

## Paternalism and Manipulation

Danny Scoccia, Dept of Philosophy, NMSU  
2/18/14; word count (w/ notes and references): ~12,400

Richard Thaler and Cass Sunstein have recently defended state paternalism that uses non-coercive “nudging” as consistent with libertarianism’s insistence on the inviolability of negative liberty rights.<sup>1</sup> Among the policy changes they recommend is making employee contribution to retirement savings the default option. Such a change would increase saving by exploiting people’s inertia. Many liberals, especially the libertarians whom Thaler/Sunstein seem most eager to win over, have not warmed to their proposals.<sup>2</sup> Some have rightly observed that paternalistic nudges are manipulative and claimed that this makes them wrong. Is manipulative paternalism *pro tanto* wrong? If so, why? What considerations if any might override this wrongness to render some of it all things considered justified?

This paper offers answers to these questions. It begins with a brief account of what manipulation is, distinguishing it from coercion, incentives, and compulsion. It then considers four possible objections to a manipulative paternalism aimed at getting people to quit smoking or the like: i) it fails to benefit its targets b/c it underestimates the value of rational autonomy; ii) it fails to benefit them b/c it overestimates the value of goods like health and longevity; iii) it disrespects its targets, implying that they lack the capacity to make good choices; and iv) it infringes their right to autonomy. I’ll argue that i)-iii) fail. iv) identifies the wrongness making feature of paternalism that is *pro tanto* wrong, namely, “hard” paternalism. But the right to autonomy, rather than being “absolute,” has two limits. Because of these limits, it is—I’ll argue in this paper’s final section—quite possible that some public health paternalism by the state is all things considered justified, despite its infringing the autonomy rights of some in the group it targets. If state paternalism is easier to justify when it uses manipulative rather than coercive means, the reasons why that is so are largely utilitarian.

### Manipulation, Coercion, and Compulsion

I suppose that the concept of manipulation is purely descriptive, like the concepts of deception and paternalism and unlike the concepts of murder and theft. If manipulation is *pro tanto* wrong, that must be shown *via* an appeal to the best, most defensible moral theory, not *via* conceptual analysis.

Three conditions seem necessary for most if not all instances of manipulation. First, like coercion it is an *intentional* attempt to influence another’s choices *via* “non-rational” means rather than “rational persuasion” or a plea for a favor. A coercive threat to cut off your pinky finger unless you quit smoking is

---

<sup>1</sup> Thaler, R. and Sunstein, C (2008) and (2003).

<sup>2</sup> Libertarian critiques include Crampton, E. (2009), Glaeser, E. (2006), Gordon, D. (2008), Mitchell, G. (2004), Rizzo, M. and Whitman, D.G. (2009), Sugden, R. (2008), and Wilkinson, W. (2008).

“non-rational” inasmuch as it does not present you with any of the independently valid, moral and prudential reasons for quitting. Second, unlike coercion, manipulation does not limit one’s options or reduce one’s negative liberty.<sup>3</sup> Two clear examples of manipulation are deception (intentionally causing others to have beliefs that one knows are false) and “withholding” information or letting another act on a mistaken belief (e.g. about the efficacy of his chosen means to his ends) instead of warning him of his mistake. If I bamboozle you into believing that you will die of cancer unless you buy my elixir when in fact you are not sick at all, I haven’t reduced your negative liberty because you still have the option of holding on to your money and not dying. It’s just that you don’t know you have it.

Examples of attempts at manipulation that don’t involve deception or withholding information include:

- “Persuasive advertising,” such as the Marlboro ads that featured a rugged, macho cowboy smoking that brand of cigarettes. The ads increased the brand’s appeal among male smokers either by exploiting a fallacy in causal reasoning (“that manly male smokes Marlboros; he’s manly because he smokes them; thus, my smoking them will make me manlier”) or by *conditioning* viewers to associate the brand with manliness.
- Insincere flattery given to a vain or insecure person to curry favor with him.
- The manager of a non-profit cafeteria arranges the various food options in an order that encourages diners to select the more healthful ones.<sup>4</sup>
- A female college student wears an especially tight fitting, low cut blouse to show off her ample cleavage during meetings with male professors in which she appeals for higher grades.
- Creating a temptation to backslide in someone known to have a weak will.

The manipulator’s intention in all of these examples is to elicit a certain choice by *exploiting* the target’s *ignorance and/or irrationality*. This, I think, gives us a third necessary condition for manipulation: the means used exploit a deficiency and would be wholly ineffective on someone who lacks the deficiency—that is, someone who is well informed, in perfect self-control, and has full economic and epistemic rationality.<sup>5</sup> Creating a temptation to backslide is manipulative only

---

<sup>3</sup> See Raz, J, (1986), pp. 377-8. Thaler/Sunstein also assume that “nudges” are non-coercive or only minimally coercive. In this paper I use “nudge” and “manipulate” interchangeably even though some of Thaler/Sunstein’s examples of nudges are not manipulative (e.g. health warnings on cigarettes, requiring employers to inform employees of workplace dangers), and nudging for them needn’t be intentional. Assuming that nudges can be unintentional allows them to claim that they are ubiquitous, unavoidable, and thus (since “ought implies can”) morally unproblematic. For criticism of that move, see Coons, C. and Weber, M. (2013), pp. 16-17 and Hausman/Walsh (2010), p. 133.

<sup>4</sup> Thaler, R. and Sunstein, C. (2008), pp. 1-3.

<sup>5</sup> By “economic” rationality I mean what Derek Parfit called the “present aim” theory of rationality. Economically rational people maximize the weighted sum of their expected present preference satisfaction (values being a kind of

because the ideally rational person never succumbs to *akrasia*. Iago can manipulate Othello into believing that Desdemona's innocent actions are evidence of infidelity only because Othello's extreme jealousy makes him epistemically unreasonable.

Suppose that we offer a smoker a large, cash payment to quit because we think that while there are much stronger, health-related reasons for him to do so, citing them won't move him; only the offer of an immediate and tangible benefit will. Because we are spurning a rational persuasion that cites those stronger, health-related reasons, our offer is an attempt to influence his choices via "non-rational" means. Since it also does not reduce his options, it satisfies the first two conditions in our account of manipulation. But incentivizing good choice making is not manipulative (at least if not intended to create a temptation to backslide) because it does not satisfy our third condition. Well-informed and economically rational persons are responsive to incentives.

Deliberation and the gathering of information are activities that themselves have costs, and with trivial choices (as well as some urgent life-or-death choices where immediate action is required) the costs of engaging in them may outweigh the benefits. That is why it is rational for creatures with limited cognitive powers like us to rely on intuition and heuristics in much of our decision-making. The first and third conditions for manipulation should be understood in a way that's consistent with this point. Thus, in cases where absorbing information or assessing arguments through some time consuming ratiocinative process would have high costs, and going with one's gut feelings would probably produce a better choice, it is not manipulative for others to encourage one's use of the short cut.<sup>6</sup>

It is sometimes claimed that manipulation has to occur "behind one's back" while coercion is transparent. That is doubly mistaken. Coercion that operates by way of threats may be transparent, but coercion needn't employ threats. The state can coerce a smoker of limited means into smoking less by imposing a steep "sin tax" on cigarettes, with the smoker not realizing that their prohibitively high price is due to the tax rather than market forces. And while successful deception requires ignorance of the deceiver's intention, creating a temptation to backslide can be effective even when the target knows what the manipulator is up to.

Manipulation should not be confused with psychological compulsion, which closes off options. Suppose that I subject you to a drug/brainwashing regimen (think *A Clockwork Orange* or *The Manchurian Candidate* here) that causes in you a craving for some good. If the craving is so strong as to be literally "irresistible," then the compulsion is obvious. But even if it's not that strong—you could resist it with enough effort if you had strong enough reason

---

preference), given their beliefs about the world. Note that "hyperbolic discounting" need not be irrational on the "present aim" theory.

<sup>6</sup> Wilkinson T. (2013) p. 350 seems to reject what I've claimed is the third condition for manipulation because he supposes that it ignores the costs of acquiring information and then deliberating.

to—there is something analogous to the imposition of a tax here, since I’ve increased the psychological costs to you of not acting on the craving. Both manipulation and psychological compulsion are incompatible with full autonomy, and that may explain why it is easy to conflate them. But one closes off (or increases the costs attached to) options, while the other doesn’t.

