PDF version of the entry
Paternalism
https://plato.stanford.edu/archives/fall2020/entries/paternalism/

from the Fall 2020 Edition of the

Stanford Encyclopedia
of Philosophy

Library of Congress Catalog Data
ISSN: 1095-5054

Paternalism

First published Wed Nov 6, 2002; substantive revision Wed Sep 9, 2020

Paternalism is the interference of a state or an individual with another person, against their will, and defended or motivated by a claim that the person interfered with will be better off or protected from harm. The issue of paternalism arises with respect to restrictions by the law such as anti-drug legislation, the compulsory wearing of seatbelts, and in medical contexts by the withholding of relevant information concerning a patient's condition by physicians. At the theoretical level it raises questions of how persons should be treated when they are less than fully rational.

1. Introduction

The government requires people to contribute to a pension system (Social Security). It requires motorcyclists to wear helmets. It forbids people from swimming at a public beach when lifeguards are not present. It forbids the sale of various drugs deemed to be ineffective. It forbids the sale of various drugs believed to be harmful. It does not allow consent to certain forms of assault to be a defense against prosecution for that assault.

The civil law does not allow the enforcement of certain kinds of contracts, e.g., for gambling debts. It requires minors to have blood transfusions even if their religious beliefs forbid it. Persons may be civilly committed if they are a danger to themselves.

Doctors do not tell their patients the truth about their medical condition. A physician may tell the wife of a man whose car went off a bridge into the water and drowned that he died instantly when in fact he died a rather ghastly death.

A husband may hide the sleeping pills from a depressed wife. A philosophy department may require a student to take logic courses.

A teacher may be less than honest about telling a student that he has little philosophical ability.

All of these rules, policies, and actions may be done for various reasons; may be justified by various considerations. When they are justified solely on the grounds that the person affected would be better off, or would be less harmed, as a result of the rule, policy, etc., and the person in question would prefer not to be treated this way, we have an instance of paternalism.

As the examples indicate the question of paternalism is one that arises in many different areas of our personal and public life. As such, it is an important realm of applied ethics. But it also raises certain theoretical issues. Perhaps the most important is: what powers it is legitimate for a state, operating both coercively and in terms of incentives, to possess? It also raises questions about the proper ways in which individuals, either in an institutional or purely personal setting, should relate to one another. How should we think about individual autonomy and its limits? What is it to respect the personhood of others? What is the trade-off, if any, between regard for the welfare of another and respect for their right to make their own decisions?

This entry examines some of the conceptual issues involved in analyzing paternalism, and then discusses the normative issues concerning the legitimacy of paternalism by the state and various civil institutions.

2. Conceptual Issues

The analysis of paternalism involves at least the following elements. It involves some kind of limitation on the freedom or autonomy of some agent and it does so for a particular class of reasons. As with many other concepts used in normative debate determining the exact boundaries of the concept is a contested issue.

And as often is the case the first question is whether the concept itself is normative or descriptive. Is application of the concept a matter for empirical determination, so that if two people disagree about the application to a particular case they are disagreeing about some matter of fact or of definition? Or does their disagreement reflect different views about the legitimacy of the application in question?

While it is clear that for some to characterize a policy as paternalistic is to condemn or criticize it, that does not establish that the term itself is an evaluative one. As 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 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 only because *X* believes *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*.

Condition one is the trickiest to capture. Clear cases include threatening bodily compulsion, lying, withholding information that the person has a right to have, or imposing requirements or conditions. But what about the following case? A father, skeptical about the financial acumen of a child, instead of bequeathing the money directly, gives it to another child with instructions to use it in the best interests of the first child. The first child has no legal claim on the inheritance. There does not seem to be an interference with the child's liberty nor on most conceptions the child's autonomy.

Or consider the case of a wife who hides her sleeping pills so that her potentially suicidal husband cannot use them. Her act may satisfy the second and third conditions but what about the first? Does her action limit the liberty or autonomy of her husband?

The second condition is supposed to be read as distinct from acting against the consent of an agent. The agent may neither consent nor not consent. He may, for example, be unaware of what is being done to him. There is

also the distinct issue of whether one acts not knowing about the consent of the person in question. Perhaps the person in fact consents but this is not known to the paternaliser.

The third condition also can be complicated. There may be more than one reason for interfering with *Y*. In addition to concern for the welfare of *Y* there may be concern for how *Y*'s actions may affect third-parties. Is the "just for" condition too strong? Or what about the case where a legislature passes a legal rule for paternalistic reasons but there are sufficient non-paternalistic reasons to justify passage of the rule?

If, in order to decide on any of the above issues, one must decide a normative issue, e.g., does someone have a right to some information, then the concept is not purely descriptive. Ultimately the question of how to refine the conditions, and what conditions to use, is a matter for philosophical judgment. The term "paternalism" as used in ordinary contexts may be too amorphous for thinking about particular normative issues. One should decide upon an analysis based on a hypothesis of what will be most useful for thinking about a particular range of problems. One might adopt one analysis in the context of doctors and patients and another in the context of whether the state should ban unhealthy foods.

Given some particular analysis of paternalism there will be various normative views about when paternalism is justified. The following terminology is useful.

