski, T. 2002. Incidence of alcohol dependence among drunken drivers. *Drug and Alcohol Dependence* 66: 7-10.

Coben, J.H. & Larkin, G.L. 1998. Effectiveness of ignition interlock devices in reducing drunk driving recidivism. *American Journal of Preventive Medicine* 6: 81-87.

Decker, S. 2002. *The field guide to human error investigations*. Aldershot: Ashgate.

Dworkin, G. 1981. Voluntary health risks and public policy. *The Hastings Center Report*, 11(October): 26-31.

Dworkin, G. 1983. Some second thoughts. In *Paternalism*, ed. R. Sartorius, 105-111. Minneapolis: University of Minnesota Press.

Elder, R.W., Shults, R.A., Sleet, D.A., Nichols, J.L., Thomson R.S., & Rajab, W. 2004. Effectiveness of mass media campaigns for reducing drinking and driving and alcohol-related crashes. *American Journal of Preventive Medicine* 27: 57-65.

Ekelund, M. 1999. *Varning – Livet kan leda till döden! En kritik av nollvisioner* (Warning – Life can lead to death! A critique of Vision Zero:s). Stockholm: Timbro.

Fahlquist J.N. 2006. Responsibility ascriptions and Vision Zero. *Accident Analysis and Prevention* 38: 1113-1118.

Feinberg, J. 1986. *Harm to Self*. Oxford: Oxford University Press.

Grill, K. 2007 The Normative Core Of Paternalism. *Res Publica* 13: 441-458.

Houston, D.J., & Richardson, L.E. Jr. 2004. Drinking-and-driving in America: A test of behavioural assumptions underlying public policy. *Political Research Quarterly* 57: 53-64.

Husak, D. 1994. Is drunk driving a serious offence? *Philosophy and Public Affairs* 23: 52-73.

Husak, D. 2003. Legal Paternalism. *The Oxford Handbook of Practical Ethics*. Oxford: Oxford University Press.

Husak, D. 2004. Vehicles and crashes. Why is this moral issue overlooked? *Social Theory and Practice* 30: 351-370.

Larkin, G.L. 1998. Effectiveness of Ignition Interlock Devices in Reducing Drunk Driving Recidivism. *American Journal of Prevention Medicine* 16: 81-87.

Mill, J.S. 1991 (1859). *On Liberty*. In *On liberty and other essays*. Oxford: Oxford University Press.

National Highway Traffic Safety Administration. 2004. *Traffic Safety Facts 2004 – A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System*. Washington D.C.: U.S. Department of Transportation.

Nozick, R. 1974. *Anarchy, State, and Utopia*. New York: Basic books.

Perrow, C. 1999. *Normal accidents*. Princeton: Princeton University Press.

Raub, R.A., Lucke, R.E. & Wark, R.I. 2003. Breath Alcohol Ignition Interlock Devices: Controlling the Recidivist. *Traffic Injury Prevention* 4: 199-205.

Svensson Smith, K., Nilsson, M., & Schönning, O. 2006. *Öppna möjligheter med alkolås – Slutbetänkande av Alkolåsutredningen* (Open possibilities with alcohol interlocks – Final report of the Commission on Alcohol Interlocks). Stockholm: SOU 2006:72.

Svensson Smith, K., Nilsson, M., Schönning O., & Sjöström, L. 2005. *Alkolås – Nyckeln till Nollvisionen – Delbetänkande av Alkolåsutredningen* (Alcohol Interlocks – The key to Vision Zero – Partial report of the Commission on Alcohol Interlocks). Stockholm: SOU 2005:72.

Swedish government. 1996. Regeringsproposition 1996/97:137 (Governmental bill). Stockholm.

Swedish National Road Administration. June 2006. *Alkohol, droger och trafik* (Alcohol, drugs and traffic). Stockholm.

Swedish National Road Administration 2006, Yttrande TR65 A 2006:22883 (Statement). Stockholm.

Traffic Responsibility Commission. 2000. *Ett gemensamt ansvar för trafiksäkerheten – Betänkande av Trafikansvarsutredningen* (Collective responsibility for traffic safety – Report by the Traffic Responsibility Commission). Stockholm: SOU 2000:43.

Wald, M.L. 2006. A New Strategy to Discourage Driving Drunk. New York: *The New York Times*, November 20.

World Health Organization. 2004. *World report on road traffic injury prevention*. Geneva.