Some view “persuasive advertising” as a form of psychological compulsion that uses “subliminal suggestion.”<sup>7</sup> “Subliminal suggestion” in the original sense of the term—movie theaters projecting popcorn images so quickly onto a screen that they are only unconsciously noticed yet still manage to induce a craving for popcorn in viewers—is pure urban legend.<sup>8</sup> But if “subliminal suggestion” (whatever it is) were real and akin to brainwashing, then it would be compulsive rather than manipulative. In that case it would not count as a “nudge,” and Thaler/Sunstein needn’t worry about finding other grounds (e.g. “publicity”) to rule out using it in the paternalism that they favor.<sup>9</sup>

Manipulative advertising creates a mental association at an unconscious level between the product and something else that’s desired but is in reality causally unrelated. The mere fact that the process is unconscious is not enough to absolve those exposed to it of all responsibility for being manipulated. The male professor who is more receptive to grade appeals from female students wielding their feminine charms is surely blameworthy for his unfair grade decisions. He may be unaware of his manipulation, but he *ought* to have been on his guard against it. Can we reasonably expect all consumers to maintain a vigilance that would prevent *any* manipulative advertising from *ever* being effective on them? Surely not.<sup>10</sup> Many of the “cognitive biases” that leave us vulnerable to manipulation—status-quo bias, risk aversion, loss aversion (the endowment effect), framing effects, anchoring effects, and so forth—are “hard wired” into the automatic, “fast” thinking or “system 1” part of our cognitive apparatus.<sup>11</sup> Generally speaking we are harder to manipulate when we engage in “slow” or “system 2” thinking. Given our cognitive limitations, we simply cannot engage in system 2 thinking prior to all of our economic and lifestyle choices. At most we can be expected to do so—and be blamed for not doing so—with the more consequential ones.

### The wrongness of manipulation; paternalism defined; possible objections to paternalism

What is the normative status of manipulation? Is it *pro tanto* wrong? An act type is *pro tanto* wrong if there is always a reason not to do it, a value or good

---

<sup>7</sup> Haworth, L (1984), esp. pp. 6-9 and Crisp, R. (1987).

<sup>8</sup> For the origins of the legend in Vance Packard’s 1950’s work of pop sociology *The Hidden Persuaders*, see Rogers, S. (1992).

<sup>9</sup> For their worries, see Thaler, R. and Sunstein, C. (2008), pp. 247-9. Hausman D. and Welch B. (2010) pp. 131-2 argues that they have no grounds on which to rule out this method of influencing choice.

<sup>10</sup> Benn, S. (1967) p. 267 seems to suppose otherwise.

<sup>11</sup> See Kahneman, D. (2011).

that its performance would sacrifice. When a person has been manipulated or coerced her rational autonomy is compromised. One is rationally autonomous to the extent that one's choices accord with and are guided by reasons that one does or would count as good, justifying ones. When I influence your choices via open and honest rational persuasion, they reflect your rational autonomy. When I deceive you, withhold relevant information from you, blackmail you, or take advantage of your weak will, gullibility, or emotional insecurities, I get you to make choices that you yourself would not see as having good reason to make, if you were well informed, free from duress, and fully rational; thus, I hinder or reduce your rational autonomy. Since rational autonomy is a good, manipulation is *pro tanto* wrong.

Of course the *pro tanto* wrongness of manipulation is consistent with its being all things considered justified in many cases. After all, there are other goods besides rational autonomy, and sometimes safeguarding them may justify some sacrifice of a person's rational autonomy. An obvious example is lying to the Gestapo who inquire whether one knows the whereabouts of any Jews in hiding. A harder case is the use of a manipulative PR campaign to reduce littering or increase voting in the general public. Supposing that such a campaign were more effective than any effort based on education and a straightforward appeal to civic duty, one may still doubt whether the sacrifice of the public's rational autonomy for increases in these social goods is justified.<sup>12</sup>

Important and difficult as questions of this kind are, my concern in this paper is with whether a person's welfare is ever a good of sufficient value, that it is sometimes right to compromise his rational autonomy to safeguard it. I take paternalism in the "strict" sense to involve the use of "non-rational" means (i.e. means other than rational persuasion) to hinder someone's already made choices, preempt the choice that he would make if he could make one (e.g. giving an unconscious Jehovah's Witness a life-saving blood transfusion), or influence his prospective choices, *for his own prudential good*.<sup>13</sup> These means include not just compulsion, coercion, deception, and non-deceptive manipulation, but offering incentives and creating new options (e.g. giving a homeless person in kind benefits rather than cash, to avoid subsidizing drug or alcohol consumption). Is the paternalistic coercion or manipulation of competent adults (who alone are presumed to possess sufficient capacity for rational autonomy) ever justified? If so, when?

---

<sup>12</sup> I note that Thaler/Sunstein laud the Texas anti-littering campaign that made an appeal to civic pride ("Don't Mess With Texas!") as an effective, non-coercive nudge. But if pride in Texas is something that the majority of Texans accept as a good reason to avoid littering in their state, the campaign did not compromise their rational autonomy. It merely draws one's attention in a vivid and effective way to a good reason not to litter.

<sup>13</sup> A "looser" sense of paternalism would replace "for his own prudential good" with "for his own good," where the latter means "for his own prudential good *or* to advance his own conception of the good."

In the discussion that follows, I will revert often to the following two examples of paternalism:

*Brother Francis* Biomedical researchers wish to test a new drug with the potential to save thousands of human lives. The test will kill many lab rats, but the researchers will spare the rats if Brother Francis agrees to serve as the guinea pig in their place. He believes that he ought to agree to this, because while one person's life has more value than one rat life, twenty or so rat lives have more value than one person's life. A friend or loved one knows that Brother Francis has a mother complex, and saying "your mommy would be ashamed of you if you volunteer!" will manipulate him into choosing not to do so, contrary to his own view of what he ought to do. The friend applies this nudge to him, for his own good.<sup>14</sup>

*Young Smokers* Vincent is a 20-ish adult of moderate means who has recently started smoking. He is not yet a "nicotine addict"; quitting would take some effort, but he could do it fairly costlessly. He knows that if he continues his one pack a day habit, he runs the risk of various smoking-related illnesses and reduced longevity. But he values the pleasures of smoking now and in the immediate future—in solitude but especially when socializing with friends who also smoke—more than his long term health. Thus, he has no wish to quit. However, recent increases in state taxes on cigarettes force him to choose between smoking and several hundred dollars per month. As a result of the state's coercion, with much resentment, Vincent and many other smokers like him choose to quit.

In the first example the paternalist is a private person who targets a single individual, while in the second the paternalist is the state whose target is a class of persons. In both cases the opportunity for rational persuasion exists and the paternalist either resorts to "non-rational" means after persuasion has been attempted but failed (the state's taxation of smokers like Vincent) or (as with *Brother Francis's* friend) believes that since it is so likely to fail it is best not even to try it.

What explains the *pro tanto* wrongness of the paternalism in these two examples? Is the paternalism in them wrong not just *pro tanto* but all things considered? We can distinguish a few possible ways of defending an affirmative answer to latter question, including:

- i) The "rational autonomy as a good" objection: the paternalism in our two examples leaves its targets on balance worse off, because for everyone the value of rational autonomy trumps the value of goods like health and longevity.
- ii) The "subjectivity of well-being" objection: the paternalism leaves Brother Francis and Vincent on balance worse off, because health and longevity have for them (owing to their preferences) much less value than they have for others. The paternalist assumes an indefensible

---

<sup>14</sup> The example (modified to include the nudge) comes from VanDeVeer, D. (1986), p. 126.

“objective list” theory of the good, which fails to recognize the extent to which “one’s good” is a function of one’s preferences or “conception of the good.”

- iii) The “egalitarian objection”: even if the interference benefits them on balance, it wrongs them by treating them as inferior sorts of person. Indeed, even some paternalism that does not use either coercion or manipulation, and thus, does not compromise its target’s rational autonomy, is wrong for this reason.
- iv) The “right to autonomy” objection: even if the paternalism does benefit them, it wrongs them by interfering with their ability to make choices that reflect their own “conception of the good.” In doing that it infringes their right to personal sovereignty or autonomy. The duty to respect this right of competent adults trumps the duty to promote their welfare.