2.1 Hard vs. soft paternalism

Soft paternalism is the view that the only conditions under which state paternalism is justified is when it is necessary to determine whether the person being interfered with is acting voluntarily and knowledgeably. To use Mill's famous example of the person about to walk across a damaged

bridge, if we could not communicate the danger (he speaks only Japanese) a soft paternalist would justify forcibly preventing him from crossing the bridge in order to determine whether he knows about its condition. If he knows, and wants to, say, commit suicide he must be allowed to proceed. A hard paternalist says that, at least sometimes, it may be permissible to prevent him from crossing the bridge even if he knows of its condition. We are entitled to prevent voluntary suicide.

2.2 Broad vs. narrow paternalism

A narrow paternalist is only concerned with the question of state coercion, i.e., the use of legal coercion. A broad paternalist is concerned with any paternalistic action: state, institutional (hospital policy), or individual.

2.3 Weak vs. strong paternalism

A weak paternalist believes that it is legitimate to interfere with the means that agents choose to achieve their ends, if those means are likely to defeat those ends. So if a person really prefers safety to convenience then it is legitimate to force them to wear seatbelts. A strong paternalist believes that people may have mistaken, confused or irrational ends and it is legitimate to interfere to prevent them from achieving those ends. If a person really prefers the wind rustling through their hair to increased safety it is legitimate to make them wear helmets while motorcycling because their ends are irrational or mistaken. If one is a weak but not a strong paternalist we may only interfere with mistakes about the facts not with mistakes about values. So if a person tries to jump out of a window believing he will float gently to the ground we may restrain him. If he jumps because he believes that it is important to be spontaneous we may not.

2.4 Pure vs. impure paternalism

Suppose we prevent persons from manufacturing cigarettes because we believe they are harmful to consumers. The group we are trying to protect is the group of consumers not manufacturers (who may not be smokers at all). Our reason for interfering with the manufacturer is that he is causing harm to others. Nevertheless the basic justification is paternalist because the consumer consents (assuming the relevant information is available to him) to the harm. It is not like the case where we prevent manufacturers from polluting the air. In pure paternalism the class being protected is identical with the class being interfered with, e.g., preventing swimmers from swimming when lifeguards are not present. In the case of impure paternalism the class of persons interfered with is larger than the class being protected.

2.5 Moral vs. welfare paternalism

The usual justification for paternalism refers to the interests of the person being interfered with. These interests are defined in terms of the things that make a person's life go better; in particular their physical and psychological condition. It is things like death or misery or painful emotional states which are in question. Sometimes, however, advocates of state intervention seek to protect the moral welfare of the person. So, for example, it may be argued that prostitutes are better off being prevented from plying their trade even if they make a decent living and their health is protected against disease. They are better off because it is morally corrupting to sell one's sexual services. The interference is justified, therefore, to promote the moral well-being of the person. This then can be called moral paternalism. Still another distinction within moral paternalism is between interferences to improve a person's moral

character, and hence her well-being, and interferences to make someone a better person—even if her life does not go better for her as a result.

Finally, it is important to distinguish paternalism, whether welfare or moral, from other ideas used to justify interference with persons; even cases where the interference is not justified in terms of protecting or promoting the interests of others. In particular moral paternalism should be distinguished from legal moralism, i.e., the idea that certain ways of acting are morally wrong or degrading and may be prohibited. So, for example, the barroom "sport" of dwarf tossing (where dwarfs who are paid, and are protected with helmets, etc. participate in contests to see who can throw them furthest) might be thought to be legitimately prohibited. Not because the dwarf is injured in any way, not because the dwarf corrupts himself by agreeing to participate in such activities, but simply because the activity is wrong.

To be sure it is not always easy to distinguish between legal moralism and moral paternalism. If one believes, as Plato does, that acting wrongly damages the soul of the agent, then it will be possible to invoke moral paternalism rather than legal moralism. What is important is that there are two distinct justifications that are possible; one appealing to the mere immorality of the conduct interfered with, the other to the harm done to the agent's character.

3. Normative Issues

Is there a burden of proof attached to paternalism? Does the paternalist or anti-paternalist have to give a reason for their action? As we have seen the analysis of paternalism seems to cut both ways. It is an interference with liberty which might be thought to place the burden of proof on the paternalist. It is an act intended to produce good for the agent which might be thought to place the burden of proof on those who object to

paternalism. It might be thought, as Mill did, that the burden of proof is different depending on who is being treated paternalistically. If it is a child then the assumption is that, other things being equal, the burden of proof is on those who resist paternalism. If it is an adult of sound mind the presumption is reversed.

Suppose we start from the presumption that paternalism is wrong. The question becomes under what, if any, circumstances, can the presumption be overcome? The possible answers are "under no circumstances", "under some circumstances", and "under any circumstances"

The last seems very implausible. Essentially it is the view that the fact that an act is (intended to be) beneficial for a person, and does not affect or violate the interests of others, settles the question of whether it may be done. Only a view which ignores the means by which good is promoted, and the ethical status of such means, can hold this. Any sensible view has to distinguish between good done to agents at their request or with their consent, and good thrust upon them against their will.

So the normative options seem to be just two. Either we are never permitted to aim at doing good for others against their wishes, and in ways which limit their liberty, or we are permitted to do so.

Why might one think that at least the state may never do so? One might think so because of various beliefs about the impossibility of in fact doing good for people against their will or because one thinks that although possible to do good it is in fact inconsistent with some normative standard which ought to prevail.

