WHAT'S SO BAD ABOUT LEGAL PATERNALISM? (OR WHAT'S SO GOOD ABOUT AUTONOMY?)[1]

William J. Talbott
University of Washington

Prefatory Note: This paper is a reworking of two chapters from a larger book project (*Human Rights and Human Well-Being*, forthcoming from Oxford). I have tried to make the paper self-contained enough to stand on its own. Because the format of the seminar does not require me to read the paper, I have included more material than I could have included in a single presentation. My hope is there is something to interest almost everyone. The paper has three parts: Introduction; The Pure Theory of Legal Paternalism; and Legal Applications Of The Most Reliable Judgment Standard. Those with more practical interests may want to jump from the introduction to the legal applications in Part III.

PART I. INTRODUCTION

The Evolution of Rights Against Legal Paternalism

Though there is no right against legal paternalism as such currently enacted anywhere in the world, the historical development of human rights cannot be understood unless it is seen as, in part, the development of rights against legal paternalism. The most important event in the historical development of rights against legal paternalism is the development of a right to religious freedom. The reason is simple: There is no greater harm a person could do to herself than to bring it about that she suffers unbearable torment for all eternity. Suppose I believe that will be your fate if you do not practice my religion. I propose to save you (and others like you) from eternal suffering by making it illegal for you to practice any religion but mine. This legal establishment of my religion would be an example of legal paternalism, because enforcing it would involve my overruling your own judgment about what is good for you.[2] A right to freedom of religion represents a rejection of this kind of paternalism. Once it is allowed that people should be free to make and follow their own judgments of what will be to their eternal benefit and harm, it is hard to see why they should not be equally free to make and act on less momentous decisions about what is good for them.

After a right to religious freedom, the second most important step in the development of a right against legal paternalism is the development of the rights that guarantee the necessary background conditions for autonomy, especially the rights to civil liberties. In addition to being essential background for autonomy, rights to freedom of expression, freedom of association, and freedom of the press are important steps in the development of rights against legal paternalism, because they involve a recognition that those in authority should not be deciding which ideas it is bad for people to be exposed to or to think about or to discuss.

With all these rights in place, there are many potential pathways to a more robust right against legal paternalism. The typical pathway involves the recognition of other choices that, like the choice of a religion, are deeply personal choices, the effects of which are borne primarily by the person making the choice. In spite of opposition from almost all major religions and in spite of laws to the contrary, in the U.S., a right against

legal paternalism has gradually developed around personal choices concerning sex, love, and death.

Freedom From Interference In Choices Concerning Sex And Love

The main legal development has been the U.S. Supreme Court's articulation of a right not explicitly found in the U.S. Constitution, first introduced as a "penumbral" right to privacy. The leading case in the development of the right to privacy was *Griswold v. Connecticut*, in which the Supreme Court declared unconstitutional a law forbidding the use of contraceptive drugs or devices.[3] Declaring that the Bill of Rights creates a penumbra which includes a right to privacy, the Court declared the statute an unconstitutional limitation on a right to marital privacy.

Because the statute in *Griswold* held out the possibility of searches of a couple's bedroom for contraceptive devices, it was reasonable to think the implicated right was a right of privacy. In fact, the issue raised by the statute was whether people should be free to make certain sorts of decisions and to act on them without the government overruling their judgment.[4]

Two years later, the Supreme Court expanded the protection against legal paternalism when it struck down as unconstitutional a Virginia law against miscegenation in *Loving v. Virginia*.[5] Though the law was struck down on the grounds that it involved invidious racial discrimination that could not survive strict scrutiny, the Court also realized it involved government interference in an important sphere of personal autonomy: "The freedom to marry has long been recognized as one of the vital personal rights essential to the orderly pursuit of happiness by free men."[6]