I shall argue that all of these views are mistaken. iv) is correct insofar as it identifies the wrongness-making feature in the two cases as the infringement of the right, but incorrect if it insists on the right’s absoluteness. I’ll then propose a couple of limits on the right that allow us to condemn the paternalism in *Brother Francis* as all things considered wrong but also allows for the possibility that the state paternalism in *Young Smokers* is justified.

#### The “autonomy as a good” objection

Given the above definition of paternalism, there are instances of it that do not compromise their targets’ capacities for rational autonomy, and thus, instances to which this objection is inapplicable. Rational autonomy seems to us an important value only when some threshold capacity for it has been surpassed, as we think it has been with only competent adults, not small children, the cognitively disabled, paranoid schizophrenics, and so on. Indeed, even young teens are thought to fall short of the threshold when it comes to many important life choices, such as whether to withdraw from school or take up smoking; paternalism toward them regarding momentous matters of this sort seems not just permissible must an obligatory part of responsible parenting.

Even paternalism toward adults whose capacity for rational autonomy ordinarily surpasses the threshold involves no real sacrifice of their rational autonomy if they are temporarily impaired by extreme fatigue or emotion or the influence of drugs or alcohols, etc. or they are acting in ignorance of important relevant facts and there is no opportunity to warn or inform them. An example that illustrates such ignorance is:

*Pedestrian Shove* A pedestrian walks into the path of a speeding car, which he doesn’t see because he’s distracted by text messages on his phone. (He *chose* to cross the street but did so carelessly). At the last moment a good Samaritan shoves him out of harm’s way.

The paternalism here and in cases involving incompetents is what Joel Feinberg called “soft” paternalism, because it interferes with substantially non-voluntary choices.

Another example of paternalism that does not compromise or diminish the rational autonomy of its target is:

*Incentivizing non-use of drugs* A son lives at home with his father while attending college. Despite the son's assurances that he is "not into drugs," the father worries that the son, while sincere, may lack the willpower needed to resist peer pressure when socializing with drug-using friends. To provide his son with some extra motivation, father offers to pay him \$50 for each week that he's drug free. (Suppose father can tell whether son has used just by carefully examining the pupils of his eyes). Son welcomes the offer, since it pays him to do what he intends to do anyway.

Since the father's offer does not compromise or impair his son's capacity for rational autonomy, the "autonomy as a good" objection does not apply to it.

But Brother Francis and Vincent are competent adults with unimpaired capacities for rational autonomy. The manipulation of one and coercion of the other diminishes their rational autonomy. If rational autonomy has as much value as the "autonomy as a good" objection alleges, then it leaves them worse off. If that's right, then the paternalism is surely all things considered and not just *pro tanto* wrong.

But the objection surely exaggerates the value of rational autonomy as a good. If the slightest loss of autonomy outweighed the largest gain in other goods, it would follow that you are better off continuing to believe in your religion than being manipulated into converting to mine, even if mine is the true religion sincere belief in which is necessary for the salvation of your soul. The right objection to my manipulation is not that it can't *possibly* benefit you, but rather that, because religious skepticism is true, I'm not warranted in my belief that there are immortal souls and a God who eternally damns the ones had by people who reject my theological beliefs.

Perhaps the defender of this objection needn't make a claim on behalf of rational autonomy's value as strong as lexical priority; perhaps it is enough if rational autonomy is only a *very* valuable good. But there is a deeper problem with the objection: it is *myopic*. If the value of rational autonomy lies in its being a good, prudential or other, then more of it is better than less not just interpersonally but intrapersonally and across time. Thus, even if every manipulative or coercive interference with another's "self-regarding" (see below) choices does compromise her *present* rational autonomy, that won't be dispositive. Suppose that I threaten you with punishment if you try heroin, and as a result you grudgingly choose not to try it, but also that if you tried it you would have become slavishly addicted to it. In that case while my coercion reduced your *present* autonomy it also prevented a large loss of your *future* autonomy, and thus, may have increased your over-all present and future autonomy. Indeed since it is very difficult to exercise rational autonomy if one is dead or severely ill, it would seem that many interferences with choices that pose an undue risk of either may also maximize rational autonomy over time. The elevation of rational autonomy as an extremely valuable good turns out not to pose the obstacle to the justification of paternalism that it may have seemed to.

Finally, does all manipulative or coercive interference with one's self-regarding choices have to reduce one's present rational autonomy? Suppose

that you have powerful desires that sometimes lead you to act in ways inconsistent with your rational autonomy. Then any coercion or manipulation that had the effect of eliminating those desires would increase your rational autonomy.<sup>15</sup> One way of canceling them would be to create desires of the same strength but opposite valence. If the choices of many to consume copious amounts of alcohol and junk food are due in part to manipulative advertising by the industries that produce and market them, then the current objection should permit quite a bit of state manipulation aimed at encouraging healthier lifestyle choices, to cancel the autonomy-reducing effects of that prior manipulation.<sup>16</sup>

### Two senses of “one’s good”; preferentism and subjectivism

To see what’s wrong with the second objection, it will help to draw some distinctions. Millian defenders of the “harm principle” claim that society has no business interfering with “self-regarding” choices. Many have noted that if “self-regarding” means “no significant effects on the welfare or interests of others” then only fairly trivial choices (e.g. should I shave this morning or not?) will be self-regarding, rendering the harm principle toothless. What defenders of the harm principle need is a normative conception of “self-regardingness”: a choice is self-regarding if and only if it does not wrong others by wrongly harming them or wrongly posing a risk of harm to them. Both Brother Francis’s choice to volunteer to be the guinea pig and Vincent’s choice to continue to smoke are “self-regarding” in this sense.

Using this definition of “self-regarding,” we can draw the first distinction we’ll need between two senses of “one’s good”: i) one’s prudentially best choice in the circumstances (i.e. the choice one would make if one were a rational egoist, concerned only with one’s own welfare or interests), and ii) the all things considered best self-regarding choice that one could make in the circumstances. (Hereafter I’ll call i) one’s “prudential value” or PV good, and ii) one’s “all things considered” or ATC good).<sup>17</sup> That one’s PV and ATC good might differ is evident

---

<sup>15</sup> This is the purpose of mandatory “cooling off periods,” which Hausman and Welch (2010) find them acceptable because they “merely counteract foibles in decision-making without in any way pushing individuals to choose one alternative rather than another.” p. 132.

<sup>16</sup> Hausman and Welch (2010) claim that “what makes the cacophony of invocations of irrational responses by non-governmental agents tolerable (to the extent that it is tolerable) are, we suggest, the limits to its effectiveness and the extent to which these invocations conflict with ... and cancel one another” p. 131. Their claim is tenable where there are many manipulators with conflicting aims and roughly equal resources. But whether those conditions are met in any capitalist society seems doubtful.

<sup>17</sup> Note that one’s PV good is “prior” to the morally right in the sense that we don’t need to have a theory of right/wrong to identify one’s PV good. So it could turn out, as the amoralist maintains, that much wrongdoing serves one’s PV good. By contrast, one’s ATC good is not “prior” to the morally right, since our definition of it used a right-based conception of “self-regarding.”

in the case of Brother Francis, who admits that volunteering is contrary to his PV good but insists that it is required by his ATC good. The paternalist claims that Brother Francis has a mistaken belief about his ATC good. The mistake is not in thinking that it is sometimes best to sacrifice one's PV good for other values. Surely if the harm Brother Francis had to suffer were much smaller—the equivalent, say, of a punch on the arm that is painful for a few seconds and leaves a bruise but does no more damage—and suffering were necessary to prevent the painful deaths of hundreds of dogs, no sane paternalist would judge such self-sacrifice foolish. Brother Francis's belief is mistaken, according to the paternalist, because the good achieved by his sacrificing his life is not large enough to justify the extreme sacrifice of his very life.