With respect to the impossibility question one might believe either that it is not possible to do any good by acting paternalistically or that although it is possible to do some good the process will (almost) always produce bads which outweigh the good.

If one thought that (almost) always more harm than good is done by the state when it acts paternalistically this raises the question of whether we can distinguish the conditions in which (rarely) more good than harm is done and build that into our guidelines. If this is possible, and allowing paternalism in these exceptional cases does not create further harms which outweigh the good produced, then we should sometimes be paternalists. If it is impossible to distinguish the "good" from the "bad" cases then, at least if we are rule consequentialists, we ought not to have such a rule; and we ought not to try and make the distinctions on a case by case basis.

But one might believe that the question of whether more good than harm is produced is not simply an empirical one. It depends on our understanding of the good of persons. If the good simply included items such as longer life, greater health, more income, or less depression, then it makes it look like an empirical issue. But if we conceive of the good of individuals as including items such as being respected as an independent agent, having a right to make decisions for oneself, or having one's autonomy not infringed, then the issue of whether the agent is better off after being paternalised is partly a normative matter. One might believe that one cannot make people better off by infringing their autonomy in the same way that some people believe one cannot make a person better off by putting them in a Nozickian experience machine (one in which they are floating in a tank but seem to be having all kinds of wonderful experiences). Compare Mill's statement that "…a man's mode of laying out his own existence is best not because it is the best in itself, but because it is his own mode…" (1859: Chapter III).

Kantian views are frequently absolutistic in their objections to paternalism. On these views we must always respect the rational agency of other persons. To deny an adult the right to make their own decisions, however mistaken from some standpoint they are, is to treat them as simply means to their own good, rather than as ends in themselves. In a way anti-paternalism is already incorporated into Kantian theories by their prohibition against lying and force—the main instruments of paternalistic interference. Since these instrumentalities are already denied even to prevent individuals from harming others, they will certainly be forbidden to prevent them from harming themselves. Of course, one may object to the former absolutism while accepting the latter.

If one believes that sometimes paternalism is justifiable one may do so for various kinds of theoretical reasons. The broadest is simply consequentialist, i.e., more good than harm is produced. A narrower justification is that sometimes the individual's (long-run) autonomy is advanced by restricting his autonomy (short-run). So one might prevent people from taking mind-destroying drugs on the grounds that allowing them to do so destroys their autonomy and preventing them from doing so preserves it. This is essentially Mill's argument against allowing people to contract into slavery. Note that if the theory of the good associated with a particular consequentialism is broad enough, i.e., includes autonomy as one of the goods, it can be equivalent to the autonomy theory (assuming that the structure of the autonomy view is a maximizing one).

A different theoretical basis is (moral) contractualism. On this view if there are cases of justified paternalism they are justified on the basis that we (all of us) would agree to such interference, given suitable knowledge and suitable motivation. So, for instance, it might be argued that since we know we are subject to depression we all would agree, at least, to short-term anti-suicide interventions, to determine whether we are suffering from such a condition, and to attempt to cure it. More generally, we might accept what Feinberg called "soft paternalism." This is the view that when we are not acting fully voluntarily it is permissible to intervene to provide information, or to point out defects in our rationality, but that if we then do make a voluntary choice it must be respected. Or we might agree to being forced to wear seat-belts knowing our disposition to discount future

benefits for present ones. The justification here is neither consequentialist nor based simply on the preservation of autonomy. Rather either kind of consideration may be taken into account, as well as others, in determining what we would reasonably agree to.

4. Libertarian Paternalism

In recent years there has been a new, influential, strand of thought about paternalistic interferences. It has been referred to as New Paternalism or Libertarian Paternalism. It is influenced by research in the behavioral sciences on the many ways in which our cognitive and affective capacities are flawed and limited.

The first theorists to emphasize these findings for making social policy were the Nudgers—Cass Sunstein and Richard Thaler (2003). They argued that since people were such bad decision makers we should nudge them in the direction of their own desired goals by orchestrating their choices so that they are more likely to do what achieves their ends.

The claim is that, unlike traditional paternalism which rules out choices by compulsion or adds costs to the choices by coercion, nudges simply change the presentation of the choices in such a way that people were more likely to choose options that are best for them. In addition they argue that any arrangement of choices will make some choices more or less likely so that some decision about the choice architecture is inevitable.

The first issue is what exactly distinguishes Nudges from other ways of influencing people's choices. Here are various examples of Nudges. These were given in the earliest discussions of nudging and have tended to be the ones focused on.

> *Cafeteria*. In order to influence students to make healthier food choices as they pass the cafeteria options place the healthy foods at eye level and place the less healthy choices higher or lower than eye level. Sometimes the nudge involves putting the healthful food earlier in the line.
>
> *Opt-In vs Opt-Out*. Given that many employees often fail to enroll (opt-in) in retirement plans, employers make the default automatic enrollment in such programs, allowing employees to easily opt-out. Such programs have been shown to increase savings rates.
>
> *Save More Tomorrow*. Employees are asked to commit now to having a portion of their salary increase in subsequent years put directly into their pension plan. People are loss averse, and thus more willing to forgo a raise in take-home income than they are to actively re-direct the additional funds they have already received to their retirement accounts each year.
>
> *Size of Plates*. Using smaller plates in a cafeteria cuts down on the amount of food consumed.
>
> *Painting Traffic Lanes*. In order to get drivers to slow down on a sharp turn the lane lines are painted closer together than usual. This creates the illusion for the drivers that they are driving faster than they actually are and they slow down as a result.

