

Author's version before copy-edit. Published version available at:
http://link.springer.com/chapter/10.1007/978-94-007-6374-6_3

Normative and non-normative concepts: Paternalism and libertarian paternalism

Kalle Grill

Department of Philosophy, Uppsala University, Sweden

Abstract This chapter concerns the normativity of the concepts of paternalism and libertarian paternalism. The first concept is central to evaluation of public health policy, but its meaning is controversial. The second concept is equally controversial and has received much attention recently. It may or may not shape the future evaluation of public health policy. In order to facilitate honest and fruitful debate, I consider three conceptual approaches to these concepts, in terms of their normativity. Concepts, I claim, may be non-normative, normatively charged, or normative in that they involve more complex relationships between values or duties. While the last approach is often best, other approaches may be appropriate depending on the context and purpose of discussion. The chapter's conceptual investigation is illustrated in application to two public health policies – a tax on consumption of fat and the encouragement of health-promoting food displays in restaurants and supermarkets.

Introduction

Few concepts are as fundamental to the evaluation of public health policy as the concept of paternalism. The modern welfare state as such may be seen as an infringement on individual liberty from a libertarian perspective. Even the social liberal, however, who sees no problem in taming market forces for the good of all, typically finds much more problematic the promotion of population health through regulation of private life-style choices. Against this background, the promise of libertarian paternalism to greatly improve people's health by making their everyday decisions more healthy, without affecting their liberty, is rather spectacular.

No moral or political argument is defended in this essay. Instead, the focus is on the concepts of paternalism and libertarian paternalism and how they can be understood in ways which are fruitful for moral and political debate, whatever one's normative commitments. In particular, the focus is on the normativity of

these concepts – i.e. whether or not they are in themselves normative or non-normative (descriptive), and in what sense.

Moral and political debate typically takes place against a conceptual background that is taken for granted. Even technical terms such as paternalism and libertarian paternalism are often simply assumed to be clearly delimited and well understood. The truth is that these terms are used with different meanings in different contexts by different people. Anyone involved in evaluation of public health policy, whether to defend it to the public or question it in parliament or discuss the details of its proper implementation within some government agency or organization, would do well to consider what the terms paternalism and libertarian paternalism may most fruitfully mean, in order to better understand other people's arguments and in order to better formulate one's own. By "better" here, I do not mean strategically efficient in the short term, but theoretically sound and so useful in promoting mutual understanding and honest debate.

I will first provide a short background to the concepts of paternalism and libertarian paternalism as they have developed in the scholarly debate. I will then give an even shorter background on what I mean by normative concepts in this context. Next follows a more detailed discussion of the debate on the proper definition of paternalism, especially in terms of the normativity of the concept. This discussion, important in its own right, will then be used to shed light on the more recent concept of libertarian paternalism. Before concluding, I will provide two illustrations to further clarify the different conceptual approaches surveyed.

Background

Much of contemporary philosophical debate on paternalism takes as its starting point John Stuart Mill's *On Liberty*. Mill did not himself use the term paternalism, but famously formulated the principle 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." (Chapter I, 9th paragraph) This principle has later become known as the harm principle.¹ Interest in this aspect of Mill's philosophy was reignited by the debate between H.L.A. Hart and Patrick Devlin over the relationship between law and morality and the proposal to decriminalize private homosexuality in the UK in the late 1950s. Hart (1963) made a point of distinguishing legal moralism – "to enforce positive morality" – from paternalism – "to protect individuals against themselves" (p. 31). Hart rejected moralism (and so Devlin's rationale for keeping homosexuality criminalized), but accepted paternal-

¹ To be precise, the harm principle sometimes denotes the principle that we may interfere with liberty (or exercise power over someone against her will) in order to prevent harm to others, and sometimes the more ambitious principle that this is the only acceptable reason to interfere with liberty.

ism. His main argument for paternalism was that "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" (p. 33). As a result, Hart argued, people make choices that harm them, and when they do we should sometimes stop them for their own good.

Our understanding of the many ways in which we are products of our environments have only deepened since Hart's times. It is no coincidence that Joel Feinberg, Mill's most distinguished follower in the 20th century, made voluntariness a central concept in his critical treatment of paternalism. Feinberg was, in most of his writings, a fierce anti-paternalist, arguing that preventing harm to a person is normally not even a good reason for limiting her liberty, at least not by law (1971, 1986). However, Feinberg supported soft paternalism – the principle that the state may limit a person's liberty for her own good when her conduct is insufficiently voluntary, or if intervention is needed to establish how voluntary it is (1986, p. 12). Feinberg went to great lengths to attempt to explain how voluntary is voluntary enough. His 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. (1986, p. 12)

Other authors have picked up and developed the idea of soft paternalism (e.g. Hodson 1977; VanDeVeer 1986) and it is now generally thought that such paternalism is compatible with a liberal normative outlook. Several authors argue that even hard paternalism – i.e. benevolent interference with voluntary choice, can be justified on a liberal basis if the motive is to secure the person's long term liberty or autonomy (Dworkin 1972; Kleinig 1983; Sneddon 2001; Husak 2003), though with this Feinberg would certainly disagree, and probably Mill as well (though he famously rejected involuntary slavery as incompatible with the very idea of liberty). Feinberg was ambivalent about calling his position soft paternalism, stating that he would have preferred "soft anti-paternalism" if the competing term was not already well established (1986, p. 15-16).

Libertarian paternalism, introduced in a 2003 article by Richard Thaler and Cass Sunstein and discussed more widely since the publication of their book-length defence in 2008 (new edition 2009), resembles soft paternalism in some ways. However, while soft paternalism is based on observations of well known obstacles to voluntary choice such as compulsion, lack of information and extreme emotions, libertarian paternalism is based on observations of more subtle psychological mechanisms. A large and growing body of research in psychology and behavioural economics shows that the decisions people make are heavily influenced by such seemingly irrelevant factors as what is the default alternative, how the alternatives are described (framing), and even by obviously arbitrary comparison

points (anchoring). People also display apparently irrational preferences and beliefs such as a general preference for the status quo, aversion to losses relative to gains, unrealistic optimism, overconfidence, and regularly mistaking random patterns for non-random ones (for an overview of this research, see Thaler and Sunstein 2009).