Journal of Medical Ethics 31: 648-653

Epistemic paternalism in public health

Kalle Grill & Sven Ove Hansson

ABSTRACT: Receiving information about threats to one's health can contribute to anxiety and depression. In contemporary medical ethics there is considerable consensus that patient autonomy, or the patient's right to know, in most cases outweighs these negative effects of information. Worry about the detrimental effects of information has, however, been voiced in relation to public health more generally. In particular, information about *uncertain* threats to public health, from—for example, chemicals, are said to entail social costs that have not been given due consideration. This criticism implies a consequentialist argument for withholding such information from the public in their own best interest. In evaluating the argument for this kind of epistemic paternalism, the consequences of making information available must be compared to the consequences of withholding it. Consequences that should be considered include epistemic effects, psychological effects, effects on private decisions, and effects on political decisions. After giving due consideration to the possible uses of uncertain information and rebutting the claims that uncertainties imply small risks and that they are especially prone to entail misunderstandings and anxiety, it is concluded that there is a strong case against withholding of information about uncertain threats to public health.

Keywords: epistemic paternalism; public health; withholding of information; uncertain information; MCS (multiple chemical sensitivity)

It is usually taken for granted that access to information is a good thing. To be informed means to be closer to the truth and to be able to make informed decisions. However, information can also have negative effects. A much discussed example is the negative psychological effects on a patient of information about a bad prognosis. Previously, a paternalistic approach to such information had a strong influence on medical practice and on medical ethics. According to this approach, a physician often best serves a patient's interests by withholding negative information that the patient has asked to be told. In the last few decades, such epistemic paternalism[1] has for the most part been replaced by a strong emphasis on the patient's right to know.[*]

Today, the prevalent presumption is that the patient has a right to know what the physician knows about her condition. This right to know is explicitly related to the right to make decisions about one's own welfare. Treatment should normally be given, only with the patient's informed consent, which is only possible if she has full access to information about her condition.[2]

[*] The term epistemic paternalism is borrowed from Goldman. He first introduces the term as referring exclusively to withholding of information in the subject's best epistemic interest, (Goldman,[1] pp 118–19) but later includes extra-epistemic reasons for withholding under the same concept (Goldman,[1] p 127).

Contrary to this development in medical ethics, there is in some quarters a growing concern over information to the public about general health threats. Sceptics about the value of such information argue that it can be extremely costly both in monetary terms and in terms of the social costs of needless worrying. These apprehensions are aggravated by the increased capacity to detect potential risk factors. As was noted by Viola Vogel, "[n]ew nanoanalytical tools are pushing detection limits down to the single molecule level, which is scientifically a huge success but could be a potential headache to regulators. Ultrasensitive detection of toxins and pollutants will alarm the public."[3]

In a recent article on the regulation of chemicals, Bill Durodié suggests that the dissemination of information about potential adverse effects of chemicals to the public has predominantly negative effects, since it contributes to making people anxious and depressed.[4] These views are expressed in relation to the European Commission's proposed new system for the *R*egistration, *E*valuation, and *A*uthorisation of *Ch*emicals (REACH).[5]

In his criticism of REACH, Durodié attacks what he rightly describes as a contemporary consensus on the public's right to know:

> One common assumption in much of the current debate on issues relating to scientific reporting and decision making is that the public have a "right to know" and should be informed whenever and wherever there is any scientific uncertainty associated with products and processes (Durodié,[4] p 393).

According to Durodié, "this 'right' would appear to suggest that consumers should be permanently bombarded with reams of information" (Durodié,[4] p 393). In his view, this will have both short run and long run negative effects. In the short run, "enhanced risk awareness" could "leave us feeling more sorry than safe" (Durodié,[4] p 394). In the long run, public anxiety and fear in the face of technological development could stifle progress:

> [B]ringing up a generation of people in fear of everyday products, questioning the ability of science to improve their lives, and hence doubting the desirability of innovation and change, has a social cost which has yet to be calculated (Durodié,[4] p 389).

It is not obvious exactly what Durodié is critical of. Misgivings about certain public awareness campaigns are seemingly taken to imply a negative appraisal of risk awareness in general and the REACH proposal in particular (Durodié,[4] p 393–4). Such campaigns have, however, little in common with the commission's proposal to make information about chemical testing available in a central database for "free and easy access". According to the current proposal of the commission, non-confidential information on registered substances will be kept available on the internet by a proposed new European chemicals agency. This information will include chemical nomenclature; physicochemical data; the results of each submitted toxicological study; no effect levels and no effect

concentrations when available; most of the information on the safety data sheet, and guidance on safe use provided by the company.[6]

The main focus of Durodié's critique is on the "social and hidden cost that these proposals [by the European Commission] entail" (Durodié,[4] p 393). Central to the critique is the (reasonable) assumption that many of the risks that will be inferable from the proposed database will be uncertain, and will remain so even after extensive research. Durodié calls for "a more measured approach to risk communication", as opposed to "feeding the climate of risk aversion" (Durodié,[4] p 394, 396).