It was reasonable to think the right articulated in *Griswold* was a right to privacy, because it concerned decisions and actions by married couples in their bedroom. The doctrine was expanded to cover decisions and actions outside the bedroom when the Court struck down legal prohibitions on abortion in *Roe v. Wade*.[7] The real issue in *Roe*, as in *Griswold*, was to define an area of individual autonomy where people are free to act on their own judgment about what is good for them without having it overruled by the government. In a subsequent case, the court defined this sphere as "involving the most intimate and personal choices a person may make in a lifetime, choices central to personal dignity and autonomy."[8]

In 1986, the Supreme Court refused to accept a clear implication of its own prior decisions, when, in *Bowers v. Hardwick*, it upheld the constitutionality of a Georgia prohibition of sodomy. The court upheld the prohibition though there could hardly be a more intimate and personal choice than the choice of a sexual partner made in the privacy of one's own bedroom.[9] Ironically, twelve years later the Georgia Supreme Court did rule the Georgia sodomy law unconstitutional on the basis of the same penumbral privacy right rationale the U.S. Supreme Court had articulated in *Griswold*.[10] It took five more years for the U.S. Supreme Court to accept the implications of its earlier decisions, by overruling *Bowers* in *Lawrence v. Texas*.[11] In *Lawrence*, the Court for the first time defined the right as a "liberty" rather than a "privacy" right, and defined the sphere of protected liberty to include "personal decisions relating to marriage, procreation, contraception, family relationships, child rearing, and education."[12]

In his dissenting opinion in *Lawrence*, Justice Scalia pointed out that a clear implication of the *Lawrence* decision is that prohibitions on homosexual marriage are unconstitutional. Unfortunately, it seems very unlikely the Supreme Court will

acknowledge this implication anytime soon. Thus, in matters of sex and love, Europe and Canada have now surpassed the United States in the working out of a sphere of personal autonomy free of paternalistic government interference.[13]

Freedom From Interference in Choices Concerning One's Own Death

In 1979 the New Jersey Supreme Court interpreted the right to privacy to include a right to refuse extraordinary measures to save one's life and a right of a guardian to refuse them for an incompetent patient.[14] In 1990, the U.S. Supreme Court extended the right to include a right to refuse ordinary life support. Where a patient was incompetent, the Court required clear and convincing evidence of intent.[15] As a result of these decisions, normal adults in the U.S. can exercise their right to refuse extraordinary measures or to refuse life support by executing an advance directive or living will. Because the decisions to refuse extraordinary measures or to terminate life support are typically made in a hospital, not in the privacy of one's own home, and because they typically involve interactions with strangers, these cases show even more clearly the right to "privacy" is a misnomer. It is really a right to a sphere of personal autonomy free of paternalistic interference.

In 1997, the U.S. Supreme Court declined to extend the right to "privacy" to include a right to assisted suicide, even though an *amici curiae* brief was filed in support of such a right by six of the most prominent political philosophers of the last half of the 20th century.[16] The six philosophers invited the Court to take the step of explicitly announcing a right to a sphere of autonomy protected from paternalistic interference that would include a right to assisted suicide. Not surprisingly, the Supreme Court declined their invitation. Nonetheless, the momentum for such a right continues to build. Oregon voters have twice endorsed it. In addition, even though his actions in assisting suicides were clearly illegal under existing law, juries consistently refused convict Dr. Kevorkian of any crime until prosecutors were able to find a case in which there were genuine doubts about whether the suicide Kevorkian assisted with was fully voluntary.

I believe the six philosophers were correct to urge the Court to reconceptualize the right to privacy as the guarantee of a sphere of individual autonomy free from paternalistic interference. In this paper, I explain why the sphere should include much more than the personal choices about love, sex, and death that have been the focus of recent Court decisions.

What Laws are Paternalistic?

In order to articulate and evaluate a right against paternalism, it is necessary to be clear on just which laws are paternalistic. A law is paternalistic if it is enacted to promote the good of the target audience by *overruling* their own judgment about what is good for them. Generally speaking, when a paternalistic law is enacted, some members of the target audience will regard themselves as worse off than they would be without the law and practically no one will regard themselves as significantly better off. The reason is simple. Consider a prohibition on going to movies on Sunday. Those who think it is better for them not to go to movies on Sunday can refrain from going whether or not there is a law. So the law does not make them any better off.[17] Those who would go to movies on Sunday but for the law will regard themselves as worse off with the law than without it.