In the spirit of soft paternalism, we might conclude that to the extent that decisions are affected by these psychological mechanisms, they are not truly our own and so can be overridden without interference with our liberty. However, this is not the conclusion drawn by the advocates of libertarian paternalism. Instead, they propose that since these psychological mechanisms affect what choices are made, we may as well make sure that they tend to promote physical, economical and mental well-being. This can be achieved by active design of the choice situation. The proponents argue that this does not limit our freedom to choose, since no alternative is blocked or removed and no one is forced to change her preferences. In sum, while based on similar observations about the deficiencies of human decision-making, soft and libertarian paternalism have quite different implications. Soft paternalism allows outright interference with insufficiently voluntary choice (including physical coercion and legal prohibitions). Libertarian paternalism allows only active design of choice situations, which is allegedly not interfering at all.

The connection between traditional anti-paternalism, soft paternalism and libertarian paternalism can also be highlighted by noting what view of the decision-maker and her preferences is linked with each approach. Mill mentions in passing that his principle is only intended to apply to people "in the maturity of their faculties" (chapter I, 10th paragraph) or of "the ordinary amount of understanding" (Chapter IV, 3^d paragraph), but in the main disregarded the frailties of the human psyche, assuming that adults normally act on their settled preferences. Feinberg, in contrast, had a deep understanding of many of the ways in which we are affected by choice situations. However, he assumed in a Millian fashion that generally competent people have settled preferences which direct them to the extent that they are free from voluntariness-reducing circumstances. The point that Thaler and Sunstein have driven home is that because of the great impact of such things as default rules, starting points and framing effects, most of the time we simply have no preferences independently of the choice situations we face.

Normative concepts

There are many ways in which a concept can be normative. I propose that calling a concept normative most often means either of two things: First, a concept can be *normatively relevant* in the sense that whether or not something falls under this concept matters, to some degree, for what we ought to do or how things ought to be. This does not require that the concept is defined in normative terms. Whether

or not a concept is normatively relevant is context-dependent. For example, what gender someone belongs to is often not normatively relevant but it is relevant when accepting applications for an exclusive mentoring program for women (perhaps set up to address structural inequality between the genders). Because of this context-dependence, it makes no sense to say quite generally that a concept is normatively relevant. I will, however, use the idea of *non-normative* concepts, meaning concepts that are almost never normatively relevant, though since there will always be at least hypothetical exceptions, the lack of normativity is really a matter of degree.

Second, a concept can be *normatively charged* in the sense that whether or not something falls under the concept always counts for or always counts against doing that thing or being that way, if it counts at all. For example, the concept of harming may be negatively normatively charged in the sense that if something is harming, that always counts against doing that thing or having things that way. The qualification “if it counts at all” is needed because on many normative theories some concerns have priority and so make other concerns irrelevant, even though these other concerns would typically be relevant. For example, on some moral theories, oath-breaking is absolutely forbidden. On such theories, if an action both harms someone and means breaking an oath, then the action is forbidden just because it is an oath-breaking; the harming is irrelevant. Even on such theories, however, if harming is negatively charged it can never turn positive in the sense that it counts in favour of an action that it harms someone. In contrast, a concept that is merely normatively relevant may matter in different ways depending on the context. For example, that a person has a certain gender may obviously count either for or against treating her in some way, or be normatively neutral.

A concept’s normativity is one of its properties. Establishing that a concept is or is not non-normative or normatively charged is to restrict its domain along one dimension. It is not to give a precise definition, but it is to delimit what definitions may be correct. For an analogy, establishing that a chair is a functional object restricts what definitions may be correct: Correct definitions cannot cover chair-like objects constructed for display only. However, what is the correct precise definition of the concept chair is still an unsettled matter. Similarly, establishing that paternalism is a negatively charged concept is not to give a precise definition, but it is to establish that correct definitions cannot cover behaviour that is morally unproblematic.

In the following, I will investigate paternalism and libertarian paternalism as properties of actions or policies. Both concepts could alternatively be understood as moral or political principles. Few people call themselves, or even others, “paternalists”, unless in a pejorative sense in the public debate. Paternalism is something you may accept, but it is not a political philosophy. Libertarian paternalism, in contrast, is to some extent a political principle, that I have already compared to the moral-political principles of anti-paternalism and soft paternalism. However, we may also say that a policy is libertarian paternalist, not in the sense that certain people endorse it, but in the sense that it has certain properties in and

of itself. It is in this non-principle sense that I will treat libertarian paternalism in the following. Policies are libertarian paternalist policies in virtue of their content or aims, people or political principles may perhaps be libertarian paternalists in a derived sense to the extent that they endorse such policies.

While I talk throughout of normative and non-normative concepts rather than moral and non-moral, I will focus on normativity in the sense of moral relevance and moral charge – i.e. relevance and charge related to moral oughts.

The concept of paternalism

For and against a non-normative approach

In the contemporary debate on the adequate definition of paternalism, authors disagree on whether or not the concept is normative, and in what sense. At one extreme, Donald VanDeVeer (1986) considers it an advantage of his definition that paternalistic behaviour need not be morally problematic in any way (e.g. p. 35). His definition is, abbreviated, that behaviour is paternalistic if it aims to benefit a person contrary to her current preference, intention or disposition (p. 22). To exemplify, people may prefer not to be told that their smoking is harmful and so providing this information for their benefit is paternalistic. However, if the information is provided in a non-intrusive and neutral way, one may reasonably claim that such provision is not morally problematic. There are many cases in which paternalistic behaviour does raise moral issues, but, VanDeVeer argues, this is because it involves some presumptive wrong over and above the paternalism.