Though not explicitly advanced in favour of withholding of information, Durodié's sharp critique of the proposed extent of communication implies an argument for some kind of limiting of the dissemination of information about uncertain threats to public health. Being supposedly in the best interest of the people concerned, such a proposal amounts to epistemic paternalism. It is the purpose of the present paper to develop the argument for epistemic paternalism and to evaluate it from an ethical point of view. In section two, we further develop the notion of making information available and distinguish it from actively communicating information. Section three introduces our methodology and some preconditions for the analysis that follow from it. After that we investigate the major types of positive and negative effects of withholding information on public health—namely effects on knowledge (section four); psychological effects (section five); effects on individual decision making (section six), and effects on political decision making (section seven). Our overall conclusions are summarised in section eight.

2. WITHHOLDING, MAKING AVAILABLE, AND ACTIVELY COMMUNICATING

Our focus of attention is on the dissemination of information about public health—that is, general health information that has not been adjusted to the individual recipient's health status, but nevertheless has (prognostic or preventive) relevance at least for segments of the population. Typical examples are "smoking is bad for your health" and "acryl amide in fried potatoes may increase the risk of contracting cancer". We will pay particular attention to public health information that is based on less than conclusive scientific evidence, and therefore uncertain.

Our discussion presupposes that a distinction can be made between actively providing a person with a piece of information and (just) keeping that information available for her to seek out herself. In clinical medicine, a reasonable case can be made that the physician has at least a prima facie obligation to actively inform the patient about her health status and ensure that she has understood the information. Inessential details, such as individual laboratory values, should be made available to her upon request. In public health, authorities and companies have corresponding obligations to actively disseminate the more important information. All affected members of the public should be actively informed of serious risks. Hence, if the tap water in an area becomes undrinkable, all those who have access to this water have an obvious right to be informed about this. Lesser risks, such as a small increase in the concentration in tap water of some contaminant that is still far below agreed upon levels of concern, should

perhaps not entail active informing, but such information should none the less be made available so that those who seek it may find it.

How to draw the line between information that should be actively communicated, and information that should not be actively communicated but only made available is an interesting ethical issue. That is not, however, the topic of the present contribution. Instead, we will discuss whether there is another line to be drawn, namely between on the one hand public health information that should be made available, and on the other hand public health information that should not even be made available, but rather be withheld from the public. The proposal in REACH to post information about registered substances on the internet is a clear example of making public health information available without necessarily communicating it actively to potential readers.

Whether or not information should be made available may seem to depend on how the information is managed once made available. We will, however, not discuss the management of information. Our narrow focus on making available or withholding is motivated by several circumstances. First, the problem has practical significance, since information holders such as scientists, companies, and public officials often face the choice of whether to make some piece of information available for further dissemination—for example, by putting out a press release or not without having much influence on how the information is managed once released. Second, although ineffective awareness campaigns should obviously be avoided, questions of research priorities and information management are largely empirical and very complex, demanding—for example, a discussion of the role of the free press. The more limited issue of whether or not information should be withheld can, we believe, be successfully dealt with somewhat independently and in the abstract. Finally, a clear presentation and evaluation of the case for withholding information will hopefully shed some light on related, more complex issues.

It is of course not crystal clear what it means to make information available without actively disseminating it. Information can be made available in quite different ways. At one extreme, the information holder can actively inform the public about what type of information is available and how one can find it (without spreading the information itself). At the other extreme, information may be kept available simply by being buried in some unindexed but publicly accessible archive. It is, however, doubtful whether the last mentioned practice qualifies as making the information available. The REACH proposal of providing information for "free and easy access" on the internet is at the more open end of this spectrum.

An individual scientist cannot keep information to herself without sacrificing the fundamentally important scientific discussion. Strict withholding of scientific information would therefore demand not only that individual scientists avoid actively providing information to people outside of the scientific community by—for example, answering questions from journalists or other non-scientists (however the division into scientist and non-scientist would be made). The scientific community would also have to develop codes of behaviour that exclude the publication of uncertain public health information, or else the government would have to impose legal restraints on reporting scientific

results indicating uncertainties about public health. Such arrangements would have to be internationally agreed upon in order to be efficient. Clearly, such a system would hamper scientific progress by reducing the flow of information within the scientific community as well as in the wider world.

Information may, however, be withheld to some extent without going to such lengths. Questions may be answered, but incompletely. Publications can be made, but not popularised and not discussed in public fora. We will discuss indirect effects of withholding information with the understanding that its effects will in practice depend on how strict the withholding is and how smoothly the practice of withholding is defended, denied or downplayed in communication with the public. In the process of defending a practice of withholding, there is always the risk of resorting to outright deception. Deception destroys trust much more thoroughly than the withholding of information. Having been kept in the dark, you could still trust what information has come and will come from the information holder. Confronted with liars, however, there is nothing left to trust.[7] Government information or scientific results will be met with deeper scepticism the more the government or the scientists are engaged in deceitful practices. The same thing applies to these groups as to doctors: "Were trust to decline so that patients did not believe what was being said to them, not only reassurance but also genuine support during an illness would become impossible."[8]