Legal paternalism is the enactment and enforcement of paternalistic laws. It is important to distinguish legal paternalism from legal solutions to collective action

problems (CAPs). Paternalistic laws resemble legal solutions to CAPs in that both sorts of laws are aimed at promoting the good of those they coerce. However, a legal solution to a CAP *gives effect* to the judgments of the target audience about what is good for them by bringing about an overall outcome they generally regard as better for them than what the outcome would be without the law. For example, if there were no legally enforced traffic signals, almost all drivers would regard themselves worse off, because driving would be so much more hazardous.

In previous chapters, I have discussed legal solutions to CAPs that are the basis of rights, for example legal solutions to the security CAPs (security rights and procedural rights), legal solutions to productivity CAPs (e.g., property and contract rights), legal solutions to information CAPs (truth-in-labeling laws and other rights to information), legal solutions to some workers' CAPs (e.g., rights to collective bargaining and minimum wage rights), and legal solutions to CAPs that are involved in attempting to enact legal solutions to other CAPs (e.g., democratic rights). The category of rights that are legal solutions to CAPs also includes product safety laws, including laws requiring the testing of drugs and other potentially hazardous products, product liability law, workers' compensation laws, and indeed the entire system of tort law (i.e., the law of liability for personal injury). As discussed in a previous chapter, it is important to realize that solving the relevant CAPs often requires that the relevant rights not be waivable, that is, that they be inalienable.

Not all legal solutions to CAPs create rights. Some just create duties. For example, traffic control laws, licensing laws, anti-pollution laws, zoning laws, building codes, occupational safety laws, securities laws, and anti-trust laws. These categories include laws enacting regulatory agencies (e.g., the Securities and Exchange Commission or the Federal Trade Commission) and a large amount of administrative law enacted by regulatory agencies. Professional licensing requirements are sometimes thought to be paternalistic. It is true that in some cases the avowed justification may be paternalistic. It need not be. Where the law makes it easier for an ordinary person to obtain the services of a competent practitioner and to avoid the services of incompetent ones, something almost everyone desires, the rationale need not be paternalistic. Because professional licensing laws do involve the delegation of authority to some licensing group to distinguish who is or is not competent, they are sometimes confused with paternalistic laws. However, delegations or authority need not be paternalistic. I return to this topic below.

Other legal solutions to CAPs provide public goods, for example, government investment in medical and other scientific research, streets and highways and mass transit, sewers and utilities, public radio and TV, and parks and other protected areas.

Another category of legal solutions to CAPs are solutions to what I have referred to as *status CAPs*. Codes of honor can solve CAPs, but they can also generate others. Thus, for example, codes of revenge can generate potentially endless cycles of violence that can be terminated by a criminal justice system with laws against exacting revenge. Laws against dueling have a similar justification.

Men's status CAPs usually involve what might be called a *macho* code of honor. So, for example, before a rule requiring all NHL hockey players to wear helmets was adopted, there were powerful motivations not to be among the first to wear one, because to do so would be regarded as a sign of weakness, and result in a great loss in status.[18]

Similar motivations can come into play in the decision whether or not to wear a motorcycle helmet or to use a seat belt when riding in a car. To the extent that mandatory helmet laws or seatbelt laws solve a status CAP, they are not paternalistic. I discuss other justifications for them below.

In a previous chapter I provided other examples of solutions to status CAPs, including laws against foot-binding. Female status CAPs typically do not involve macho codes of honor but what might be called *patriarchal* codes of honor. Foot-binding improved prospects for marriage, probably the most important kind of status for a woman in a patriarchal society. Female genital cutting is a similar case. In all societies in which it is practiced, women who do not undergo this procedure have much lower marriage prospects. The case of female genital cutting is complicated by the fact that the practice is typically supported by beliefs which, if true, might well be thought to justify it—for example, that a baby will suffer brain damage or death if its head touches the mother's clitoris during birth. Even without these sorts of justification, the parents of female children in a village in which female genital cutting is generally practiced may face a CAP that can be solved by a joint declaration banning the practice (G. Mackie 1996).