VanDeVeer presents this argument in favour of his approach: "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 claim, however, is too strong. Even if paternalism is morally wrong by definition, we can still argue rationally about whether or not anything is paternalistic in this sense. The labelling can, in other words, involve much more than a simple assumption of wrongness. VanDeVeer points out 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'. Even if, by definition, justice is good and betrayal is bad, this does not exhaust the discussion on these important concepts, nor does it necessarily beg any moral questions.

Gerald Dworkin (2008) takes a less categorical stand in favour of a non-normative approach, stating in his *Stanford Encyclopedia of Philosophy* entry on

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." Dworkin presents no further argument for this position. I believe the standard rationale for preferring non-normative definitions is a belief that it makes normative discussion more tidy and so more fruitful. If we can first, the thinking goes, agree on which things are which, we will then be in a better position to clarify our normative disagreements and perhaps even resolve them. Though this sounds reasonable, I propose that it is somewhat naive and not well suited to the issue at hand.

Any definition of a moral-political concept is bound to be controversial. This goes for normative as well as non-normative definitions. Much discussion on the concept of paternalism appeals to linguistic intuitions. For example, in Bernard Gert and Charles Culver's much quoted (1976) article on the definition of paternalism, the authors list a number of cases that they take to be obvious examples of paternalism, presumably in force of the linguistic reactions we have to those cases, simply 'feeling' or 'seeing' that they must involve paternalism, quite independently of the clearly normative definition that the authors then present to accommodate the listed cases. Needless to say, these cases have led to much discussion and there is no sign that even proponents of non-normative definitions are about to agree on one. The preparatory establishment of a non-normative definition is never completed and so stands in the way of normative discussion rather than facilitating it.

More to the point, the concept of paternalism is especially ill suited to non-normative definitions. We are interested in this concept precisely because of its alleged normative properties. Authors like Mill and Feinberg denounce paternalism more or less categorically. It would be convenient if we could find a non-normative description of the focus of their concern, but since this is embedded in values like liberty, autonomy, individuality and sovereignty, such a description would be hard to come by even under more favourable epistemic circumstances.

Our linguistic intuitions may tempt us to regard some public health policies as paternalistic even though they are morally unproblematic. This fact seems to point in the direction of a non-normative concept. For example, the smoking information policy mentioned in connection with VanDeVeer's definition above may seem paternalistic. Or, for a more private example, borrowed from Dworkin (2008), it may seem paternalistic for a dying person to bequeath money to a benefactor for use in a third person's best interest, instead of bequeathing it directly to that person. However, linguistic intuitions are neither consistently shared, nor even consistent within each person's linguistic universe. I would be prepared, for example, to count both these instances as non-paternalistic, though perhaps paternalism-like in certain respects.

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 moral and political discussions of paternalism, however, we are mainly interested in benevolent interfer-

ence, not in material provision or other areas of parenthood. Paternalism is a term of art and we should define it to serve our purposes rather than our intuitions. These purposes are, I propose, mainly to understand and critically evaluate the traditional liberal resistance to some forms of benevolent involvement in other people's lives.

Like most authors, both Dworkin and VanDeVeer are mainly interested in the normative aspects of paternalism. While VanDeVeer strongly prefers a non-normative conceptual approach, he devotes most of his investigation to those cases of paternalism that involve a "presumptive wrong". Dworkin leaves it an open question whether or not his own analysis is normative, and proposes that we "decide upon an analysis based on a hypothesis of what will be most useful for thinking about a particular range of problems". I propose that almost all such analyses will be normative in some sense, though there may be exceptions.

For and against a normatively charged approach

In the political debate, the term paternalism is often used in a pejorative sense - paternalism is assumed to be unacceptable, without further argument. It is perhaps this form of normative approach that inspires VanDeVeer's critique. In the philosophical and bioethical debate, a common approach is to assume, somewhat more cautiously, that paternalism is morally wrong unless certain conditions are fulfilled (e.g. Dworkin 1972, Kleinig 1984). In public health ethics, a similar assumption underlies several influential treatments of the justifiability of coercive public health measures, which are typically paternalistic, though not always described using that term (e.g. Kass 2001, Upshur 2002, Childress et al., 2002).

One way to understand this widespread assumption is in terms of the conceptual assumption that paternalism is negatively normatively charged. This means that the fact that, for example, a public health policy is paternalistic always counts against that policy (and this is part of what it means for the policy to be paternalistic). This is consistent with the possibility that some other property of the policy, such as its great benefits in terms of population health, outweighs the negative charge.

Seana Shiffrin (2000) has more explicitly proposed what seems like a normatively charged definition of paternalism, arguing that "paternalist doctrines and policies convey a special, generally impermissible, insult to autonomous agents" (p. 208). One of the conditions in Shiffrin's definition is that paternalistic behaviour towards a person B is "directed at B's own interests or matters that legitimately lie within B's control" (p. 218). Despite the disjunctive formulation, B's interests are supposed to be among the things that B should have control over. Therefore, paternalism always involves controlling what someone else should control, which seems morally wrong. Shiffrin does not, of course, claim that paternalism is always morally wrong all things considered. Parts of her argument indi-

cates, however, that the fact that some behaviour is paternalistic always counts against it, morally speaking (that paternalism is, as she says, “pro tanto morally problematic”) (p. 220, footnote 25).

The normatively charged approach to paternalism rests on a strong commitment to autonomy or some such liberal value. However, since paternalism is not merely an infringement of autonomy, if it even necessarily is such an infringement, this background value is not sufficient to explain the negative charge. In fact, the normatively charged approach is seldom explicitly defended (Shiffrin is an exception).

One advantage of the normatively charged approach is that it directs our attention towards the moral controversy involved in paternalism. However, as it stands this approach means that people who disagree on the moral status of paternalism will disagree on the nature and even existence of paternalism. Those who find quite acceptable what Shiffrin and others call paternalism will have to claim that there exists no paternalism. This may often be conceptually impractical or confusing, especially when the assumption is implicit. It is also conceptually radical, since it implies that the substantial amount of authors who think that they think that paternalism is acceptable must be mistaken.