One of J.S. Mill's anti-paternalism arguments was that paternalists' judgments about what's best for competent adults who are well-informed about their options are usually mistaken, because such adults are usually the "best judge" of their own "good." The "best judge" principle comes in two versions, one about one's PV good, the other about one's ATC good. Mill supposed that the PV version was correct, at least as a "presumption."

Another distinction that we'll need to evaluate ii) is between preferentism as a theory of one's PV good and "subjectivism" as a theory of one's ATC good. Let's begin with preferentism, the idea that one's prudential good is a function of one's level of preference or desire satisfaction. Preferentism has two main rivals: hedonism and "objective list" theories of welfare. (There are also possible theories of prudential value that combine features of any two or even all three of these). I assume that any theory of prudential value is temporally neutral regarding present and future: if something benefits you slightly in the short run but harms you greatly in the long run, then it is all thing considered contrary to your PV good. This means that preferentism has to count future and not just present preference satisfaction and frustration as affecting one's PV good. But which types of preferences should count? Here it has to be noted that to be remotely plausible as theory of one's PV good, preferentism has to exclude preferences that are insufficiently "personal" in some sense. A preferentist theory of prudential value that did not exclude "moralized," "external" (e.g. altruistic and envy-based), and "impersonal" preferences would have the absurd consequence that self-sacrifice is impossible, i.e. psychological egoism is true.<sup>18</sup>

---

<sup>18</sup> See the "success theory" in Parfit (1984), pp. 494-6, and Griffin J. (1986), who agrees that the "best account of utility makes it depend on some desires and not on others" p. 33. I should note that "remotely plausible" doesn't mean "correct." Whether any version of a preferentist theory of one's PV good is tenable, or all are embarrassed by the phenomenon of "adaptive preference formation" or refuted by Rawls' "bare persons" objection, is a question beyond the scope of this paper. Whether the phenomenon of indeterminate and/or unstable personal preferences affects the tenability of preferentism is not so clear. Perhaps preferentism's defenders can dig in their heels and insist that in those cases where someone's personal preferences really are indeterminate,

Thus, the theory should identify one's prudential good with whatever maximizes the satisfaction of one's present and future personal, intrinsic (rather than instrumental), "non-moralized" preferences, weighted according to their strength.

Subjectivism about one's ATC good is the view that one's ATC good consists in whatever maximizes one's present preference satisfaction, not excluding any of the preference types that the preferentist theory just described would exclude. Or to restate the idea in what seem to me nearly equivalent terms: one's ATC good is determined by one's currently held "conception of the good."<sup>19</sup> Note that subjectivism is not a skeptical view. It does not say that it is impossible for anyone to know where someone's ATC good lies. It is, however, a "comprehensive moral doctrine," which disqualifies it from figuring in the justification of principles governing the limits of state paternalism, according to those liberals who suppose that the principles can and should be "neutrally" justified.

Both subjectivism and skepticism about one's ATC good are untenable for many reasons. Perhaps the main one is that they render pointless any first person practical reflection about whether one's conception of the good is correct or defensible. But of the two, subjectivism is by far the less tenable. Consider the question whether the Jehovah's Witness who refuses a life-saving blood transfusion is making the best choice. Surely he is if there *really is* a God (omniscient and all-good) who forbids all transfusions. The skeptic's claim that nobody can know whether his choice is best, because no one can know whether such a God exists, makes clear sense. By contrast, the subjectivist's claim—that the JW's choice is best just in case he holds the belief in question and *regardless of whether it is true*—makes little sense, least of all to the JW who holds it.

#### Problems with the "subjectivity of welfare" objection

This objection like the previous says that any coercive or manipulative paternalism that targets competent adults who are well-informed about the options will fail to achieve its goal of making them better off. But for the previous one the reason why such paternalism won't achieve its goal is that it underestimates the value of rational autonomy, whereas for the present one the reason why it won't is that it overestimates how much value goods like health, longevity, and long term financial security have for everyone. This objection like the previous one is not supposed to tell against the paternalism in *Pedestrian Shove* or *Incentivizing non-use of drugs*.

Despite its air of plausibility and wide appeal (especially among libertarian-minded economists), the objection seems to me to be hopelessly muddled. In particular, it confuses one's PV good with one's ATC good, the two

---

there is no fact of the matter as to whether a paternalistic intervention benefits or harms him.

<sup>19</sup> Included in one's "conception of the good," I assume, are metaphysical and religious beliefs the truth or falsity of which cannot be established by ordinary empirical means.

corresponding versions of the “best judge” principle, and preferentism as a theory of one’s PV good with subjectivism as a theory of one’s ATC good.

The confusion becomes apparent when we ask how the objection could apply to the paternalism in *Brother Francis*. In that case the paternalist and Brother Francis disagree not on where his PV good lies but rather on where his ATC good lies. Anyone who presses the objection to this paternalism has to appeal to the version of the “best judge” principle about one’s ATC good. The problem with that version of the principle is that it presupposes subjectivism about one’s ATC good, which as we’ve seen is absurd.

The objection fares no better when leveled against the paternalism in *Young Smokers*. I suppose that in this case the disagreement between the paternalist and Vincent is over the question whether smoking is consistent with or contrary to his PV good. The defender of the objection is alleging that preferentism is the correct account of prudential value, it supports a version of the “best judge” principle about one’s PV good, and thus, a paternalism that forces Vincent to quit probably harms him on balance. The problem with paternalists, according to this line of thinking, is that they assume “healthism,” the view that improved health is always better for all of us.<sup>20</sup> “Healthism” might be consistent with an objective list theory of prudential value, but it is inconsistent with preferentism. The problem with defenders of objective list theories is that they disrespect other persons by not deferring to their own judgments about what’s good for them.

The reply to this argument is that the paternalism in *Young Smokers* is entirely consistent with the preferentist theory of prudential value described in the previous section; it simply is not true that a defender of the paternalism here must assume an objective list theory. While it certainly denies that health has lexical priority over other prudential goods, the preferentism described in the previous section implies that any 20-ish adult who judges smoking to be in his best interests is probably mistaken, because the amount self-regarding preference satisfaction his smoking provides him in the present is probably exceeded by the amount that it deprives him of in the future.

Of course, Vincent’s choice to continue smoking, even where quitting is an available, low cost option, need not be *irrational*, since what’s best for him *in the long run* might not matter to him all that much in the present. But the question whether a self-regarding choice is economically rational is separate from the question whether it is prudentially good for the person making it. Even if Vincent’s indifference to his distant future well-being is not irrational (i.e. a Humean or “present aim” theory of practical rationality is correct), it remains the case that very unhealthy lifestyle choices are probably on balance bad for him, causing as they do significant and foreseeable harm to self in the future. The

---

<sup>20</sup> See Crampton, E. (2009), from whom I borrow the term “healthism.” Crampton’s “healthists” demand that the state attach *lexical priority* to health over *all* other goods. Mitchell G. (2004) seems to me to defend this objection. He refers to the view that “individuals are the best judges and protectors of their own welfare” as a “key libertarian assumption.” (p. 1249)

judgment that smoking is in the long run bad for the *average* 20-ish adult seems *clearly* true no matter which account of prudential value—preferentism, hedonism, an objective list theory, or some hybrid combination of them—turns out to be best. It is compatible with limited commensurability among some prudential values, such as the pleasures of smoking vs. a longer life, but not full incommensurability among all prudential values.<sup>21</sup> It is also compatible with the possibility that smoking and other unhealthful lifestyle choices may be in the best interests of a minority—e.g. those tortured artists and bohemian intellectuals whose creative output would wane if they adopted a more staid, bourgeois lifestyle.