The initial examples above served as the focus of much of the early literature. The early critics attempted to distinguish interventions such as Cafeteria from interventions such as merely providing information. It is clear from more recent writings that the category of nudges is intended to be quite broad. According to Sunstein all the following are nudges: reminders, warnings, a GPS, disclosure of the interest rate of a bank card, any information about what people like you do, simplification of government forms, default rules, subliminal messages urging people to eat healthy food.

Thaler and Sunstein also provided a characterization of a nudge as

> any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentive. (2008: 6)

4.1 Definitional issues

Why is the view labeled by its introducers as Libertarian Paternalism? In the paradigmatic definitions of nudging the intervention is always by way of what is called "choice architecture." Choice architecture is the design of different ways in which choices can be presented to an agent. Examples include the number of choices, whether the choice is opt-in or opt-out, the way in which alternatives are described or presented, the incentives attached to the choices, etc.

Their view is Libertarian because it preserves freedom of choice. No choice is eliminated or made more difficult. Nobody is coerced. The choice set remains the same. No significant costs or incentives are attached to the choices the agent faces.

Their view is Paternalistic because it seeks to promote the good of the agent being nudged. And it is the good as viewed by the agent herself. We are not nudging towards ends she does not hold. Nudging is about means not ends.

Their definition of Paternalism is very weak in the sense that it allows many more acts to count as paternalistic than would be under almost all traditional definitions of paternalism.

In terms of the analysis of Paternalism given in this entry is Nudging paternalistic? The first condition in the definition is: the action (or its omission) interferes with the liberty or autonomy of *Y*. Nothing

corresponds to this in the definition above. Putting a warning label on a cigarette pack does not interfere with the liberty of autonomy of any cigarette smoker.

Basically, the definition of paternalism in Libertarian Paternalism is focused solely on the fact that nudges are being used to make the agents being nudged better off. We could replace "paternalism" with "benevolence" and nothing important would be left out since the "libertarian" aspect picks up everything else that is significant. Whether this expansion of the definition of paternalism is warranted or not is a matter of what issues are being explored and whether such an expansion makes things clearer or more confused.

There are nudges which are not paternalistic (on their definition) because the aim is to promote the general good—even if the chooser is not benefitted. Nudging building managers to put in elevators with braille buttons, influencing people to contribute to Oxfam by putting up pictures of starving infants, are examples where the good to be promoted is the welfare of people other than those being influenced.

However one comes out on the issue of whether the definition of paternalism is useful or not we turn to the more important issues about whether, and in what circumstances, nudges are justifiable ways of influencing persons to make certain choices.

4.2 Normative Issues

Given that nudges are not coercive, that they are intended to promote the good of individuals as they themselves perceive that good, that they have been shown to often be effective, are there any plausible normative objections to their use?

As with any policy intervention, either by the state or by private organizations, there are possible misuses to worry about. Perhaps there are slippery-slopes to be avoided. Perhaps proponents of nudging over-estimate the amount and seriousness of faulty reasoning by agents; mistakes that nudgers wish to harness to promote the agents welfare. Perhaps they are mistaken about what agents really value when they claim people prefer health to more sugary beverages.

But these objections are not objections to nudging but to the misuse of this type of behavioral intervention. Are there objections to the very nature of nudging itself?

There is one feature of many nudges that must be considered which, although not intrinsic to the concept of a nudge, is often present in the background as a crucial feature. One author actually links these background conditions to the definition of Libertarian Paternalism.

> Libertarian Paternalism is the set of interventions aimed at overcoming the unavoidable cognitive biases and decisional inadequacies of an individual by exploiting them in such a way as to influence her decisions (in an easily reversible manner) towards the choices she herself would make under idealised conditions. (Rebanato 2012: 6)

An example of this use of cognitive biases is changing opt-in to opt-out. It is because of cognitive bias to doing nothing to change the status quo that there are relatively fewer opt-outs than might be expected.

Given this background, there are at least three objections that have objected to features intrinsic to some—by no means all—nudges. The first is that nudging often occurs without the nudged being aware they are being nudged. The second is that nudging often works by harnessing defects in the thinking of those being nudged. The third is that some

nudges (besides those subject to the first two objections) are forms of objectionable manipulation.

4.3 Transparency

One issue with many nudges is what the person being nudged knows about the nudge. In the Cafeteria example the students are aware that food has been placed at different levels of eyesight. In that sense the nudge is transparent to them. It is not like subliminal messaging in which they are not aware of messages directed to them. Let us call nudges which are transparent in this sense narrow nudges.

For an example of a nudge that is not even narrowly transparent consider the experiment where, in order to increase office worker's payments for the coffee they take from the coffee machine a painting of a pair of eyes is hung above the coffee pot. This did increase the rate of payment. Most workers questioned about the painting either had not noticed it at all or made no connection with the issue of payment.

In the cafeteria example while aware that the food is on different levels the students are not aware that the placing of the food has been done in order to promote a certain end—eating more healthy foods. The placement of the food is not random, nor motivated by aesthetic considerations. It is deliberate and motivated by a particular set of considerations. Some nudges are more transparent in the sense that it is obvious they have been deliberately introduced and their motivation is also clear. For example, the warnings on cigarette packages that smoking is dangerous to one's health. Call these broad nudges.