Sometimes, paternalism is discussed as if Mill’s views, or some other set of liberal views, defined the very concept. This suggests that the normatively charged approach could be relativized in the sense that we take an essential property of paternalism to be that it is *held to be* morally wrong by some relevant class of people. This approach would have the appealing consequence that talking about paternalism simply is talking about certain positions, which we may then endorse or dispute. As this is just what we normally do when we debate normative issues, there would be no separate question of the nature of paternalism. Notwithstanding this advantage, the approach is probably too subjective. In referring to Mill’s views, we typically suppose that Mill successfully identified something – paternalism – which can be defined independently of his discovery or description. To allow the description to define what is described is to put the cart before the horse and implies for example that if Mill was inconsistent in his views, the concept itself is inconsistent.

While both the objective and the subjective version of the normatively charged approach points our attention in the direction of an (alleged) moral wrong, they say very little about this wrong. Full understanding of the concept of paternalism, on this conceptual approach, need not reveal anything about *why* it is (held to be) wrong, nor why many people find paternalism quite acceptable in many cases. The third approach I will consider is an attempt to provide more normative substance, without too radical and impractical conceptual consequences.

A value-focused approach

Shiffrin is unusually thorough in explaining the particular moral wrong involved in paternalism, in terms of the insult it delivers to a person's autonomous judgment and agency. Such explication of values is at the heart of an alternative approach that I will call the *value-focused* conceptual approach.

Unlike typical candidates for negatively normatively charged concepts, like cruelty, deception and unfairness, paternalism is almost universally thought to involve, essentially, the promotion of some good, or at least the honest intention to promote some good. Shiffrin's definition is a notable exception to this norm. She is led to the normatively charged approach by her unorthodox position that paternalism need involve no benevolence. On some other, libertarian-leaning accounts, the aim for good is thought of as a bad excuse, or as a further insult on top of the interference, with the assumption that no real good will come of it. Such understandings too fit nicely with a normatively charged definition. However, the most interesting ethical problems in this area arguably arise when the aim to benefit is effective or at least likely to be effective. From a policy perspective, policies that interfere with people's liberty and do no good are obviously to be avoided, while policies that interfere with people's liberty and promote their health or other interests raise difficult questions.

Several contemporary definitions of paternalism try to capture the conflict between interference and benevolence. For example, David Archard (1990) proposes that the core of paternalism is "the usurpation of one person's choice of their own good by another". According to Archard, this usurpation is not always even pro tanto morally wrong. It is rather that "we have duties to promote the well-being of others, but we also have a duty – at least on the standard anti-paternalist account – to respect the choices of others." (p. 41-42) In other words, paternalism involves a conflict between two duties and the justifiability of paternalistic action and policy depends on how those two duties are balanced against each other.

This points to a middle way between non-normative and normatively charged definitions. The concept of paternalism may essentially involve two conflicting duties or values, where one has to do with respect for liberty or autonomy, and the other has to do with promotion or protection of things like health, happiness, security, financial wellbeing, or even achievement, self-fulfilment and other aspects of the good life quite generally.² Indeed, the good can even include long term liberty and long term autonomy, which are threatened by the agent's free and autonomous choice or action in the present. When this conflict is resolved in favour of overall good, or wellbeing, at the expense of present liberty, we have an

² In the following, I will for simplicity talk only of values, and not of duties, with the assumption that "value" may refer either to good things as such (which should be promoted according to various moral theories incorporating a consequentialist element), or to duties relative to these good things (figuring in deontological moral theories).

instance of paternalism. To denounce paternalism is to hold that such conflicts should never be resolved in that way. The concept as such, however, carries neither a positive nor a negative charge.

The value-focused approach clearly distinguishes paternalism from other normative concepts. Various proposed definitions can now be interpreted as attempts to capture more precisely the nature of the values which conflict, as well as the way in which they conflict. Two values conflict in an objective sense when they cannot both be realized or fulfilled. In this objective sense, paternalism is the promotion of good at the expense of liberty, regardless of anyone's thoughts, intentions or feelings. Most authors hold that paternalism must involve a motive or intention to promote good at the expense of liberty. This means the conflict is more subjective or intentional. Arguably, evaluation of laws and policies should more often be focused on objective effects, while evaluation of behaviour may perhaps be focused on motives and intentions. The value-focused approach can accommodate both of these varieties (as can the normatively charged approach).

For a very subjective take on the conflict of values, we could say that it is paternalism if someone attempts to promote what she thinks is a person's good, by limiting that person's freedom, even if she does not in fact do any good. It could also be paternalism if, conversely, someone promotes a person's good by means which she thinks limits the person's freedom, even if it does not in fact do so. Somewhat surprisingly, it is almost standard practice to define paternalism along these lines. Both VanDeVeer's and Archard's explicit definitions say that paternalism involves interference only in the sense that the agent believes that she will act contrary to the person's preference (VanDeVeer) or diminish her choice (Archard), regardless of whether or not she actually does so (another influential example is Gert and Culver 1976). However, I believe that this is a weakness of these definitions, which should be corrected. Some of the surrounding discussion suggests that VanDeVeer and Archard think that paternalistic behaviour must at least aim to do something which is in fact interfering (as opposed to something that is completely innocuous but that the agent in her possibly confused state of mind dreams up to be interfering). However, I propose that even such slightly less subjective paternalism, when completely ineffective, is irrelevant for the evaluation of public policy.

If anyone should hold that liberty is not a value, she must, if the value-focused approach is accepted, claim that there exists no paternalism. This might seem to lead to the same radical and impractical terminology that I said would follow from the normatively charged approach objectively understood. However, most people do find liberty valuable, even though they may find other values more important.³ Those who insist that liberty is not even a rather trivial value are certainly fewer than those who insist that paternalism is sometimes morally right. The

³ I assume that utilitarians can hold, with Mill, that liberty is a value, instrumental to utility or an aspect of utility.

value-based approach, therefore, will not confuse or limit our normative language to the same extent as the objective normatively charged approach.