### The egalitarian objection

Defenders of this objection include Seana Shiffrin and Jonathan Quong.<sup>22</sup> According to Shiffrin, paternalism represents an “affirmative disvaluing of the agent”, “a positive... effort by another to insert her will and have it exert control merely because of its [perhaps only alleged] superiority.”<sup>23</sup> Quong defends a slightly different definition of “paternalism,” but one that requires a similar motivation on the part of the paternalist: she must make a “negative judgment about [the target’s] ability (assuming [he] has the relevant information) to make the right decision or manage the particular situation in a way that will effectively advance [his] welfare, good, happiness, needs, interests, or values.”<sup>24</sup> This motivation, according to Shiffrin and Quong, disrespects the target by impeaching his moral standing. Since a motivation of this sort is probably absent from the good Samaritan in *Pedestrian Shove*, Shiffrin and Quong would probably deny that it’s a case of paternalism at all, *contra* the definition of paternalism I gave earlier. But they would certainly hold that the motivation is present in the paternalists in *Brother Francis* and *Young Smokers*. The paternalistic friend of Brother Francis, who would use manipulation rather than attempt to persuade him that his view of the relative value of rat and person life is mistaken, impugns or belittles his *capacity* to form, revise, and act on a

---

<sup>21</sup> For an appeal to what looks like full incommensurability to rebut the paternalist’s claim that “the badness of ill health outweighs ... the goodness of the pleasure of smoking,” see Gaus, G (2003), p. 22.

<sup>22</sup> Anderson, E. (1999) seems to be another supporter of it. She supposes that people ought to feel insulted if paternalism is offered as the rationale for mandatory participation in social insurance programs but not for “trifling” restrictions of liberty like seat belt laws (p. 302).

<sup>23</sup> Shiffrin, S. (2000), p. 220.

<sup>24</sup> Quong J (2011), p. 80. Though both require a similar motivation, the last part of Quong’s definition makes it narrower than Shiffrin’s, which counts as paternalism any interference in “matters that lie legitimately within [the target’s] control” motivated by the belief that one’s judgment or agency are superior to his. Quong (pp. 79-80) is rightly critical of this feature of Shiffrin’s definition, which makes it overinclusive.

conception of the good (what Rawls called the “second power of moral personality”).

Indeed, the egalitarian objection if sound would apply to *Incentivizing non-use of drugs* too. Father may not harbor doubts about son’s ability to identify his ATC good, but he does harbor doubts about his ability consistently to pursue it. In making the offer he wrongly “infantilizes” his son. Thus, the objection has a wider range of application than the previous two, which applied only to paternalism that interferes with one’s ability to make or implement one’s own decisions.

Notice that the objection does *not* apply to coercion or manipulation motivated by pure self-interest rather than a judgment about how the target ought to choose. While the businessman who uses “persuasive advertising” to increase his sales may be guilty of treating his customers as “mere means,” he does not take himself to be a better judge of what they ought to do than they are.<sup>25</sup> Finally, the objection, at least as understood by the non-libertarian liberals who endorse it, is not supposed to apply to preventing the denigration of others—treating them as if they were one’s inferiors—*with their consent*. Alleged examples of “consensual moral denigration” include voluntary slavery and maybe prostitution and “dwarf tossing.”<sup>26</sup> If we prevent voluntary slavery arrangements for paternalistic reasons, we are (according to these defenders of the objection) affirming our moral superiority to others. But if our motive is to uphold the inalienably equal moral status all persons, we do not leave ourselves open to the charge of elitism.

The fundamental problem with the egalitarian objection is that it confuses agent and act evaluation. Evaluating agents as blameworthy or not is a separate matter from evaluating their actions as justified or not. I concede that paternalists are blameworthy when they harbor elitist attitudes about the abilities of their targets to identify and act on the good. Many males in patriarchal societies support restrictions on the liberty of women for paternalistic reasons, in the sincere belief that they serve the best interests of women. The restrictions are thought to “protect” females from the bad consequences of their lesser ability to reason (e.g. Aristotle thought that they possess “passive” but not “active” intellect) or to control themselves and their emotions (“hysteria” and so on). I agree that when paternalists are motivated by attitudes of this sort toward women, racial minorities, the disabled, etc., they disrespect their intended beneficiaries and are blameworthy. But I deny that a patronizing or condescending attitude is an essential part of the motivation of any paternalist.

---

<sup>25</sup> For the argument that the Kantian Principle of Humanity implies that all manipulation is *pro tanto* wrong, see Korsgaard, C. (1986), esp. p. 336. Though the Kantian principle has wider application than the egalitarian objection, the objection might be defended on the basis of the principle.

<sup>26</sup> For slavery contracts see Anderson, E (1999), p. 319. Shiffrin (2000) rejects a paternalistic rationale for the unconscionability doctrine in contract law in favor of a nonpaternalistic rationale that is supposed to imply moral disapproval of the debaser but not of the party consenting to be debased.

The egalitarian objection rightly condemns the motives of *some* paternalists, but that condemnation is irrelevant to whether paternalism *per se* is *pro tanto* wrong and if so, why.

Consider again the paternalism in *Incentivizing non-use of drugs*. Defenders of the egalitarian objection might say that the paternalism there is somewhat less objectionable than the paternalism in *Brother Francis*, because the claim that someone's volitional capacities don't pass muster is less denigrating than the claim that his cognitive capacities don't.<sup>27</sup> But does the father in this example have to believe that his own volitional capacities are superior to his son's? Maybe he does believe that it would be easier for him to refuse the offer of drugs at a rave than for his son to do so. But that belief is consistent with the humble belief that with regard to *most* other matters his son's powers of self-control are equal to or even greater than his own. It might be urged that a father who makes the offer while holding that humble belief cannot plausibly be accused of "infantilizing" his son.<sup>28</sup>

A reply to this is that it is possible to disrespect others in ways that don't suggest one's superiority to them.<sup>29</sup> Telling a woman that her "function" is biological reproduction, so she should stay at home to raise her children and not pursue any career outside of it, disrespects her, and the disrespect remains if it is another woman who is telling her this. Or consider an oncologist who defends her medical paternalism on the grounds that ordinary adults, after learning that they have malignant cancer and being informed of their treatment options, are simply incapable of making a rational choice due to the strong emotions that are likely to feel; the incapacity remains even weeks after receiving the bad news. This physician disrespects her patients if she makes the choices for them without their informed consent, and the disrespect remains even if she would include herself within the scope of her generalization. Isn't the same true of the father in *Incentivizing non-use of drugs* who holds the humble belief?

I think not. A patient of the oncologist ought to feel insulted upon learning of the paternalism; not feeling insulted would betray the moral vice of insufficient self-respect.<sup>30</sup> If the father's offer were "non-comparatively" disrespectful, then the son too *ought* to feel insulted even if the father has the humble belief described above; not feeling insulted would reveal a moral failing on his part. But do we really think that the son must be culpably deficient in self-respect if he's not offended by the offer? It seems to me that we don't. The difference between the two cases is that oncologist's view of the autonomy capacities of ordinary

---

<sup>27</sup> Michael Cholbi defends such a view in his "Paternalism and Our Rational Powers" (manuscript).

<sup>28</sup> DeMarneffe, P. (2006), p. 80 makes a similar point when he notes that that while the judgment that you are prone to certain defects of practical rationality might wrong you, the judgment that we are all prone to such defects needn't.

<sup>29</sup> In this connection Quong, p. 101 distinguishes a "comparative" and "noncomparative" wrong.

<sup>30</sup> Hill, T. (1973).

adults seems unduly pessimistic, while the father's doubts about his son's ability to control himself when friends offer him drugs may well be warranted on the basis of past experience.

The *pro tanto* wrongness that Shiffrin/Quong claim to find in all paternalism need not be present in *Incentivizing non-use of drugs*, and it is certainly not present in forbidding young teens to smoke, overruling the drunk who thinks he's sober enough to drive home safely, and many other instances of "soft" paternalism. Perhaps, however, it is present in all instances of "hard" paternalism. In *Brother Francis* the paternalist is supposed to be vulnerable to the charge of elitism because she holds that Brother Francis has made an evaluative mistake *and* since there is little hope of convincing him of this *via* purely rational argument, the use of non-rational means to stop him from volunteering is justified. Our paternalist might believe that she would succeed in convincing Brother Francis to change his mind if there were time for it, but there isn't so that's why she must resort to manipulation. Or she might believe that any attempt at rational persuasion would probably fail even if there were ample time for it. The egalitarian objection seems most promising in the latter case, so that's the one we should consider. Perhaps our paternalist believes that Brother Francis is so heavily psychologically invested in his current views about the relative value of rat life to person life as to make him close-minded and thus unable to appreciate the force of strong objections to them. If she believes this, isn't she assuming that he possesses a lesser capacity for practical rationality than she does—leaving her open to the egalitarian objection?