There is some evidence that making nudges broad does not interfere with their efficacy. A recent study in the context of end-of-life care showed the

effect of a default is not weakened when people are told that a default was chosen because it is usually effective (Lowenstein et al. 2014).

Note that there could be an even more transparent feature of nudges—call them very broad nudges. This would be when the mechanism by which the nudge influences is made public as well. Suppose we presented an opt-out set up and said (1) we are doing this to increase participation in the retirement program, and (2) if this is effective it is because people have a tendency to stick with the status-quo.

Nudges which are neither narrow nor broad—such as subliminal messages to movie-goers to buy fruit instead of popcorn—might be an effective way of encouraging consumption of healthier food. But they seem to have a morally dubious character. Even if the facts are that such messages have rather weak efficacy we object to their by-passing any possibility to avoid or resist them.

Sunstein has put forward what he calls a transparency condition:

> Choice architecture should be transparent and subject to public scrutiny, certainly if public officials are responsible for it. At a minimum, this proposition means that when such officials institute some kind of reform, they should not hide it from the public…If officials alter a default rule so as to promote clean energy or conservation, they should disclose what they are doing. (2015: 19)

This formulation leaves open a number of issues. Are "transparent" and "subject to public scrutiny" different conditions? The "minimum" interpretation is only a condition of publicity about what officials are doing. The government could announce in advance that they are going to use subliminal messages on television programs to promote health. Sunstein believes that this would satisfy his transparency condition but that it might be objectionable on grounds of manipulation.

It remains open for discussion how to formulate a norm which is closer to the intuitive idea of transparency, to distinguish various sense of transparency—such as narrow and broad—and to debate whether such transparency is a necessary component of legitimate nudges. Must, for example, the public utility which hopes to encourage energy conservation preface its informational message about the average consumption of your neighbors with the fact that they are sending this information because they think it will encourage conservation? Does it have to disclose they believe it may do so because people have a tendency to conform to what their neighbors are doing?

Why is transparency required? One possible objection to non-transparency is that it interferes with the autonomy of those influenced. It seems too strong to think that this is true of all non-transparent nudges but a more limited claim is that a nudge limits autonomy when it influences an agent's choices and were she aware of the nudge she would reject the influence and the influence would no longer be effective in her choices.

Another issue concerning autonomy is whether it is affected by both intentional and unintentional nudges. If the cafeteria manager places the food at random she will still be influencing people's choices. Are the choices more autonomous in this case than in a case in which the food is placed in exactly the same way but deliberately in order to affect the choices?

4.4 Harnessing Bad Reasoning

The second objection to nudges has to do with a specific mechanism through which the end of nudging—promotion of agent ends—is sometimes accomplished. Consider the cafeteria example. The reason we place the healthy foods at eye level is because there is a tendency to

choose what is at eye level over options that are not. The thought is that nudgers can harness this tendency by putting healthy foods at that level.

Since the positioning of foods is not a rational ground for choosing nudgers use this non-rational tendency so that healthy foods are chosen. Note in this case we get both a lack of transparency and the harnessing of non-rational tendencies.

Some argue that taking advantage of our non-rational tendencies, even for good ends, is objectionable.

Consider the opt-out nudge. It relies upon, and works in virtue of, the fact that we tend to go with the given even if there are better options easily available. It is because we irrationally choose the worse option that we present the better option for the agent to choose irrationally.

Framing effects: It is one of the most confirmed findings of empirical decision theory that subjects decisions are affected by different ways of presenting information. For example, in deciding on whether to have an operation which can cure one's disease but has the possibility of causing death how one chooses is affected by whether one is told (A) or (B):

A. 90% of patients who have the operation survive. (survival rate)
B. 10% of the patients who have the operation die. (mortality rate)

This is exactly the same information but those told (A) are more likely to choose the operation than those given (B). It is irrational to make the decision differently depending on how it is worded.

This harnessing of the irrational for our own good is not paradoxical but it strikes some as problematic in the same way getting children to read by offering them financial incentives is problematic. We are getting them to read for the wrong reasons. At least in these cases there is the idea that once reading they will come to appreciate the pleasures and importance of reading for its own sake. But do people who stick with opt-in out of a tendency to stick with the given learn to change their faulty heuristic? If anything, it is reinforced because their faulty heuristic has a good consequence.

If we think of cases of rational persuasion, then in the ideal case, we would find that the agent chooses because she believes she has been given reasons, these reasons support her choice, and she acts because of those reasons. In the case of harnessing non-rational tendencies for nudges these conditions are not satisfied.

It is a good thing that, usually, we act not simply in accordance with the reasons there are to act, but also out of, in recognition, of those reasons. This desirable feature may be out-weighed if the goods accomplished by nudges are important, and alternative interventions are much less effective and/or require much costly or difficult to operate.

It is clear that while many nudges (as defined) harness bad reasoning, most do not. Some do not harness reasoning at all, e.g., the eyes/coffee case.

4.5 Manipulation

Since nudges are defined to exclude coercion, and they usually are not cases of outright deception (as opposed to a lack of transparency) the concept that is often used to criticize nudges is that of manipulation. The charge of manipulation is raised often against the acts of others even when, like nudging, they are benevolently motivated. We think of manipulation, like other forms of paternalism as failing to respect us as rational and capable choosers. After all, if we were capable choosers why not just present us with the reasons which favor our acting in particular way?