3. METHODOLOGICAL CONSEQUENTIALISM AND THE RIGHT TO KNOW

At the core of liberal political thought is the individual's right to direct her own life. Though a choice between two unknowns would still be a choice, what we value is people's right to make *informed* choices about their own lives. The right to direct your life in a meaningful way, to make informed choices about your own life, thus implies some version of a right to know. Much liberal thought, of course, goes on to argue that the right to direct your own life is an absolute right that cannot justifiably be infringed. If the same goes for the right to know, then the case for justified withholding of harmful information uninterestingly fails, so that no amount of harm can cancel it. We will not pursue this line of thought further, but rather consider the more difficult case of consequentialist ethics.

Consequentialism is the "hard case" for the defence of the public's right to information. In order to explore the best arguments for the withholding of information, we will therefore assume a form of *methodological consequentialism*. In other words, we will assume that the issue of whether or not information on public health should be held available for the public has to be determined by weighing the positive against the negative consequences of making it available.

Methodological consequentialism does not preclude a discussion in terms of rights and duties. Deontological language is convenient to describe the moral status of disseminating information. We use this terminology with the understanding that from a consequentialist point of view, rights and duties are not foundational moral principles, but rather action guiding rules that are adequate to the extent that they tend to maximise

the good. Clearly, from that point of view, to the extent that there is a right to know, this is only a prima facie right that can be overridden by other considerations. For reasons of terminological convenience, we use the phrase "right to know" to denote the right to have access to information rather than the right to actually receive it. A right to know is, in this sense, in principle compatible with a possible right not to actually receive the information (if one chooses not to ask for it). Such a right is sometimes inferred either from the right to direct one's own life or from some version of a right not to be harmed. For an overview, see Chadwick.[9]

Furthermore, we will not reduce the value of the consequences to some unitary value, such as happiness or preference satisfaction. Instead, we will operate with values that have intuitive appeal and that may either be considered intrinsic or instrumental in contributing to more basic values. As mentioned in the introduction, four types of effects will be considered. Corresponding to Durodié's concern with short run negative effects, we will consider effects on people's psychological wellbeing. Corresponding to his concern with long run negative effects, we will consider effects on political decisions. We will furthermore consider effects on knowledge, primarily for their instrumental value. Finally, we will consider effects on private decisions, where the connection to informed control over one's own life is most obvious.

We take none of these areas to be supremely important. Rather, we consider each area in its own right, on the assumption that they can in some way be incorporated into a wider consequentialist framework. Should the making available of some pieces or types of information lead to great enough harm, not countered by positive consequences of the same magnitude, such information could justifiably be withheld. This is a paternalistic stand that we find quite reasonable from a consequentialist point of view. The question is thus whether the balance will in practice be in favour of withholding information about public health, or in favour of making it publicly available.

4. EFFECTS ON KNOWLEDGE

Apart from its possible intrinsic value, improved knowledge typically contributes to a person's decision making capacity. Positive practical effects of public health information will ensue only if access to (correct) information of this kind gives rise to better knowledge. It is therefore important to consider the effects of public health information on a person's state of knowledge.

We take for granted that under most circumstances access to more accurate information will improve a person's knowledge. There are cases, however, in which (true) information confuses the recipient or causes her to have a more distorted view of the state of the world. This can happen if the new information connects with false beliefs already held. There seem—for example, to be persons who believe that (ionising) radiation is "contagious" so that an irradiated object will itself emit radiation. If a person with this incorrect belief is told that a certain spice product has been subjected to irradiation, then she may conclude that the product is radioactive. Similar mistakes are

the subject of numerous studies of risk communication between experts and lay persons. For an interesting exposition see Thomson P B.[10]

This problem may also arise in relation to uncertain information. Hence, in our example, if this person receives information saying that there is uncertainty whether or not a specific product has been irradiated, she will probably conclude that there is uncertainty whether or not it is radioactive. It is essential to note, however, that this problem is not specific to uncertain information. A policy of withholding information in order to avoid negative effects of information on knowledge would imply withholding both certain and uncertain information in those areas where misinterpretation is most likely. These would be areas where the meaning of terms differs between expert and lay use, where there is widespread prejudice, or where there is a general lack of knowledge which is compensated for through superstition or wild guesses. The sheer magnitude of prevalent misconceptions and possible misunderstandings, and—not least—their unpredictability, makes such a project of withholding a rather implausible undertaking.

Let us, however, for argument's sake, consider the possibility that information about uncertain threats to public health is one of those areas where misunderstandings are most likely to have negative effects on knowledge, or where the effects are most severe. There would then be a case for the withholding of such information in order to avoid adding to existing confusion. How strong the case is depends on the likely outcome of withholding.