Not necessarily. Our paternalist might admit that all human beings, herself included, are prone to the foible of under-appreciating the force of objections to theories in which they are heavily invested. But even supposing she does believe that her capacity for practical rationality is on whole greater than Brother Francis's, such a belief need *not* be objectionably elitist. It won't be so long as she admits that Brother Francis's capacities surpass the threshold necessary for moral personhood. Everyone whose capacities surpass it is equally a moral person with a right to autonomy (as well as to own property, seek/refuse medical treatment, enter binding contracts, vote, etc). But that doesn't mean that every person's capacities are equal; some may surpass the threshold by a much wider margin than others. Plainly some people are better able to spot errors in statistical and causal reasoning, distinguish good from bad analogies, avoid formal logical fallacies, etc. than others. More importantly, some persons are much more willing to examine critically their own evaluative beliefs and admit to errors or inconsistencies in them than others are. The paternalist who claims that her capacity for practical rationality is greater than her target's need *not* claim that her target's capacity falls below the minimum threshold—as does the capacity of a child or deranged person. The main problem with the egalitarian objection to paternalism is that it confuses these different types of claim.

### The "right to autonomy" objection

Whereas the first objection appealed to rational autonomy as a good, this one appeals to some account of autonomy as a right. The right, whatever it is, can only be respected or infringed by others; one can't infringe one's own rights. The good, by contrast, can be promoted or compromised by oneself or others. If I get very drunk or have myself locked in a room for a day, I do not infringe my right to autonomy but I do temporarily reduce my autonomy by impairing my capacity to make autonomous choices or limiting my ability to execute them.

The objection alleges that the paternalism in *Brother Francis* and *Young Smokers* wrongs its targets by infringing one of their rights. The right is supposed to protect only "self-regarding" (in the sense spelled out above) choices, to be had only by competent adults, and to be "negative" rather than "positive." The claim that the right is "negative" just means that to infringe it one must *interfere* with another's choices in some way. A refusal to expand his options or help him achieve his goals wouldn't infringe it. Paternalism is "hard" when it infringes the right, "soft" when it doesn't. Thus, according to this objection it is only hard paternalism that is *pro tanto* wrong, not all paternalism. Since the paternalism in *Pedestrian Shove* and *Incentivizing non-use of drugs* is soft, this objection finds no *pro tanto* wrongness in them.

Different theorists defend different views of the content and scope of the right, leading to some disagreement about which cases of paternalism are soft and which hard. On the standard libertarian view of the right, interference with the self-regarding choices of competent adults is permissible only if they previously gave or currently give *actual* consent to it.<sup>31</sup> Moreover, the only form of "interference" that could infringe the right is one that restricts liberty. Thus, noncoercive nudges cannot infringe the right; on some libertarian views not even deception can infringe it.<sup>32</sup> This means that for the libertarian paternalism is soft if it uses nudges and hard if it uses taxes or the threat of criminal sanctions. Use of the "soft"/"hard" terminology in this way by both Thaler/Sunstein and many of their libertarian critics reflects a tacit acceptance of this libertarian account of the right.<sup>33</sup> The libertarian account of the right to autonomy provides grounds for opposing manipulative state paternalism only if there are good grounds for fearing that state nudging would put us on a slippery slope leading to state

---

<sup>31</sup> This appears to be Nozick's view; it is present in Gordon, D. (1980). An example of a paternalistic restriction of liberty that this view permits is the Missouri law that excludes from casinos self-identified "gambling addicts" who've agreed to the exclusion to help them combat their addiction. Note that since the man in *Pedestrian Shove* never gives such consent, this view has the counter-intuitive implication that the shove infringes his right to autonomy. Vallentyne P. (2007) defends another supposedly libertarian account of the right intended to avoid this implication. On his view coercive interference infringes the right only if the target "validly dissents" to it.

<sup>32</sup> See Rothbard, M. (1985); Childs, J. (1994).

<sup>33</sup> These critics include Glaeser E. (2006) and Sugden R. (2008).

puches. Several libertarian writers seem to object to the Thaler/Sunstein program on this feeble basis.<sup>34</sup>

An alternative to the libertarian's account (as well as Feinberg's, which I won't discuss) is one that focuses on preferences rather than choices and on hypothetical rather than actual consent.<sup>35</sup> I'll call it the HC (for "hypothetical consent") account. It holds that whether interference with a competent adult's self-regarding choices infringes her autonomy all depends on whether she *would* consent to it, given her *presently held* preferences, values, and religious or metaphysical beliefs, if she were thinking clearly and had full, accurate, empirically ascertainable information about her situation. According to HC the paternalism in *Pedestrian Shove* is likely to be the morally less problematic "soft" variety because it is likely that the pedestrian has preferences such that if he were well informed about his circumstances, he would want others physically to prevent him from unintentionally placing his life in imminent and grave peril. If we knew that he is even more averse to others' interfering with any of his self-regarding choices than to placing himself in such peril, but we interfered anyway, then our paternalism would be hard. Thus, soft paternalism is justifiable by appeal to the target's own preferences and values, whereas hard paternalism "imposes" on its targets a conception of the good—of prudential value, virtue, religion, etc.—that they do not share.<sup>36</sup>

HC's account of the right differs from the libertarian's in another respect, namely, what it counts as "interference" that could infringe the right. At the very least it counts deception as a form of interference. Suppose that I lie to my friend about the personal qualities of a woman he's dating because I'm sure that they make a poor match but he won't realize this till after his fragile heart has been broken. Surely my paternalistic deception is at least *pro tanto* wrong because it infringes his autonomy; I'm using non-rational means to control his self-regarding choices in an area where it is unlikely that he would consent to my interference.

If it is the infringement of his right to autonomy that explains the wrongness of my deceiving or coercing him, then it seems natural to suppose that non-deceptive manipulation can also infringe the right and is *pro tanto* wrong

---

<sup>34</sup> Glaeser E. (2006), Wilkinson W. (2008), and Rizzo, M. and Whitman, D.G. (2009) all voice slippery slope worries about the use of nudges by the state. These writers should be reminded that they rightly reject similar slippery slope arguments based on speculative worries about future rights violations, offered by social conservatives in support of restrictions on liberty—e.g. that if physician-assisted suicide and voluntary euthanasia for the terminally ill are legalized, eventually we'll find ourselves condoning the involuntary killing of anyone with a disability.

<sup>35</sup> I criticize Feinberg's account in Scoccia (2013).

<sup>36</sup> Something like the HC account of the right is present in Dworkin, G. (1972), esp. his defense of seat belt laws, in Hodson, J. (1977), and in Vandever, D. (1986). The defense of some state paternalism in Rawls, J. (1971), pp. 248-9 that appeals to what we would consent to if rational and behind a veil of ignorance is also of this type.

when it does.<sup>37</sup> I take HC to imply as much. If in some case coercive interference with a self-regarding choice is wrong while manipulative interference isn't, or the reverse, this will be so either for consequentialist reasons or because the target would consent to the one but not the other.

It seems to me that an objection to the paternalism in *Brother Francis* and *Young Smokers* based on the HC account of the right has much intuitive appeal and explains the *pro tanto* wrongness present in those cases. Indeed, I suspect that many who press the first two objections considered earlier find them plausible only because they confuse them with this one.<sup>38</sup> This objection unlike those explains why paternalism can be *pro tanto* wrong even in some cases where it interferes with genuinely foolish self-regarding choices—ones that are contrary to both one's PV and ATC good. I suppose that defenders of a right to autonomy can remain, like Feinberg, agnostic about whether its justification lies in some indirect consequentialism, Rossian deontology, Kantian deontology, etc.<sup>39</sup> Certainly it should be possible to justify the right in a way that does *not* presuppose either subjectivism or skepticism about one's ATC good or the view that rational autonomy as a good trumps health, financial security, etc.

As an account of the moral demand that we respect others' autonomy, HC invites a few objections. One is that hypothetical consent lacks the "moral magic" that only actual consent has. Another is that when a person's conception of the good is the product of adaptive preference formation or an oppressive social environment, it is implausible to suppose that "autonomy" in any meaningful sense requires deference to it. I won't try to respond to either of these. I suspect that the first one is mistaken, while the second can be met with a minor amendment in HC's account of the right.