It is possible to hold that wellbeing has no value, I suppose, though this seems very exotic. Authors who are opposed to government benevolence typically believe either that such benevolence does no good, or that what good it does is outweighed by accompanying bads. Of course, the first group of authors often want to label alleged but ineffective wellbeing-promoting government activities “paternalism”. This is not possible on the value-focused approach, since such activities do not involve any positive value. In this sense, the value-focused approach is radical – policies that are trivially undesirable because they have no positive effects are not paternalistic. This is an immediate and unavoidable consequence of delimiting the concept to what is arguably the normative core of paternalism – the conflict between on the one hand respect for liberty and autonomy and on the other hand the protection and promotion of wellbeing. There can be such a conflict only if there is something for liberty to be in conflict with. I propose that limiting the use of “paternalism” in this way does not cause much conceptual confusion, but merely deprives critics of misconceived government policy of a theoretically-sounding pejorative term.

The concept of libertarian paternalism

The concept of paternalism has been formed over a long period of time as part of the process of understanding the conflict between individual liberty and benevolence, and the traditional liberal insistence that the former take priority. In contrast, the term libertarian paternalism was introduced in an academic article in 2003. The authors, Richard Thaler and Cass Sunstein, present libertarian paternalism both as a political position or principle and as a form of policy strategy. The strategy, as noted above, is focused on exploiting predictable human biases and other irrational tendencies to design choice situations in ways that promote wise choices. This strategy seems both practically useful (as indicated by the interest at the highest levels of government both in the US and in the UK) and philosophically interesting (as indicated by the academic interest among legal scholars, economists and, more recently, philosophers). Since Thaler and Sunstein do little to specify the concept of libertarian paternalism, and since many of their examples seem to contradict what they do say in the way of specification, the concept is rather open to interpretation. This is of course potentially problematic, as people debate the strategy with no clear view of what it is. As possible remedies, I will now consider how the three conceptual strategies that have been proposed for understanding paternalism may be applied to libertarian paternalism.

The non-normative approach

Like paternalism, libertarian paternalism is very much a political-philosophical concept, to be used in normative, political debate. Just as with paternalism, however, we may consider whether the concept as such is normative or non-normative. According to Thaler and Sunstein, 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." (Sunstein and Thaler 2003, p. 1162; cf Thaler and Sunstein 2009 p. 5) It is libertarian in the sense that it aims to ensure that "people should be free to opt out of specified arrangements if they choose to do so." (2003, p. 1161; cf 2009 p. 5) The possibility to opt out is said to "preserve freedom of choice" (2003, p. 1182) – i.e. to be "liberty-preserving" (2009, p. 5). This can be understood non-normatively: That an action or policy aims at improvement and ensures opt-out possibilities need not be either morally problematic or morally benign.

I argued above that a non-normative approach is ill suited to the concept of paternalism, because this concept is strongly tied to traditional liberal denunciation. Something of the opposite is true for libertarian paternalism. The strategy is put forth as a non-intrusive and therefore unproblematic form of benevolent policy-making. I will discuss this claim and some critique below in considering a value-focused approach. However, I propose that what makes libertarian paternalism interesting is to a large extent the novel strategies proposed for influencing choice by non-rational means, that is by exploiting (in a non-pejorative sense) people's biases, misconceptions and predictable, irrational responses to irrelevant aspects of choice situations. Luc Bovens (2009, p. 209) similarly proposes as one defining characteristic of libertarian paternalism that "some pattern of irrationality is being exploited"⁴. Thaler and Sunstein often have a wider focus than this, recommending all sorts of policies that will improve choices, including providing relevant information (which for Mill and others is paradigmatically non-paternalistic). However, it is the non-rational influencing that is more novel and typical of the strategy. A possible non-normative understanding of libertarian paternalism, therefore, is 'benevolent non-rational influence', or perhaps, more narrowly, benevolent non-rational influence that does not make any alternative inaccessible or too difficult to pick – i.e. 'mild benevolent non-rational influence'.⁵

The meaning of "benevolent" here is ambiguous. We know from the debate on the definition of paternalism that it is notoriously difficult to determine

⁴ Bovens mostly talks of "nudge" rather than "libertarian paternalism" but means the same thing by these terms.

⁵ This is perhaps still too wide, since it would seem to include subliminal advertising, which Thaler and Sunstein reject (2009, p. 247). In rejecting subliminal influence, Thaler and Sunstein refers to what they call a publicity condition, which would exclude "invisible" manipulation. Perhaps some such transparency condition should complement the non-normative specification.

even in the abstract whether a policy aims to benefit only the people directly targeted, or also others, and if so to what degree. Libertarian paternalism could either inherit this difficulty, or the restriction to the chooser's interests could simply be dropped, allowing that the benevolence could be directed at anybody, the chooser and/or others. Though Thaler and Sunstein say the strategy aims at making choosers themselves better off, some of their policy proposals (such as presumed consent to organ donation (2009, p. 179)) are aimed exclusively at benefiting other people than the targeted choosers. One reason to keep the distinction, however, is that some authors find it more morally problematic to influence choice for the sake of the chooser. This is not to allow this alleged moral problem to define the concept, but simply to delimit the concept non-normatively in order to simplify normative debate. If we go ahead and drop the restriction to the chooser's interests, of course, the term "paternalism" is arguably less appropriate (though as noted Shiffrin's definition does not include benevolence).

What is the point of a non-normative concept of libertarian paternalism? It may be to point out certain public health strategies that are salient because they have not been properly recognized and so not properly utilized, or because they are likely to be politically feasible, in contrast to other strategies. Whether or not such strategies are morally acceptable is, on this approach, another matter.