> Nudging uses the clever tricks of modern psychology and economics to manipulate people. We don't like manipulation when it's done to sell us things; we shouldn't like manipulation when our governments do it to us. (Wilkinson, see Other Internet Resources)

The problem is that manipulation seems a very amorphous and ill-understood concept. There is widespread disagreement about what kinds of influence are manipulative and the conditions under which they are wrong.

There seems to be only general agreement on the idea that in some way or other manipulation interferes, or perverts, or takes advantage of factors that people would not want to influence, the decision making of agents. Whether manipulation must be intentional, whether it must be hidden, whether the motive of the manipulator matters, whether there has to be a gap between the way in which the influence causes behavior and the reasons which justify it , whether there has to be a manipulator if one is manipulated, all are contested in the literature (see papers in Coons and Weber 2013).

Even if there were a consensus on the widely shared view that interpersonal manipulation is unjustified only when there is a bypass or subversion of the rational capacities of the person being influenced there will be much disagreement about what "bypass" and "subverts" come to. Does making use of framing effects in conveying information "bypass" or "subvert" rational capacities? Is harnessing a non-rational propensity of a person bypassing or subverting rational capacities?

Perhaps people perceive a process as manipulative only if they already disapprove of it for other reasons. A recent series of experiments discovered that people's views of whether a given nudge was manipulative

varied with whether the nudge was in the direction of their political convictions or not (Fox and Tannenbaum 2015).

In almost every case, respondents on the left of the political spectrum supported nudges when they were illustrated with a liberal agenda but opposed them when they were illustrated with a conservative one; meanwhile, respondents on the political right exhibited the opposite pattern.

Given the very different conceptions of manipulation there is disagreement about why, when it is, manipulation is wrong. Because it violates dignity? Because it violates autonomy? Because it violates a conception of liberty?

It is clear that many nudges are not plausibly examples of manipulation. Warning labels, default rules such as opt-out, providing caloric information on menus cannot count as manipulation without using such an expansive conception of manipulation as to deprive it of any use.

Nudgers are clear that they want the influence they use to be easily avoidable. That is why they do not consider making choices very costly or difficult to be a nudge. But manipulations vary in their strength or effectiveness. Perhaps certain subliminal messages are quite weak in their force; only people who already are thinking about buying popcorn are affected. Would only the influences that are difficult to evade or avoid be considered manipulative?

In the case of paternalistic acts there seemed to be only one or two concepts which figure in the normative objections—e.g., autonomy, liberty—and they belong to a similar class of values. In the case of objectionable nudging there seem to be a greater diversity of normative values at stake, and they seem to have no overarching conceptual unity.

Until more refined notions of manipulation and of subverting rational decision-making are developed it may be more fruitful to look at specific nudges which strike one as problematic because of some identifiable features they have, and to distinguish them from other nudges which lack such features. There may be no common features which explain why those nudges that are wrongful all fall under a plausible concept of manipulation.

5. Paternalistic Lies

There is one class of manipulation which clearly raises issues about whether the person being manipulated had had their autonomy and/or liberty interfered with--lies. Lies motivated by the claim that the person being lied to will have her interests advanced are the most common examples of paternalism that are claimed to be, at least sometimes, justified.

Paternalistic lies, and other forms of deception such as truthful but misleading assertions, raise interesting conceptual as well as normative issues. Unlike coercion which makes certain choices more costly, and hence restricts agents' options, many lies make it more difficult for the agent to take certain options without increasing their cost. Further this difficulty is hidden from the person being lied to. This raises the issue of what exactly are the features of the person being lied to which are affected. Is it her liberty, her freedom, her autonomy? Is it the voluntariness of her action? Is it none of these things but something like her ability to accomplish her aims? Thinking about these questions is important in evaluating whether, and when, paternalistic lies are justifiable.

This discussion is about lying not deception. One can deceive without lying, e.g. by asserting something which you believe true but from which

you know that the hearer will infer something false. Or by not asserting at all as when a husband replaces his sleeping pills with cough medicine to prevent his wife from harming herself. By a lie I mean an assertion *P* by *A* who does not believe that *P* and intends a listener to either come to believe that *P* or at least to believe that *A* believes that *P*.

A lie may not in fact deceive the listener for many reasons—*P* might actually be true or the listener may be aware that she is being lied to—but it is still pro tanto wrong as an abuse of assertion whose function is to provide evidence of what the speaker takes to be the truth. Paternalistic lies are prima facie wrong, as Kantians would claim, because they are violations of our right as rational creatures to determine our ends, and the means to obtain them, for ourselves.

Since lying is pro tanto wrong any specific lie can turn out to be justified if there are reasons that outweigh the ones making it wrong. And defenders of paternalistic lies argue there are sometimes harms avoided or goods produced that outweigh the harms of lying. Anti-paternalists argue that such calculations fail to take into account that lies interfere with important values such as the autonomy, liberty, and rationality of the person and these values may outweigh the harms avoided or goods produced.

Let us start with autonomy. Does being lied to interfere with, or diminish the autonomy of the person being lied to? Obviously to answer this we need to be given an understanding of what it is for a person to be, or to act, autonomously. For example, if we understand autonomy as a general characteristic of a person's relationship to their goals, ideals, preferences, values, and so forth then the lie must affect that relationship. If the lie, for example, asserts that some value, e.g. politeness, causes mostly misery, then their autonomy understood as self-reflection on how they should live is undermined.