PART II. THE PURE THEORY OF LEGAL PATERNALISM

Almost everyone agrees that some paternalism is morally justified—for example, paternalism toward young children and toward the severely mentally ill or severely mentally impaired. Some of those who think some paternalism can be justified nonetheless oppose all *legal* paternalism, either because of the potential for abuse (e.g., the involuntary commitment of political opponents as mentally ill) or because of the unavoidable side effects of legally enforcing paternalistic laws (e.g., the large number of those currently incarcerated in the U.S. for selling or using small amounts of illegal drugs, where the laws are aimed at protecting drug users whether they want to be protected or not).

These concerns with potential abuse and potential negative side effects are very real practical concerns that would be relevant to any attempt to justify legal paternalism. However, I wish to temporarily set these practical concerns aside, so I can focus on what might be called the *pure theory of legal paternalism*, that is, the question of what sorts of paternalism could be justified if there were no concerns about potential abuse or potential negative side effects. Because my account carries no presumption against legal paternalism for children and non-autonomous adults, in the remainder of this paper I focus primarily on paternalism targeted at autonomous adults—that is, adults with normal development of normal cognitive and emotional capacities.

Two Rationales for a Right Against Legal Paternalism

The simplest rationale for a right against legal paternalism is the Kantian rationale that paternalistic intervention is incompatible with respect for individual autonomy. The Kantian rationale is *nonconsequentialist*, because it does not depend on the truth of any claim about the goodness or badness of the consequences of respecting individual autonomy. For most nonconsequentialists, the moral problem with paternalism is that it involves interfering with autonomous, self-regarding choices. Thus, for most nonconsequentialists, all autonomous self-regarding choices would be covered by a right against legal paternalism.

In this paper, my focus is on the *consequentialist* case against legal paternalism. From a consequentialist perspective, the focus shifts from the *autonomy* of the target's choice to the *reliability* of the target's judgment. From the consequentialist perspective, the judgments of normal adults about what is good for them deserve protection because they are generally reliable and thus the policy of interfering with them can be expected to reduce general well-being.

In a previous chapter, I discussed Mill's rationale for opposing legal paternalism. I believe the consequentialist case is made stronger if Mill's utilitarian criterion is replaced with a consequentialist criterion that pays attention to the distribution of well-being. In this paper I simply assume there is some such criterion and show how it could be used to define the contours of a right against legal paternalism.

The Millian rationale for a right against legal paternalism depended crucially on what I called *the claim of first person authority*: the claim that the judgments of autonomous adults about what is good for them are generally reliable, and, in addition, are generally more reliable than the judgments of other people, including the judgments of governments. In a previous chapter, I considered some reasons for thinking the claim of first person authority is true. However, rather than assuming it is true, I suggested we understand rights respecting democracies as social experiments to determine whether it is

true. On this approach, we suppose the claim of first person authority is true and is generally taken to be true and then ask: What rights could be justified on consequentialist grounds? That is the approach I take in this chapter to determine the contours of a consequentialist right against legal paternalism.

Perhaps it is not surprising that consequentialism requires us to draw the line between justified and unjustified paternalism in a different place from where most nonconsequentialists would draw it. What is more surprising is that a good case can be made that the consequentialist account draws the line in the right place.

If the claim of first person authority were generally taken to be true, autonomous adults would be regarded as the most reliable judges of what is good for them. This would seem to make it impossible to justify *any* legal paternalism directed toward autonomous adults on consequentialist grounds. Recall that the rationale for a paternalistic law requires it to be aimed at promoting the good of those it coerces, by overruling their own judgment of what is good for them. If an autonomous person's judgments about what is good for her were acknowledged to be generally more reliable than the judgments of other people, it would seem it could never be reasonable for legislators to enact such legislation.