The value-focused approach

While one virtue of Thaler and Sunstein's writings is their detailed treatment of concrete techniques for constructing choice situations (nudging techniques), they clearly focus on these techniques because of their potential to promote wellbeing (including health) without limiting freedom of choice. The claim that libertarian paternalism does not limit freedom of choice is what makes it so politically attractive. This claim has been challenged in several ways. However, a possible conceptual approach to libertarian paternalism is to take the claim seriously and accept that policies that do limit freedom of choice are not in fact libertarian paternalist. Likewise, we may accept the claim that libertarian paternalism promotes wellbeing, or at least that it aims to do so.⁶

These assumptions amount to a form of value-focused understanding of the concept, though not in this case in terms of a conflict. Libertarian paternalism is, on this conceptual approach, essentially the combination of a promotion of good (by avoiding harm or providing benefit) and the absence of an interference with freedom or liberty (at least in terms of choice). Just as with paternalism, these two sides may appear on different levels of intentionality – they may be intentions

⁶ Thaler and Sunstein defend no particular idea of what wellbeing consists in, but rather refer to people's considered judgment, assuming that people typically prefer to live long and be healthy.

only, effects only, or effective intentions. It is libertarian paternalism in a motive sense to aim to promote wellbeing without limiting liberty. It is libertarian paternalism in an effect sense to actually do so. The approach can accommodate these varieties, though as with paternalism, motives without effects are hardly relevant for the evaluation of public policy.

In the case of paternalism, the value-focused approach in and of itself distinguishes the concept from other concepts. The value-focused approach to libertarian paternalism, in contrast, includes all forms of non-interfering benevolence, including providing help upon request. This category of actions is better described simply as non-interfering or non-paternalistic benevolence. Even if it sometimes seems that Thaler and Sunstein are in fact talking about this wide category, it would be pointless to attach a new term to this existing concept, except for rhetorical reasons. If libertarian paternalism is to be an original concept, it needs some additional non-normative content.

One such possibility is to restrict the concept to non-rational influence, thereby picking out a subclass of non-interfering benevolence. The point of this might be similar to what could motivate a purely non-normative approach, only now coupled with the specification that the non-rational influence in question is liberty-preserving. This is to make a normative claim and to back this claim up with a non-normative specification. This is arguably what Thaler and Sunstein intend to do to the extent that they focus on non-rational influence. This approach is controversial, raising the question whether or not any policy can be libertarian paternalist in this sense.

I have called this approach value-focused because it is focused on particular values – wellbeing and liberty, and a particular relationship between them. However, if this relationship means that one positive value is promoted, with no ensuing disvalue, then the concept is at the same time positively normatively charged, since being libertarian paternalist always counts in favour of a policy.

With the hybrid approach, as with the non-normative approach, the question arises of whether or not to keep the restriction to promoting the chooser's good in particular. If this restriction is dropped, the term "paternalism" is even less appropriate on this approach than on the non-normative approach, since there is by definition no limitation of liberty, which is an essential property of paternalism on most accounts. On the other hand, the term may be motivated because it refers to influencing techniques that are typically liberty-limiting, though not in this case, or that are relevantly similar to techniques which are liberty-limiting.

An alternative to the hybrid approach is to look to more specific or narrow liberal values that may be diminished by libertarian paternalist policies even if these do not restrict freedom in the sense of blocking out alternatives or making them prohibitively costly. Because libertarian paternalist policies do not engage the rational processes of their targets, but rather exploit shortcomings of rationality, they may be considered manipulative and/or an aggressive substitution of the chooser's judgment for that of the policy-maker. For example, Mark White (2010) argues that intentionally structuring choice situations in order to promote the pre-

sumed interests of choosers is manipulative in a way that neither structuring choice situations randomly or structuring them to further one's own self-interest is. Bovens (2009) suggests that libertarian paternalism may prevent the building of moral character and decrease people's sense of personal responsibility (pp. 214-15). If paternalism essentially involves some such disvalue, a value-focused approach that more closely resembles the conflict of values characteristic of paternalism proper would be appropriate. However, since critics and proponents disagree on the merits of these charges, and since this is a main division in the political debate on libertarian paternalism, this strategy is probably too confusing.

Some critics, such as White, argue not only that libertarian paternalism has its costs, but, much more strongly, that it has no benefits. Governments can do nothing to improve the wellbeing of their populations through libertarian paternalist policies, these critics claim. White, for example, believes that there is no way to know what a person prefers unless she has explicitly stated her preference or acted on it, and that there is no way to benefit a person unless we know what she prefers. Since libertarian paternalism does not even aim to satisfy such explicit preferences, it can yield no benefits. Proponents are mistaken, White seems to say, in simply assuming that most people are better off alive and in good health. This view is obviously rather extreme. My point is only that if benevolence is not desirable, for whatever reason, and manipulation is always undesirable to some extent, then libertarian paternalism is morally wrong and so the concept is negatively normatively charged.

Since libertarian paternalism is a new concept and less integral to our moral and political world views than paternalism, a charged concept is less problematic than in the case of paternalism. If, for example, we decide that libertarian paternalism is morally unproblematic by definition, critics may deny that there is such a thing as libertarian paternalism without too much linguistic confusion.

In sum, a value-focused approach to libertarian paternalism can either combine non-interference and good-promotion, which requires some further non-normative content to form an original concept, or combine good-promotion and some controversial disvalue such as manipulation. Either strategy can lead to a normatively charged concept, if it entails the realization of a positive value only, or a disvalue only.

While Thaler and Sunstein have a tendency to include everything they like under the concept of libertarian paternalism, they seem, when they focus on the concept itself, to have in mind the hybrid positively charged approach. Libertarian paternalism is, on this approach, 'morally unproblematic mild benevolent non-rational influence', where at least "mild" may need some further specification.

A terminological note on libertarianisms

As noted above, Thaler and Sunstein consider the term libertarian appropriate because the strategy preserves freedom of choice. However, the strategy is aimed at improving welfare, an aim that libertarians typically think is inappropriate for the government. There are many other aims that could inform how choices are influenced by non-rational means. For example, Gregory Mitchell (2004) has suggested that a truly libertarian paternalism would permit influencing choice only in order to either improve decision-making competence by various debiasing techniques, or to prevent liberty-restricting irrational choices, but not to improve the chooser's general welfare (p. 20). Alternatively, one could aim to maximize autonomy or liberty in some sense, perhaps in the form of self-reliance, or to promote a libertarian life-style of non-interference with others. Any of these strategies would arguably better deserve the name libertarian paternalism.