A more serious objection cites the fact of indeterminate and unstable people's preferences—a fact supposedly established by research in behavioral economics and social psychology. If this phenomenon were sufficiently pervasive, then there would often be no *fact of the matter* whether one would, if one were well-informed about relevant empirical matters and thinking clearly, consent to interference with one's choices, and thus, no fact of the matter whether it infringes one's autonomy right according to HC. If, say, a hybrid objective list/hedonistic account of prudential value is correct (a theory that

---

<sup>37</sup> Raz (1986), Hausman and Welch (2010), and Wilkinson T. (2013) agree that manipulation no less than coercion *can* reduce and/or violate another's autonomy. Wilkinson also says, "If it were consented to, in the right sort of way, then... manipulation would at least be consistent with autonomy and might count as enhancing it." p. 346

<sup>38</sup> E.g. maybe as the second objection alleges, the paternalist makes a mistake when she judges Vincent's conception of his PV good to be misguided. How does *that* "disrespect" him? Surely to disrespect him, it is *also* necessary that she coerce or manipulate him into making a different choice, so that she infringes his right to autonomy as conceived by HC.

<sup>39</sup> For Feinberg's unwillingness to take a stand on the deeper foundational questions of moral theory, see Feinberg, J. (1984), pp. 17-18.

seems to me at least as plausible as the preferentism described earlier), then the rationale for the paternalism in *Brother Francis* and *Young Smokers* would remain unscathed, and there would be no reason for thinking that it is even *pro tanto* wrong on account of its infringing the right to autonomy.

Fortunately, the phenomenon in question seems much overblown. Consider what is supposed to be an example of it: the fact that many cancer patients will opt for or against chemotherapy depending on whether the prognosis for those who undergo it is framed positively in terms of survival rates or negatively in terms of mortality rates, even though the two are logically equivalent.<sup>40</sup> This does *not* show that these patients lack “true” preferences with regard to chemotherapy vs. other forms of cancer treatment. It only shows that a bit more work is needed to determine what they are. The objection presupposes a behavioristic “revealed preferences” methodology that confuses choices, which are episodic and overt, with preferences, which are inner and dispositional.

### Limiting the strength of the right to autonomy and defending manipulative public health paternalism

I grant that paternalistic interference with Brother Francis’s decision to sacrifice himself is not just *pro tanto* but all things considered wrong, because the duty to respect his autonomy overrides beneficence or the duty to promote welfare. My aim in this final section is very briefly to sketch an account of the conditions under which hard paternalism is all things considered justified that supports this verdict about *Brother Francis* while allowing that some public health paternalism is justified.

Feinberg, Hodson, and VanDeveer hold that the right to autonomy is “absolute” vis-à-vis any interference intended to promote the right bearer’s own welfare.<sup>41</sup> It seems to me that liberals should reject this absolutism and accept limits on the right’s strength and scope. Feinberg’s is one of two extreme views that make liberalism a simple, unitary view—the other one being that the right to autonomy is not an intrinsic value, but of value only instrumentally, as a means to promoting welfare.<sup>42</sup> As I see it promoting welfare and respecting autonomy are independent values, neither reducible to the other. They are often complementary but sometimes in conflict, and the challenge faced by liberal theorists (“comprehensive” rather than “political”) is to identify guidelines for resolving the conflicts without shortchanging either value. The problem with both a purely utilitarian defense of paternalism and an absolutist, libertarian anti-paternalism is that each fails to take seriously one of these values.

The first limit on the right that liberals can and should recognize makes it stronger with regard to the more important choices that are crucial to defining who we are or what we stand for (i.e. “self defining” or “self constituting” choices),

---

<sup>40</sup> Thaler and Sunstein (2003), p. 1161.

<sup>41</sup> Feinberg, J. (1986), pp. 54-6. Hodson (1977) and VanDeveer (1986), who both accept versions of HC, make the target’s hypothetical consent to the interference a necessary condition of justified paternalism.

<sup>42</sup> Cite Feinberg’s rejection of this view.

and weaker with regard to other choices.<sup>43</sup> I take a choice to be self-constituting just in case it is *required* by one's conception of the good. Thus, for the orthodox Jew or the morally committed vegan the choice not to eat pork is self-constituting, while for the rest of us, with conceptions of the good that neither forbid nor require eating pork, it is not self-constituting. Arguably, self-constitution is an important component of what J.S. Mill meant by "individuality." I grant that the line separating choices that are self-constituting from those that aren't is sometimes indistinct. Still, there are clear cases on both sides of the line. The stubborn refusal to wear a helmet while riding a motorcycle seems for many an affirmation of a rebellious "born to be wild" self-image. By contrast, the choice to wear/not wear a seat belt does not have a remotely similar significance for most drivers.

Smoking, getting high on hard or soft drugs, and eating junk food seem to me choices that are *not* self-constituting for the *average* person who makes them. Of course the average smoker may well believe that he does no wrong to self or others when he smokes in private and that no one else has the right to dictate to him what he does to his own body. But surely not many cigarette smokers regard their smoking as a crucial part of their self-identity on a par with being a gay activist, devout Catholic, or adventurous risk-taker who laughs in the face of death. Smoking is usually merely a *habit* not required by one's conception of the good life. I grant that there are exceptions to this generalization. Some smokers (especially of cigars and pipes) are tobacco aficionados; they value their ability to savor different tobacco flavors and would regret it if they woke up one morning and found that they had lost their passion for smoking. There are drug users who value their altered states of consciousness as a way to achieve mystical union with the divine. But these cases are the exception, not the rule.<sup>44</sup>

My view is that the right to autonomy is very strong in cases like *Brother Francis* or the Jehovah's Witness who refuses a life-saving blood transfusion, because their choices are clearly self-constituting, while the right is weaker in cases involving unhealthful lifestyle choices to smoke, drink to excess, use hard drugs, etc. because those choices typically are not self-constituting. In the former cases the right is strong enough to override the duty to promote welfare, whereas in the latter cases it is not.

---

<sup>43</sup> Kleinig, J. (1983), p. 68 limits the choices protected by the right to autonomy to those important to "personal integrity." Waldron, J. (1981), esp. pp. 34-5, limits the right's scope in a similar way.

<sup>44</sup> Endorsing these generalizations about smoking and drug use puts me in agreement with George R. (1993) pp. 106-7 and apparent disagreement with Moore M. (1997), pp. 793-5. (Yes, this gives me pause). However, Moore seems to have an overly broad, almost Sartrean conception of "self-constitution." For example, he regards a choice to continue using hard drugs despite their destructive effects on one's life as a self-constituting decision to commit suicide slowly.

A second limit on the right's strength and scope is more important for the justification of state paternalism than "private" paternalism by a friend or relative. The former may be harder to justify than the latter, because in order to respect the rule of law and avoid giving public officials wide, easily abused discretionary powers, it must target persons insofar as they fall under some general description—e.g. drivers who fail to wear their seat belts, recreational hard drug users, etc. Because it targets a class of persons, it is almost certain to infringe the right to autonomy of some in the class but not others and to benefit some while harming others. These varied effects on different persons mean that state paternalism raises distributive justice concerns not raised by private paternalism.<sup>45</sup> Sorting out these concerns is no easy task, but completing it seems to me to require recognition of another limit on the right to autonomy: it does not absolutely protect even self-constituting, self-regarding choices from interference aimed at increasing the welfare or autonomy of *others*.

Suppose that i) there is some type of choice that is imprudent for some but harmless or even slightly beneficial to others who make it; ii) most of those for whom the choice is imprudent make it due to *akrasia*; some of them would consent to either a shove or a nudge that would protect them from themselves ("group 1"), while others ("group 2") prefer to suffer the harm to having their *akratic* choice thwarted by a paternalist's coercion or manipulation; iii) none of those for whom the choice is harmless ("group 3") would consent to the shove or nudge, which may leave them worse off; iv) there is no simple, cost effective way for the state to determine which of these three groups any particular individual belongs to. If the right to autonomy is limited in the second way, then the state may be justified in infringing the right to autonomy of those in groups 2 and 3 by coercing or nudging all, in order to protect those in group 1 from their weak-willed imprudence. Whether it is in fact justified depends on relative size of the three groups, as well as the size of the benefit that the paternalism confers on those in group 1 and the size of the harm it inflicts on those in group 3. The mere fact that there are *some* whose right to autonomy it infringes is not *sufficient* to make it wrong. Feinberg, I think, accepts a limit on the right of this type, whereas Nozickian libertarianism, with its insistence on rights as "absolute side constraints, does not."<sup>46</sup>

Any state paternalism is bound to leave some worse off than they would have otherwise been. But as the means of implementing it shift from criminalization *via* imprisonment to criminalization *via* fines, to sin taxes and other coercive but non-criminal measures, and finally to non-coercive manipulation, the

---

<sup>45</sup> See Arneson, R. (1989).