If, on the other hand, the lie is relevant to how, given their ends, they can best attain them then what is diminished is their likelihood of achieving their autonomously chosen ends, or of acting in accordance with their values. Their capacity for reflective acceptance remains intact.

At the level of choosing what to do in order to attain goals, or promote ends, or acting in accordance with values, being lied to diminishes the likelihood of achieving any of these. Is this failure also to be considered a diminishment of autonomy?

First, it should be noted that being lied to does not, and is not always intended to, affect choices. When a doctor falsely tells a patient that he does not have a fatal disease she is trying to prevent the patient from becoming depressed. It is mood not particular choices that are at issue.

Second, it not clear that having choices fail because of a lie affects autonomy. If it does then whenever a person, out of ignorance, has their choice fail they have acted non-autonomously. But it is not obvious that when I choose a medicine on the basis of inaccurate information, and fail to get cured, that my choice of the medicine was not autonomous. Perhaps it is the voluntariness of my choice that is affected.

One way to think about this is to consider the exercise of autonomy in a given choice as requiring freedom as non-interference. This assume the actor is autonomous in some basic capacity sense. And non-interference is then analyzed in terms of what makes one's choices more costly(coercion) or more difficult(deception) to pursue.

An interesting case to think about is the prescribing of placebos. In the typical case the doctor is mis-leading the patient about the inert nature of the medication. But he is doing so because there is good reason to think that the patient will be better off for having taken the placebo. The patient's practical reasoning is distorted by the lie so something important for autonomy is removed but the patient's welfare is increased.

In conclusion it seems clear that lying interferes with something significant concerning the choices and actions of the person lied to and in doing so acts prima facie wrongly. But being clearer about what is, and what what is not, diminished by paternalistic lies needs further investigation in order to evaluate the legitimacy of such lies.

Bibliography

Alexander, Larry, 2010, "Voluntary Enslavement", *San Diego Legal Studies Paper No. 10–042*, October 19, 2010. [Alexander 2010 available online]

Arneson, Richard J., 1989, "Paternalism, Utility, and Fairness", *Revue Internationale de Philosophie*, 43(170(3)): 409–23.

——, 2005, "Joel Feinberg and the Justification of Hard Paternalism", *Legal Theory*, 11(3): 259–284. doi:10.1017/S1352325205050147

Bergelson, Vera, 2007, "The Right to Be Hurt: Testing the Boundaries of Consent", *The George Washington Law Review*, 75: 165–255.

Boven, Luc, 2008, "The Ethics of Nudge", in Grune-Yanoff and S.O. Hansson, *Preference Change: Approaches from Philosophy, Economics and Psychology*, Chapter 10, Berlin and New York: Springer.

Brink, David O., 2013, *Mill's Progressive Principles*, Oxford: Oxford University Press.

Castro, Clinton, Adam Phan, and Alan Rubel, 2020, *Epistemic Paternalism Online in Guy Axtell & Amiel Bernal (eds.) Epistemic Paternalism.*, London: Rowman & Littlefield.

Cholby, M, 2015, "Paternalism and our Rational Powers", *Mind*, 126(501): 123–153.

Conly, Sarah, 2012, *Against Autonomy: Justifying Coercive Paternalism*,

Cambridge: Cambridge University Press.
Coons, Christian and Michael Weber (eds), 2013 *Paternalism: Theory and Practice*, Cambridge: Cambridge University Press.
Dixon, Nicholas, 2001, "Boxing, Paternalism, and Legal Moralism", *Social Theory and Practice*, 27(2): 323–344.
de Marneffe, Peter, 2006, "Avoiding Paternalism", *Philosophy and Public Affairs*, 34: 68–94. doi:10.1111/j.1088-4963.2006.00053.x
Dworkin, Gerald, 1972, "Paternalism", *The Monist*, 56: 64–84.
——, 2005, "Moral Paternalism", *Law and Philosophy*, 24(3): 305–319.
——, 2013, "Defining Paternalism", in Coons and Weber 2013: 25–39.
——, 2012, "Harm and the Volenti Principle", *Social Philosophy and Policy*, 29: 309–321. doi:10.1017/S0265052511000057
Enoch, David, 2016, What's Wrong with Paternalism: Autonomy, Belief, and Action *Proceedings of the Aristotelian Society*, April 2016
Feinberg, Joel, 1986, *Harm to Self*, Oxford: Oxford University Press.
Fox, Craig R. and David Tannenbaum, 2015 "The Curious Politics of the 'Nudge'", *New York Times*, September 26, 2015, p. SR9.
Goldman, Alan H., 1980, "The Refutation of Medical Paternalism", in his *The Moral Foundations of Professional Ethics*, Towata: Rowman and Littlefield.
Goodin, Robert E., 1991, "Permissible Paternalism: Saving Smokers from Themselves", *The Responsive Community* 1: 42–51. [Republished in LaFollette, Hugh (ed.), 2014, *Ethics in Practice: An Anthology*, Hoboken: Wiley-Blackwell.]
Gorin, Moti, 2014, "Do Manipulators always threaten Rationality?" *American Philosophical Quarterly*, 51(1): 51–61.
Gostin, Lawrence O. and Kieran G. Gostin, 2009, "A broader liberty: J. S. Mill, Paternalism and the Public's Health", *Public Health*, 123: 214–22. doi:10.1016/j.puhe.2008.12.024
Grill, Kalle and Hanna Jason, 2018, *The Routledge Handbook of the Philosophy of Paternalism*, Routledge.