As a terminological matter, therefore, the term libertarian paternalism is unfortunate. We must accept that our terminology is typically determined not by the desirability of theoretical clarity, but by contingent social and political events. Perhaps, however, it is not too late to change the term libertarian paternalism. Thaler and Sunstein put less emphasis on it in their book than in their earlier articles. In the book they talk more about nudging and choice architecture, which are stronger candidates for non-normative concepts (choice architecture may refer to all circumstances of a choice, or to those which have a non-rational influence; nudging is changing or shaping the choice architecture). This suggests the possibility of abandoning the very concept of libertarian paternalism and speak only of nudging and choice architecture. Some of the conceptual investigation in this chapter may be relevant for investigating these related concepts, but not all.

Illustrations

The conceptual approaches I have discussed can be applied to most moral and political concepts. Concepts can be non-normative, normatively charged or value-focused (where the two latter approaches partly overlap). Which approach one adopts naturally determines what one means when one uses the corresponding terms. I will now illustrate this by way of two examples, one for each concept.

Taxing fat in food – a paternalistic policy?

At the time of writing, it seems likely that Denmark will soon introduce a special tax on consumption of fat in food products. Similar measures are discussed in oth-

er countries. If someone claims that such a tax would be paternalistic, she may be giving a non-normative description, for example claiming in the spirit of VanDeVeer that the measure is aimed at providing benefits and preventing harms, but that it is contrary to people's preferences (VanDeVeer does not discuss group cases so the person would have to look elsewhere to find a reasonable specification of what she means by "people's"). Neither one of these properties necessarily makes the measure either good or bad. Though the benevolent aim is typically held to count in favour of the measure and the resistance against it, whether or not they actually do so depends on what is the correct moral and political theory. The categorization of the measure as paternalistic says nothing about this, nor about the appropriateness of the measure.

It is somewhat obscure why this person would make this statement, unless she thought that the mentioned properties of the tax were indeed morally relevant. She may of course wish to point out that the tax has these properties, which other people may find morally relevant, without any normative commitments of her own. However, there are certainly other properties that people may find morally relevant too, and so one is left with the suspicion that the person does find these particular properties especially relevant.

Now, another person may claim that the same fat tax is paternalistic, but mean by this, among other things, that the measure is morally wrong considering its immediate costs and benefits. This means that unless the tax is justified by factors unrelated to paternalism (such as raising funds necessary for the operation of the government), and unless this tax for some reason avoids a major catastrophe (larger than just saving thousands of anonymous life years by improving population health), then the tax is morally wrong all things considered. This second person's categorization of the measure as paternalistic implies a moral claim in addition to the conceptual assumption. We can therefore disagree with her in several distinct ways. If we accept her terminology but reject her moral claim we must say that the tax is not in fact paternalistic. Alternatively, we may object to both her terminology and her moral claim, insisting that though the tax is paternalistic, this does not mean that it is wrong. Less typically, we may accept her moral claim but reject her terminology, which leaves it an open question whether we agree that the measure is paternalistic, though we agree that it is wrong.

A third person may call the fat tax paternalistic in order to point out that it involves a cost in terms of liberty and a likely gain in terms of population health, giving rise to a value conflict. In other words, the categorization can help focus our attention on factors relevant for moral evaluation. If we accept this terminology but wish to deny that both of these values are involved in this particular case, we can object that the person is mistaken, because the tax is not paternalistic. If we reject the terminology, we may or may not agree that implementing it involves a conflict of values.

Needless to say, a debate between these three people over a proposed tax on fat in foods would likely be very confusing unless they recognized their differences in terminology. Sometimes one may accept whatever terminology one is

confronted with, in order to avoid conceptual disagreement and be able to focus on more important matters. However, this strategy can actually be confusing, since we expect normative and conceptual matters to be intertwined in moral and political debate. This strategy also presupposes that one is aware of what terminology one is confronted with, which is often opaque. Therefore, there may be no way around making explicit what one takes others to mean by some term, before agreeing to that use for present purposes, or declaring how one will use it instead. For most purposes, including the fat tax, I believe the value-focused approach of the third person is most conducive to fruitful moral and political debate.

Regulating display of food – a libertarian paternalist strategy?

While the term "paternalism" is used rather frequently in moral and political discussions, "libertarian paternalism" was recently introduced and it remains to be seen whether it will become an established concept or prove more temporary. More than in the case of paternalism, therefore, it is important to ask whether calling anything by this term provides a meaningful and useful categorization (though this may certainly be asked also in the case of paternalism proper).

In their book, Thaler and Sunstein introduce the libertarian paternalism strategy with the story of a (hypothetical) director of food services for a large city school system who realizes that the order and manner in which food is displayed in the cafeterias of her schools will have a large influence on what the students eat. The cafeteria case is also emphasized in both their 2003 articles (Sunstein and Thaler 2003, Thaler and Sunstein 2003) and is frequently considered a paradigmatic case of libertarian paternalism. It would not be difficult to instruct directors of public schools to implement healthy display plans, without changing what alternatives are offered. In fact, this policy could include all public workplaces where cafeterias are run by the organization (as opposed to independent entrepreneurs). Going more comprehensive, it would not be difficult to encourage healthy food display by private businesses. For example, the government could issue certificates for participating restaurants (and why not supermarkets) and this certificate could be publically marketed as a sign of health awareness and social responsibility (or certificates could entail tax exemptions though this would arguably be less libertarian).

What could be the point of calling this form of public health policy libertarian paternalism? First, it could be a non-normative claim, implying that the policy would amount to mild benevolent non-rational influence on people's food choices, or rather the encouragement of such influencing. Second, it could be a hybrid claim, implying that while the policy would amount to mild benevolent non-rational influence, this would not limit people's freedom to choose food as they like. If the expected health benefits are a positive and there are no other negatives directly associated with the policy, this is claiming that the policy is desira-

ble, and to be implemented. More generally, if all libertarian paternalism is similar in this respect, the concept as such is positively normatively charged. Third, the point of calling the healthy display policy libertarian paternalism could be to point out a value conflict between on the one hand promotion of population health, and on the other the ensuing disvalue of manipulation, or perhaps some other similar disvalue. If there is such a disvalue and if the promotion of population health (by the government) is inappropriate or impossible, then this is to claim that the policy is undesirable. More generally, all libertarian paternalism is undesirable and so the concept is negatively normatively charged.