An obvious danger with withholding is that awareness of the lack of information will create greater confusion than would the information itself. The knowledge or suspicion that information is withheld can induce false beliefs that involve the attribution of exaggerated weight to withheld information: "if they don't tell us, it must be important". Furthermore, such information as would become available under a practice of withholding would not only tend to be partial and unbalanced, but would also likely be misinterpreted due to suspiciousness and to lack of relevant background information. The attempt to limit uncertainty might thus create even greater uncertainty about the state of things. As long as information is made available about the uncertain threats to health that science discovers, at least we know (or can find out should we want to) what these uncertain threats are, and we need not speculate about other threats that scientists might possibly have discovered. If they are not made available, the scope for such speculation is wide open.

Generally speaking, when a person or group of persons have misconceptions that give rise to adverse effects of information, withholding information that can be misconceived seems to be a dubious strategy. The obvious solution is instead to provide information that counteracts the misconception, both in general and in connection with particular pieces of information that may otherwise be misunderstood. In the above example of irradiated food, this means that (i) general education about the physical characteristics of radiation should be promoted, and (ii) information about the irradiation of food products should when necessary be accompanied by information that, as far as possible, forestalls potential misunderstandings.

5. PSYCHOLOGICAL EFFECTS

Information often threatens our wellbeing and peace of mind. Bad news can be harmful simply by making us aware of displeasing facts. In everyday life, we act under the presumption that such harm is outbalanced by the usefulness of being aware of the state of things. In most cases this presumption is empirically well founded. This is why we do not in general need to hesitate to tell a student that she has failed an exam, or a patient that she has diabetes. There are cases, however, where we tend to hesitate, such as telling a friend that her partner cheats on her or a patient that she has a deadly, incurable disease. In these and some other situations it can reasonably be claimed that an affected person would be better off without the information.

How harmful, then, is information about uncertain threats to public health? This is, at least in part, an empirical issue. However, there does not seem to be an adequate information base for answering it. Durodié refers to an article by Winters *et al*, who have carried out a series of experiments in which adverse symptoms, mainly from the respiratory organs, were induced by inhalation of air with enhanced CO_2 concentration. By adding an odour to the CO_2 enriched air, subjects were conditioned to exhibit these symptoms also when exposed to the odour alone. This associative (Pavlovian) learning was substantially enhanced in subjects who were, prior to the exposure, given a leaflet that, according to the authors, contained "information similar to that found on websites and other media about environmental pollution and a description of a patient with MCS (multiple chemical sensitivity)".[11] Durodié also cites Simon Wessely as claiming that Sweden, with its restrictive chemicals regulation, has one of the highest levels of self reported sensitivities to chemicals in the developed world, but we could not find this claim in the paper referred to. (Wessely S. Psychological, social and media influences on the experience of somatic symptoms. Workshop paper, 1997. Manuscript received from author, via personal communication, in 2004.)

According to Durodié, these results indicate that official recognition of health threats contribute to the proliferation of such diseases as MCS (Durodié,[4] p 394). The experiment conductec by Winters *et al* does not, however, establish a link between information about chemical substances of the type proposed by regulatory agencies and psychosomatic diseases such as MCS, and this for at least two reasons. First, the leaflet contained sweeping formulations about chemicals in general, very different from the specific science based statements about potential problems with specific substances that are intended for instance in the REACH proposal criticised by Durodié. Second, MCS was not reported by Winters *et al*.[11] It remains to be shown that MCS or related psychosomatic disorders can be caused by Pavlovian learning or by some other mechanism for which results from Pavlovian learning are relevant.

As with effects on knowledge, the psychological harms of information need to be evaluated in relation to the harms of withholding information. Withholding must be organised in some fashion and indirect effects will ensue. Lack of information may induce not only confusion but also anxiety, perhaps in particular if one is aware of the fact that information is being withheld in order to prevent anxiety. We find it hard to believe that the type of anxieties that may contribute to MCS would be relieved if the

present situation were replaced by one in which the public knew or suspected that there was (uncertain) scientific information about possible threats to their health that where being withheld from them. Furthermore, it is difficult to see how professionals could bring these patients to trust them under such circumstances.

In conclusion, the psychological effects are to a large extent unknown. Epistemic paternalism may lead to a situation in which the public knows that there is scientific uncertainty about the health effects of certain chemicals, and also knows that the identities of these chemicals are not disclosed to them. We propose that the latter situation is likely to have more serious psychological effects than one in which the public receives full and correct information about scientific uncertainties. Hence, there is a real possibility that epistemic paternalism will be counterproductive and withholding must be evaluated with this possibility in mind.