<sup>46</sup> Feinberg regards a law requiring all motorcycle license applicants to attend an educational session on the risks of helmetless riding as permissible soft paternalism, despite the fact that it would apply to motorcyclists who don't want to attend the sessions and are already well informed of the risks. Feinberg has to say that even though the law infringes their autonomy, it is justified on account of the increased welfare and autonomy that it makes possible for *other* motorcyclists. See Feinberg J. (1986), p. 136.

costs that it imposes on those whom it doesn't benefit (both inside and outside the group it targets) are likely to decrease, and as a consequence, the likelihood that it will do enough net good to justify infringing the autonomy rights of some increases. This does not mean that coercive state paternalism is never justified. Coercive seat laws seem fully consistent with the two limits on the right that I've described. The anti-smoking measures taken in the U.S. since the 1960's—including increased taxation, restrictions on advertising, and bans on smoking in public places for the purpose of "denormalizing" it—seem to me clearly justified by their public health benefits despite the fact that some are coercive, they interfere with the self-constituting choices of some, they probably increase the stigmatization experienced by those smokers who can't or won't quit.<sup>47</sup> What it does mean is that manipulative state paternalism is in general easier to justify than coercive state paternalism. I do not see how it can plausibly be denied that a manipulative public health campaign that uses slick marketing methods to discourage hard drug use among impressionable young people does much more good than harm, especially for the worst off members of society.<sup>48</sup>

---

<sup>47</sup> Due in large part to these measures, the percentage of adults who smoke in this country has dropped from 42% in 1965 to 18% in 2012. According to a recent study in the *Journal of the American Medical Association* (January 8, 2014) this reduction means that about eight million Americans will live about twenty years longer. 44 million Americans still smoke, and it's estimated that 1% of them—about 4 and half million (!)—will die from smoking related illnesses each year.

<sup>48</sup> Bovens, L. (2008) worries that while it might be salutary in the short run, "the long-term effect of nudge may be infantilisation, i.e. decreased responsibility in matters regarding one's own welfare" (p. 11). Of course this concern applies to all forms of paternalism, not just manipulative ones. It seems to me to identify a good *consequentialist* reason why we should be careful and highly selective in our paternalistic interventions—rather than a trumping reason against any paternalism whatsoever.

## References

- Arneson, R. (1989). Paternalism, Utility, and Fairness. *Revue Internationale de Philosophie* 170, 409-23.
- Benn, S. (1967). Freedom and Persuasion. *Australasian Journal of Philosophy* 45, 259-75.
- Bovens, L. (2008). The Ethics of Nudge. Till Grüne-Yanoff and S.O. Hansson (eds), *Preference Change: Approaches from Philosophy, Economics and Psychology*. Springer.
- Childs, J. (1994). Can Libertarianism Sustain a Fraud Standard? *Ethics* 104, 722-38.
- Coons, C. and Weber, M. (2013). Introduction to *Paternalism: Theory and Practice*. Cambridge Univ. Press.
- Crampton, E. (2009). Public Health and the New Paternalism. *Policy Magazine* 25, 36-40.
- Crisp, R. (1987). Persuasive Advertising, Autonomy, and the Creation of Desire. *Journal of Business Ethics* 6, 413-18.
- DeMarneffe, P. (2006). Avoiding Paternalism. *Philosophy and Public Affairs* 36, 68-90.
- Dworkin, G. (1972). Paternalism. *The Monist* 56, 64-48.
- Feinberg, J. (1984). *Harm to Others*. Oxford University Press.
- Feinberg, J. (1986). *Harm to Self*. Oxford University Press.
- Feinberg, J. (1990). *Harmless Wrongdoing*. Oxford University Press.
- Gaus, G. (2003). Liberal Neutrality: A Compelling and Radical Principle. S. Wall and G. Klosko (eds), *Perfectionism and Neutrality*. Rowman and Littlefield.
- George, R. (1993). *Making Men Moral*. Oxford University Press.
- Glaeser E. (2006). Paternalism and Psychology. *University of Chicago Law Review* 73.
- Goodin, R. (1995). *Utilitarianism as Public Policy*. Cambridge Univ. Press.
- Gordon, D. (1980). Comment on Hospers. *Journal of Libertarian Studies* 4, 267-272.
- Gordon, D. (2008). Libertarian Paternalism. *The Mises Review*. Available on-line at <http://mises.org/daily/2965>.
- Griffin, J. (1986). *Well-Being*. Oxford University Press.
- Haworth, L. (1984). Autonomy and Utility. *Ethics* 95, 5-19.
- Hausman, D. and Welch, B. (2010). Debate: To Nudge or Not to Nudge. *Journal of Political Philosophy* 18, 123-36.
- Hill, T. (1973). Servility and Self-Respect. *The Monist* 57, 87-104.
- Hodson, J. (1977). The Principle of Paternalism. *American Philosophical Quarterly* 14, 61-69.
- Kahneman, D. (2011). *Thinking: Fast and Slow*. Farrar, Straus, and Giroux.
- Kleinig, J. (1983). *Paternalism*. Rowman and Allenheld.
- Korsgaard, C (1986). The Right to Lie: Kant on Dealing with Evil. *Philosophy and Public Affairs* 15, 325-49.

- Machan, T. (1995). *Public Rights and Private Illusion*. The Independence Institute.
- Mitchell, G. (2004). Libertarian Paternalism is an Oxymoron. *Northwestern University Law Review* 99, 1245-78.
- Moore, M. (1997). *Placing Blame*. Oxford University Press.
- Nozick, R. (1974). *Anarchy, State, and Utopia*. Basic Books.
- Parfit, D. (1984). *Reasons and Persons*. Oxford University Press
- Quong, J. (2011). *Liberalism Without Perfection*. Oxford University Press.
- Rawls, J. (1971). *A Theory of Justice*. Harvard University Press.
- Raz, J. (1986). *The Morality of Freedom*. Oxford University Press.
- Rizzo, M. and Whitman, D.G. (2009). Little Brother is Watching You: New Paternalism on the Slippery Slopes. *Arizona Law Review* 51, 684-739.
- Rogers, S. (1992). How a Publicity Blitz Created The Myth of Subliminal Advertising. *Public Relations Quarterly* 37, 12-17.
- Rothbard, M. (1985). *The Ethics of Liberty*. New York University Press.
- Shiffrin, S. (2000). Paternalism, Unconscionability Doctrine, and Accommodation. *Philosophy and Public Affairs* 29, 205-250.
- Scoccia, D. (2013). The Right to Autonomy and the Justification of Hard Paternalism, in *Paternalism: Theory and Practice*.
- Sugden, R. (2008). Why Incoherent Preferences Do Not Justify Paternalism. *Constitutional Political Economy* 19, 226-248.
- Thaler, R. and Sunstein, C. (2003). Libertarian Paternalism is Not an Oxymoron. *University of Chicago Law Review* 70, 1159-1201.
- Thaler, R. and Sunstein, C. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Penguin Books.
- Vallentyne, P. (2007). Libertarianism and the State. *Social Philosophy and Policy* 24, 187-205
- VanDeVeer, D. (1986). *Paternalistic Intervention: The Moral Bounds of Beneficence*. Princeton University Press.
- Waldron, J. (1981). The Right to Do Wrong. *Ethics* 92, 21-39.
- Wilkinson, T. (2013). Nudging and Manipulation. *Political Studies* 61, 341-55.
- Wilkinson, W. (2008). Why Opting Out is No 'Third Way.' *Reason Magazine*.