I believe all three approaches may have good uses in moral and political debate. In order to avoid conceptual confusion, we should ideally have more than one concept. The first use may perhaps be called “nudging”, which sounds less normative than “libertarian paternalism”. The third could simply be called “paternalism”, or the name of some other value conflict depending on what value one takes to be compromised. The term “libertarian paternalism”, if it should be used at all, should be reserved for the second use, which is more novel and most in line with the intentions of its originators.

Conclusion

Modern anti-paternalists are typically what Joel Feinberg called soft paternalists, which means they denounce paternalism only when directed towards sufficiently voluntary behaviour, unless the aim is to make sure that such behaviour is indeed sufficiently voluntary. In the spirit of soft paternalism, the insight that our choices are strongly influenced by irrelevant details of choice situations could warrant benevolent interference with such choices. As a political principle, libertarian paternalism instead only sanctions actively designing choice situations so that they tend to produce wise choices, without restricting available alternatives.

I have argued that paternalism is a strongly normative concept, intimately linked with the traditional liberal denouncement of various types of benevolent involvement in other people’s lives, especially by the government. Since there are no inherent problems with normative concepts, it is best to accept a normative definition of paternalism. A normatively charged definition may have its advantages, but in general a value-focused definition will be more informative and less confusing. A precise value-focused definition will say something about in what sense a person’s liberty (or autonomy or some similar liberal value) is diminished by paternalism, and in what sense her other interests are promoted.

I have argued that there are several fruitful ways to understand the new concept of libertarian paternalism. These should be distinguished and preferably used under different names, though I do not expect convergence in either conceptual approaches or terminology. A non-normative definition could be fruitful because it points to a form of influence that has not been widely used in public

health work, at least not in its full width, but that could be very useful and could be acceptable because it is effective while relatively mild. A value-focused definition could be very useful because it captures the moral and political controversy the program has stirred. A hybrid positively charged definition is arguably what Thaler and Sunstein comes closest to endorsing. This may be fruitful in identifying useful public health measures which are morally unproblematic, if there are any such measures.

Acknowledgments Thanks to Stephen John, Verina Wild, Niklas Möller, Angus Dawson, the editors and an anonymous reviewer for helpful comments. The author is a COFAS Marie Curie Fellow with financial support from the European Commission and the Swedish Council for Working Life and Social Research.

References

- Archard, D. (1990). Paternalism defined. *Analysis* 50(1), 36-42.
- Bovens, L. The ethics of nudge. (2008). In T. Grüne-Yanoff and S.O. Hansson (Eds.) *Preference Change: Approaches from Philosophy, Economics and Psychology* (pp. 207–219). Berlin: Springer.
- Childress, J. F., Faden, R.R., Gaare, R.D., Gostin, L.O., Kahn, J., Bonnie, R.J., et al. (2002). Public health ethics: mapping the terrain. *The Journal of Law, Medicine & Ethics* 30(2), 170-178.
- Dworkin, G. (1972). Paternalism. *Monist* 56(1), 64-84.
- Dworkin, G. (2008). Paternalism. In E.N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition), <http://plato.stanford.edu/archives/fall2008/entries/paternalism/>. Accessed 1 June 2011.
- Feinberg, J. (1971). Legal Paternalism. *Canadian Journal of Philosophy* 1, 105 -124.
- Feinberg, J. (1986). *Harm to Self*. Oxford: Oxford University Press.
- Gert, B and Culver, C.M. (1976). Paternalistic Behavior. *Philosophy and Public Affairs* 6(1), 45-57.
- Hart, H.L.A. (1963). *Law, Liberty and Morality*. Oxford: Oxford University Press.
- Husak, D.N. (2003). Legal Paternalism. In H. LaFollette (Ed.), *The Oxford Handbook of Practical Ethics* (pp. 387–412). Oxford: Oxford University Press.
- Hodson, J.D. (1977). The Principle of Paternalism. *American Philosophical Quarterly* 14, 61-69.
- Kass, N.E. (2001). An Ethics Framework for Public Health. *American Journal of Public Health* 91(11), 1776-82.
- Kleinig, J. (1983). *Paternalism*. Manchester: Manchester University Press.
- Mill, J.S. (1859). *On Liberty*. (Many editions)
- Mitchell, G. (2004). Libertarian Paternalism is an Oxymoron. *Northwestern University Law Review* 99(3), 1245-1278.
- Sneddon, A. (2001). What's Wrong With Selling Yourself Into Slavery? Paternalism and Deep Autonomy. *Crítica* 33(98), 97-121.
- Shiffrin, S. (2000). Paternalism, Unconscionability Doctrine, and Accommodation. *Philosophy and Public Affairs* 29(3), 205-250.
- Sunstein, C.R. and Thaler, R.H. (2003). Libertarian Paternalism Is Not an Oxymoron. *The University of Chicago Law Review* 70, 1159-1202.
- Thaler, R.H. and Sunstein, C.R. (2003). Libertarian Paternalism. *The American Economic Review* 93, 175-179.

- Thaler, R.H. and Sunstein, C.R. (2009). *Nudge: Improving Decisions About Health, Wealth and Happiness*. 2d edition. New York: Penguin Books.
- Upshur, R.E.G. (2002). Principles for the Justification of Public Health Intervention. *Canadian Journal of Public Health* 93(2), 101-103.
- VanDeVeer, D. (1986). *Paternalistic Interference*. Princeton: Princeton University Press.
- White, M. (2010). Behavioral Law and Economics - The Assault on Consent, Will, and Dignity. In G. Gaus, C. Fodor and J. Lamond (Eds.) *New Essays on Philosophy, Politics and Economics: Integration and Common Research Projects*. Stanford: Stanford University Press.