6. EFFECTS ON PRIVATE DECISIONS

Although Durodié is certainly right that the potential threats to public health from chemicals in products and processes are often uncertain, it does not follow that these uncertain risks are better ignored. Without going into detail, Durodié dismisses uncertain risks as an object of rational decision making: "The emphasis, promoted by some, on what could be, rather than on what is, removes human action, understanding, competence, and will from the equation" (Durodié,[4] p 393). This quote catches in a nutshell the assumption that uncertainties are not part of "what is".

Durodié claims to be writing "in the spirit of" an earlier article by Chauncey Starr, which deals with uncertainties in a somewhat more stringent manner. Durodié quotes with approval Starr's description of such alleged threats to health as global warming, radiation, and genetic modification as examples of "amplification of a minor popular concern into an apocalyptic dogma".[12] Starr is concerned with public fears that are "hypothetical" —meaning that "the guesstimate of either their probability, their consequences, or both has an indefinitely wide range of uncertainty". The wide range of uncertainty is implicitly assumed to imply that the risks are certainly minor.[12]

In opposition to the above, it must be emphasised that uncertainty does not equal complete lack of knowledge. When scientific findings imply that some apparently harmless substance might be hazardous, this means something more than the mere logical possibility that it may be hazardous. The range of uncertainty is not indefinitely wide. Maybe the substance has proven hazardous to rats, or maybe its molecular structure indicates that it can be harmful. Such indications, and others that scientists in the field deem to be relevant for risk assessment, do tell us something. A scientific judgment that there is an uncertain risk associated with some product or process is a qualified judgment, not to be confused with the situation where no such judgment has been made, and certainly not with the situation where a product or process has been judged harmless.

Hence, when the risks associated with the use of some chemical, for example, are uncertain, this means that the situation with respect to this substance differs from the

situation of *both* substances that are known to have serious adverse effects *and* from the situation of substances which we can be reasonably certain have no such effects. For a rational decision maker who strives to avoid negative health effects, these differences should have practical consequences. Everything else being equal, she should prefer the substance known to be harmless to the substance with unknown properties, and the latter to the substance known to have serious negative effects.

Information is directly useful to the degree that it actually guides concrete choices. Being informed, however, also contributes to your capacity to make informed choices, whether or not you use it. The capacity to make informed choices and thus direct your own life is what is usually called autonomy. If—for example, you wish to avoid uncertain health threats, but do not know which those threats are, you are effectively prevented from acting in accordance with that wish. In general, information is a prerequisite for meaningful choice.

Good reasons can be given for avoiding uncertain threats to health when the costs of doing so are sufficiently small. Avoiding uncertain threats to health will in the long run improve our chances of staying well, though we do not know to what degree. The point of balance will differ between individuals depending on the values on which they base their decisions. Starr mocks the precautionary principle, saying it is caused by "a primitive instinct to suspect the unknown" (Starr,[12] p 804). Individuals' negative evaluations of risks and uncertainties can more neutrally be referred to as aversions. Granted that individuals should be allowed to choose what products to use and what activities to take part in based on their own values, including their own degrees of risk aversion and aversion to uncertainty, they can obviously make use of information about the possible health risks connected with different products and activities.

7. EFFECTS ON POLITICAL DECISIONS

In a democratic society the people must have access to adequate information on the performance of the government. To be able to evaluate the government, people need to be informed of the consequences of government policy. Issues such as public health, consumer safety, and the status of the natural environment are common and fully legitimate citizen concerns. Another use of information about threats to public health is thus political.

Of course, citizens are never fully informed, and the government can be evaluated on the basis of what people do know even if this is not all there is to know. However, every bit of information that is concealed makes citizens less informed and thus less able to fulfil their political role. People are interested in different matters, and attach different degrees of importance to different types of facts. Each citizen should be free to look into those aspects of government that she finds most important or most neglected, while at the same time ignoring other aspects of government. To accept, however, that not everyone knows everything is one thing, excluding some matters from public scrutiny altogether is a completely different matter. (There are policy areas, such as foreign and

security policies, where information is withheld from the public for reasons that are not relevant for issues of public health. We will not discuss these practices here.)

Most obviously, citizens need information about uncertain health threats in order to be able to participate in the political process as it concerns regulatory issues. They also need such information in order to be able to discuss the government's research priorities and the priority given to research in comparison with other areas. Information about uncertain threats to public health is in fact an integrated part of the information base on which political choices should be made. To the extent that there is such a thing as a citizen right or duty to make informed choices among political alternatives (parties, presidential candidates, or alternatives in a referendum), there is also an inferred right or duty to be informed of the relevant facts.