In the remainder of this paper, I explain why this is not quite right by articulating a standard for a category of *weak* paternalism that could be justified on consequentialist grounds even if the claim of first person authority were generally recognized to be true. I refer to it as the *most reliable judgment standard for weak paternalism*. The best way for me to motivate the most reliable judgment standard is to consider the inadequacies of other ways of drawing a line between justifiable and unjustifiable paternalism.

Explicit Voluntary Endorsement Standard for Weak Paternalism

Strict libertarians would hold that paternalism is never justified, unless the target of the paternalistic interference has previously given her explicit, voluntary consent to it. Call this the *explicit voluntary endorsement standard for weak paternalism*. The strict libertarian is surely right that explicit voluntary consent can justify paternalistic interference, but most people would not believe it reasonable to define justifiable paternalism so narrowly. Suppose, for example, in a feverish state, Arnold hallucinates a pedestrian bridge outside his third story window, where there is none. Suppose he decides to walk right out the window. Fortunately, you and some of his other friends are visiting. When you realize what he is planning to do, you try to talk him out of it by pointing out the consequences of falling from a third story window. Arnold replies by telling you that you have overlooked the fact that he will be walking on a sturdy bridge that can easily support his weight. Nothing you say can convince him there is no bridge there. You and his other friends know if you forcibly prevent him from walking out the window, Arnold will be grateful to you after his fever passes. So when Arnold tries to step through the window, you and his other friends physically restrain him. Then you call the legal authorities who confine Arnold for his own protection for 24 hours until the fever resolves.

Undoubtedly, you and Arnold's other friends and the authorities who were called all interfered with Arnold's liberty for his own good. Because he disagreed at the time, your intervention required you to overrule his own concurrent judgment about what was good for him. Was the intervention justified? According to the explicit voluntary endorsement standard, it would only be justified if Arnold had previously given his

explicit voluntary consent to it, perhaps by having executed a durable power of attorney giving you the authority to make decisions for him during periods of incapacity. This is unrealistic. Most people have not given their explicit consent to all the various kinds of paternalistic interference they would endorse if they were asked. This suggests a modification to the explicit voluntary endorsement standard.

The Hypothetical Autonomous Endorsement Standard

The preceding example suggests that paternalism can be justified by hypothetical or implicit consent as well as by actual, explicit consent. This is the main idea behind what I refer to as the *hypothetical autonomous endorsement standard for weak paternalism*. The best way to explain this standard is to explain how it is an improvement on what is surely the most influential account of weak paternalism in the literature, Feinberg's (1971) voluntariness standard.

Feinberg's (1971) account of weak paternalism focuses on the choice the paternalistic intervention aims to prevent. On Feinberg's account, paternalistic intervention is weak (and therefore justifiable) only if it involves interference in a subject's substantially non-voluntary choices or if it is necessary to determine whether or not they are substantially non-voluntary (1971, 9).[19] What is distinctive about Feinberg's account is his distinction between voluntary and non-voluntary choices is not dependent on their causal antecedents, but depends only on whether they "represent the agent faithfully in some important way: they express his or her settled values and preferences."(7) Applied to the example of Arnold, Feinberg's account would hold Arnold's choice to walk out the window was substantially nonvoluntary because it did not express his settled values and preferences. Intervention to prevent him from walking out the window would qualify as weak paternalism because it would in fact promote Arnold's own settled values and preferences.

Feinberg applies his account not only to cases involving temporary incapacity or derangement, but also to cases of simple ignorance. Thus, he agrees with Mill:

> If either a public officer [Dick] or anyone else saw a person [Harry] attempting to cross a bridge which had been ascertained to be unsafe, and there were no time to warn him of his danger, they might seize him and turn him back, without any real infringement of his liberty . . . .[20]

The example of choices due to ignorance raises two important issues for Feinberg's account. One can be accommodated by a clarification in the spirit of the account, the other will require a revision of it.