Groll, Daniel, 2012 "Paternalism, Respect and the Will", *Ethics*, 122(4): 692–714. doi:10.1086/666500
Hanna, Jason, 2018, *In Our Best Interest* Oxford University Press.
Hector, Colin, 2012, "Nudging towards Nutrition: Soft Paternalism and Obesity-Related Reform", *Food & Drug Law Journal*, 67(1): 103–122.
Husak, Douglas, 2003, "Legal Paternalism", in Hugh LaFollette (ed.), *The Oxford Handbook of Practical Ethics*, New York: Oxford University Press.
Kleinig, John, 1983, *Paternalism*, Towata: Rowman and Allenheld.
Kultgen, John, 1995, *Autonomy and Intervention: Paternalism in the Caring Life*, New York City: Oxford University Press.
Le Grand, Julian and Bill New, 2015, *Government Paternalism: Nanny State or Helpful Friend?* Princeton: Princeton University Press.
Mill, John Stuart, 1859, *On Liberty*, Indianapolis: Bobbs-Merrill, 1956.
Nys, Thomas, Yvonne Denier, and Toon Vandevelde (eds), 2007, *Autonomy and Paternalism: Reflections on the Theory and Practice of Health Care*, Leuven: Peeters.
Pope, Thaddeus Mason, 2000, "Balancing Public Health against Individual Liberty: The Ethics of Smoking Regulations", *University of Pittsburgh Law Review*, 61(2): 419–498.
Pugh, Jonathan, 2015, "Ravines and Sugar Pills: Defending Deceptive Placebo Use", *Journal of Medicine and Philosophy*, 40(1): 83–101.
Rebanato, Riccardo, 2012, *Taking Liberties: A Critical Examination of Libertarian Paternalism*, London: Palgrave Macmillan. doi:10.1057/9780230391567
Saghai, Yashar, 2013, "Salvaging the Concept of Nudge", *Journal of Medical Ethics*, 39(8): 487–493. doi:10.1136/medethics-2012-100727
Salazar V., Alberto R., 2012, "Libertarian Paternalism and the Danger of Nudging Consumers", *King's Law Journal*, 23(1): 51–67. doi:10.5235/096157612800081222

Sartorius, Rolf, 1983, *Paternalism*, Minneapolis, MN: University of Minnesota Press.

Savelscu, Julian, 1995, "Rational non interventional paternalism: why doctors ought to make judgments of what is best for their patients", *Journal of Medical Ethics*, 21: 327–331.

Shiffrin, Seanna, 2000, "Paternalism, Unconscionability Doctrine, and Accommodation", *Philosophy & Public Affairs*, 29: 205–250.

Skipper, Robert A., 2012, "Obesity: Towards a System of Libertarian Paternalistic Public Health Interventions", *Public Health Ethics*, 5: 181–191. doi:10.1093/phe/phs020

Sunstein, Cass R., 2013, "The Storrs Lectures: Behavioral Economics and Paternalism", *Yale Law Journal*, 122(7): 1826–1900. [Sunstein 2013 available online]

——, 2015, "The Ethics of Nudging", *Yale Journal on Regulation*, Vol. 32: 413–450. [Sunstein 2015 available online]

Sunstein, Cass R., and Richard H. Thaler, 2003, "Libertarian Paternalism Is Not an Oxymoron", *University of Chicago Law Review*, 70(4): 1159–1202.

Thaler, Richard H. and Cass R. Sunstein, 2008, *Nudge: Improving Decisions about Health, Wealth and Happiness*, New Haven, CT: Yale University Press.

Turner, Piers Norris, 2013, "The absolutism Problem in *On Liberty*", *Canadian Journal of Philosophy*, 43(3): 322–340. doi:10.1080/00455091.2013.847346

VanDeVeer, Donald, 1986, *Paternalistic Intervention: The Moral Bounds on Benevolence*, Princeton, NJ: Princeton University Press.

Weber, Michael, and Christian Coons, 2014, *Manipulation: Theory and Practice*, Oxford: Oxford University Press.

Wilkinson, T.M., 2013, "Nudging and Manipulation", *Political Studies*, 61(2): 341–355. doi:10.1111/j.1467-9248.2012.00974.x

Academic Tools

- How to cite this entry.
- Preview the PDF version of this entry at the Friends of the SEP Society.
- Look up this entry topic at the Internet Philosophy Ontology Project (InPhO).
- Enhanced bibliography for this entry at PhilPapers, with links to its database.

Other Internet Resources

- Loewenstein, George, Cindy Bryce, David Hagmann, & Sachin Rajpal, 2014, "Warning: You Are About To Be Nudged", unpublished working paper (Mar. 28, 2014), archived at SSRN
- Wilkinson, Martin, "Nudges manipulate, except when they don't", August 2013.

Related Entries

autonomy: in moral and political philosophy | autonomy: personal | beneficence, principle of | liberty: positive and negative | limits of law | Mill, John Stuart: moral and political philosophy