In an open society, openness and sharing of information is a driver of progress not an obstacle. New discoveries are constantly made in all segments of society. The sharing of these discoveries and the building on each other's achievements is what makes possible the rapid technological development that we currently enjoy. It is this pure effectiveness of information sharing that prompted John Stuart Mill, with his otherwise sharply critical view of centralised government, to champion "the greatest dissemination of power consistent with efficiency; but the greatest possible centralisation of information, and diffusion of it from the centre". Mill believed that the central government "should have a right to know all that is done, and its special duty should be that of making the knowledge acquired in one place available to others".[13] Particularly in the age of the internet, one does not need to endorse the proposed role of central government in order to sympathise with the underlying purpose of this proposal—namely to improve the quality of decisions throughout society by making as much information as possible available to all decision makers, and, as Mill would have been eager to point out, in a democracy all citizens are decision makers.

Durodié and Starr fear that technophobia will grow and progress suffer. But we need to avoid not only technophobia, but also technomania. A proper balance must be struck between safety and development. To the extent that the public should be able to take part in political decision making, they need access to full scientific information, including information about scientific uncertainties.

8. CONCLUSION

In summary, we have found epistemic paternalism about public health to be a very problematic position, possibly to the degree of being self defeating. Withholding information about threats to public health interferes with people's ability to make informed choices in both the private and the political parts of their life. Furthermore, such a practice is likely to give rise to more confusion and unwarranted anxieties than it can prevent. This applies even if the practice of withholding is restricted to information that is uncertain.

The argument for epistemic paternalism would be stronger if great uncertainty about health effects would imply small risks. Unfortunately, things are not that simple.

History is rich with examples of threats to health that were first considered non-existent or simply not considered, later considered uncertain, and are now known to be harmful. Tobacco and asbestos are among the best known examples of this.

It should be conceded that the policy we propose has the unavoidable disadvantage that warnings will sometimes go out about things that will later be found quite safe. An instructive case is the famous Berg letter to *Science* in 1974, in which several distinguished scientists went public with their fear of the "potential hazards" of in vitro recombination of genes.[12] The letter initiated a large and sometimes untidy public and scientific debate about the risks of gene splicing. Since scientists now, including the authors of the Berg letter, agree that the potential hazards were not actual, we can conclude that the spread of these scientific doubts did little good and caused much anxiety and, possibly, false beliefs. This does not mean, however, that it was wrong to inform the public of the potential problem. It lies in the nature of uncertainties that they may later become certainties, in either direction. The appropriateness of informing about uncertainties cannot be evaluated on the basis of scientific results that were unavailable at the time. We agree with Keith Boone, writing about the controversy in retrospect, that the strong reactions toward the Berg authors, the later triumphant demeanour of the critics and the repentant attitude of the authors, were unfortunate and misplaced. Events such as the Berg letter are an integral part of an openness that has some negative consequences but is on the whole superior to secrecy. The Berg letter controversy was, it should also be noted, amplified by earlier policies of secrecy in the USA, resulting in diminished trust in public and professional figures on the whole.[15]

Withholding, like manipulation, might be the best way to achieve a short term well defined objective—for example, to avoid anxiety. In the long run, openness and truthfulness have better consequences. The REACH proposal to make information about chemical testing available to the public is a prominent example of such openness. To lessen the harm done by information, we should look at how information is communicated, doing what we can to avoid ill considered campaigns and exaggerated media reports. Even if, however, such efforts prove to be unsuccessful, it is still better to make information about uncertain threats to health available to the public than to adopt a policy of keeping them secret. As Roger Higgs so eloquently put it: "The antidote to fear is not silence but open discussion".[8]

ACKNOWLEDGEMENTS

We would like to thank an anonymous reviewer for valuable comments.

REFERENCES

1. Goldman AI. Epistemic paternalism. *Journal of Philosophy* 1991;**88**:113–31.

2. Beauchamp TL, Childress JF. *Principles of biomedical ethics* [5th ed]. Oxford University Press 2001:80–8, 283–90.

3. Vogel V. Social implications of nanotechnology in education and medicine. In: Roco MC, Bainbridge WS, eds. *Societal implications of nanoscience and nanotechnology*. NSET workshop report. Washington DC: National Science Foundation, 2001:143–8 at 145.

4. Durodié B. The true cost of precautionary chemicals regulation. *Risk Analysis* 2003;**23**:389–98.

5. European Commission. *Strategy for a future chemicals policy* [white paper]. Brussels: European Commission: 27 Feb 2001. COM (2001) 88 final.

6. European Commission. Proposal for a regulation of the European parliament and of the council concerning the registration, evaluation, authorisation and restriction of chemicals (REACH). Brussels: European Commission: 29 Oct 2003. COM (2003) 644 final.

7. Jackson J. *Truth, trust and medicine*. London: Routledge 2000.

8. Higgs R. Truth-telling. In: Kuhse H, Singer P, eds. *Companion to bioethics*. Oxford: Blackwell 1998:435.

9. Chadwick R. The philosophy of the right to know and the right not to know. In: Chadwick R, Levitt M, Shickle D, eds. *The right to know and the right not to know*. Aldershot: Ashgate Publishing Company, 1997.