The first issue can be developed out of Feinberg's discussion of the Mill example. Suppose that after Dick stops Harry from crossing the bridge, Dick explains to Harry that the bridge is unsafe. Suppose Harry replies he knows it is unsafe. He wants it to collapse because he is a movie stuntman. Feinberg would agree intervention is no longer justified and Dick should allow Harry to proceed (1971; 8).

Consider a variant Feinberg does not discuss. Suppose after Dick explains to Harry that the bridge is unsafe, Harry agrees with Dick that he would not want to try to cross an unsafe bridge, but disagrees with Dick on whether the bridge is unsafe. In this case, both Harry and Dick agree Harry's settled values and preferences favor avoiding unsafe bridges, but they disagree on whether the bridge he is about to cross is unsafe. Is Dick justified in acting so as to force Harry to do the act he (Dick) believes will best achieve Harry's settled preferences for crossing a safe bridge, or, having given Harry

whatever information he may have about the safety of the bridge, is he required to allow Harry to act on his own judgment about how to best pursue his settled preference to cross a safe bridge?

If there is no reason to believe Harry is suffering from any incapacity of judgment, I believe it would be objectionably paternalistic for Dick to intervene. It is hard to imagine a case of this kind involving unsafe bridges, but it is easy to imagine other kinds of cases. For example, it is not objectionably paternalistic for the government to publish information on the potentially harmful effects of various drugs. Subject to some exceptions I discuss shortly, autonomous adults should be allowed to gather their own information and to make their own determination of safety. If they decide the government information is slanted to overstate the dangers and if there is no issue of potential harm to others, they should be able to exercise and act on their own judgment of the dangers.

To reach this result, Feinberg's voluntariness standard must be understood to limit weak paternalism to cases in which the target's choice does not reflect *what his own autonomous judgment of how best to promote his settled values and preferences would be* (were he to consider the question). I believe this is a clarification Feinberg would accept.

The second issue for Feinberg's account is more substantial. It concerns what exactly the test for weak paternalism is to be applied to. Consider a potential target of paternalist intervention. Feinberg focuses on the target's attitude toward the choice to be intervened with. I believe he should have focused on the target's attitude toward the relevant kind of intervention. Two examples will serve to illustrate the difference:

Peter is someone who refuses to use maps and refuses to ask for directions, even when he is hopelessly lost. One day Peter volunteers to drive to a nursery to pick up a plant for his wife, Martha. Martha gives him directions and he sets off in her car.

Martha expects him back within an hour. When he does not return, she calls the nursery and finds out he has not yet arrived. Martha concludes he is lost. Fortunately, Martha's car is equipped with a GPS navigator with a special kind of autopilot. When a destination is input, the navigator will map a route to the destination. The map can be input to the autopilot, which, if activated, will take control of the car and drive it to the destination. Martha has no way to contact Peter, but she can send commands to the navigator/autopilot device from home. When she activates the autopilot, it announces to Peter that it is taking over control of the car to drive it to the nursery and then does so. When Peter arrives home from the nursery, he is furious. Although he agrees that taking the most direct route to the nursery reflected his settled values and preferences and he agrees that he did not take the most direct route, he is livid that his wife took control of the car away from him by activating the autopilot.

Like the example of Harry crossing the unsafe bridge, this is an example in which ignorance renders a choice substantially nonvoluntary. Feinberg does not want to allow this sort of paternalistic intervention to be justified, so he qualifies his account by saying: "Even substantially nonvoluntary choices deserve protection unless there is good reason so judge them dangerous."(1971, 8). By adding this proviso, Feinberg makes sure his account gives the right result in this sort of case. However, I don't believe Feinberg has identified the reason Martha's paternalistic intervention is objectionable in this case. To see why not, consider a variant on this example:

When he goes driving, even when he has a map to guide him, Mark often gets lost. He has no reluctance about asking directions and appreciates it when someone offers help. One day Mark's wife Rachel asks him to pick up a plant for her at the nursery. Because their car is in the shop, Mark borrows their neighbor's car. When Mark is not back within an hour, Rachel calls the nursery and is told he has not yet arrived. Her neighbor overhears the conversation and tells her his car is equipped with a navigator/autopilot device accessible from home. Neither Mark nor Rachel had been aware that such a device existed. Rachel has no way to contact Mark, so she simply activates the navigator/autopilot device. The device announces to Mark that it is taking over control of the car to drive it to the nursery and then does so. When Mark arrives home he is thrilled Rachel was able to activate the autopilot and have it drive him to the nursery.