10. Thomson PB. The ethics of truth-telling and the problem of risk. *Science and Engineering Ethics* 1999;**5**:489–510.

11. Winters W, Devriese S, Van Diest I *et al.* Media warnings about environmental pollution facilitate the acquisition of symptoms in response to chemical substances. *Psychosomatic Medicine* 2003;**65**:332–8, especially 336.

12. Starr C. Hypothetical fears and quantitative risk analysis. *Risk Analysis* 2001;**21**:803-806.

13. Mill JS. *On liberty and other Essays*. Oxford: Oxford University Press, 1991:126.

14. Berg P, Baltimore D, Boyer HW *et al.* Potential biohazards of recombinant DNA molecules. *Science* 1974;**185**:303.

15. Boone K. When scientists go public with their doubts. *Hastings Cent Rep* 1982;**12**:12–17.

THESES IN PHILOSOPHY FROM THE ROYAL INSTITUTE OF TECHNOLOGY

1. Martin Peterson, ”Transformative Decision Rules and Axiomatic Arguments for the Principle of Maximizing Expected Utility”, Licentiate thesis, 2001
2. Per Sandin, ”The Precautionary Principle: From Theory to Practice”, Licentiate thesis, 2002
3. Martin Peterson, ”Transformative Decision Rules. Foundations and Applications”, Doctoral thesis, 2003
4. Anders J. Persson, ”Ethical Problems in Work and Working Environment Contexts”, Licentiate thesis, 2004
5. Per Sandin, ”Better Safe than Sorry: Applying Philosophical Methods to the Debate on Risk and the Precautionary Principle”, Doctoral thesis, 2004
6. Barbro Björkman, ”Ethical Aspects of Owning Human Biological Material”, Licentiate thesis, 2005
7. Eva Hedfors, ”The Reading of Ludwik Fleck. Sources and Context”, Licentiate thesis, 2005
8. Rikard Levin ”Uncertainty in Risk Assessment - Contents and Modes of Communication”, Licentiate thesis, 2005
9. Elin Palm, ”Ethical Aspects of Workplace Surveillance”, Licentiate thesis, 2005
10. Jessica Nihlén Fahlquist, ”Moral Responsibility in Traffic Safety and Public Health”, Licentiate thesis, 2005
11. Karin Edvardsson, ”How to Set Rational Environmental Goals: Theory and Applications”, Licentiate thesis, 2006
12. Niklas Möller, ”Safety and decision-making”, Licentiate thesis, 2006
13. Per Wikman Svahn, ”Ethical Aspects of Radiation Protection”, Licentiate thesis, 2006
14. Hélène Hermansson, ”Ethical Aspects of Risk Management”, Licentiate thesis, 2006
15. Madeleine Hayenhjelm, ”Trust, Risk and Vulnerability”, Licentiate thesis, 2006
16. Holger Rosencrantz, ”Goal-setting and Goal-achieving in Transport Policy”, Licentiate thesis, 2006
17. Kalle Grill, ”Anti-paternalism”, Licentiate thesis, 2006
18. Jonas Clausen Mork, ”Is it Safe? Safety Factor Reasoning in Policy Making under Uncertainty”, Licentiate thesis, 2006
19. Anders J. Persson, ”Workplace Ethics: Some Practical and Foundational Problems”, Doctoral thesis, 2006

20. Eva Hedfors, ”Reading Fleck. Questions on Philosophy and Science”, Doctoral thesis, 2006
21. Nicolas Espinoza, ”Incomparable Risks, Values and Preferences”, Licentiate thesis, 2006
22. Mikael Dubois, ”Prevention and Social Insurance - Conceptual and Ethical Aspects”, Licentiate thesis, 2007
23. Birgitte Wandall, ”Influences on toxicological risk assessments”, Licentiate thesis, 2007
24. Madeleine Hayenhjelm, ”Trusting and Taking Risks: A Philosophical Inquiry”, Doctoral thesis, 2007
25. Hélène Hermansson, ”Rights at Risk - Ethical Issues in Risk Management”, Doctoral thesis, 2007
26. Elin Palm, ”The Ethics of Workplace Surveillance”, Doctoral thesis, 2008
27. Jessica Nihlén Fahlquist, ”Moral Responsibility and the Ethics of Traffic Safety”, Doctoral thesis, 2008
28. Barbro Björkman, ”Virtue Ethics, Bioethics, and the Ownership of Biological Material”, Doctoral thesis, 2008
29. Karin Edvardsson Björnberg, ”Rational Goal-setting in Environmental Policy: Foundations and Applications”, Doctoral thesis, 2008
30. Holger Rosencrantz, ”Goal-setting and the Logic of Transport Policy Decisions”, Doctoral thesis, 2009
31. Kalle Grill, “Anti-paternalism and Public Health Policy”, Doctoral thesis, 2009