Let us stipulate that both Peter and Mark's settled values and preferences favored taking a direct route to the nursery. On Feinberg's account, both Peter and Mark's driving choices were substantially nonvoluntary, because they did not reflect their settled values and preferences. However, on Feinberg's account, paternalistic intervention in both cases qualifies as strong, not weak, because even though their choices are nonvoluntary, there is no reason to think they are dangerous to anyone. So, on Feinberg's account, in neither case was the paternalistic intervention justified.

This seems to me to be a mistake. On the assumption both wives knew their husbands well enough to anticipate how they would respond to the intervention, it seems to me Martha's intervention in Peter's driving was not justified, but Rachel's intervention in Mark's driving was. Once we change our focus from the target's choices that are interfered with to the target's attitude toward the relevant kind of paternalistic interference, the explanation for the difference is quite straightforward.

Suppose neither Peter nor Mark had ever explicitly considered whether they would endorse such interference with their driving. Nonetheless, it might have been clear to Martha that Peter *would not* endorse such intervention were he to consider it and it might have been equally clear to Rachel that Mark *would* endorse such intervention were he to consider it. I describe the examples by saying Mark hypothetically endorses the intervention, but Peter does not. On the assumption the endorsement is autonomous, in the sense in which I use the term, it seems to me that Mark's hypothetical endorsement of Rachel's intervention makes it weak, justifiable paternalism, and that Peter's lack of hypothetical endorsement of Martha's intervention makes it unjustified, strong paternalism. This suggests the following modification of Feinberg's voluntariness standard for weak paternalism, which I refer to as the *hypothetical autonomous endorsement standard for weak paternalism*: Paternalistic interference is weak whenever the target of the interference would autonomously endorse the relevant kind of interference if she were to consider it in such circumstances and in such a manner to enable her to make an autonomous judgment about it.

The hypothetical autonomous endorsement standard gives the results we were looking for in the Peter example and the Mark example. Rachel's intervention in Mark's driving qualifies as justified weak paternalism; Martha's intervention in Peter's driving does not. Nonetheless, I believe the hypothetical autonomous endorsement standard is neither a necessary condition nor a sufficient condition for justifiable weak paternalism. It is not necessary, because some paternalism is weak even though the target would not

autonomously endorse it; it is not sufficient, because some paternalism that the target would autonomously endorse would not be weak. The most important exceptions to the standard are the exceptions to necessity—that is, cases where paternalism is weak, even though the target would not autonomously endorse it. Most of my discussion will focus on examples of that kind. I take up sufficiency later.

The Hypothetical Autonomous Endorsement Standard is not a Necessary Condition for Weak Paternalism

It is not at all obvious that there are cases of weak paternalism violating the hypothetical autonomous endorsement standard. The clue to finding them is to understand autonomy in the non-metaphysical sense in which I use the term. In that sense, a person's judgment is *autonomous* when she has developed the ability to make reliable judgments about what is good for her. Consider the Arnold example again. Perhaps the relevant difference between the non-autonomous judgment of the feverish Arnold and the autonomous judgment of the non-feverish Arnold is not a metaphysical difference having to do with causation, but simply that fact that it is reasonable to expect Arnold's non-feverish judgment to be more reliable than his feverish judgment.

Suppose it is the fact that it is reasonable to regard Arnold's autonomous non-feverish judgment as more reliable than his non-autonomous feverish judgment that gives the former priority over the latter in justifying intervention to prevent him from walking out the third story window. This raises what initially is simply a logical possibility—that it might be reasonable to believe that a person's autonomous judgments made at one time are more reliable than those made at a different time and, on that basis, that intervention endorsed by the target's more reliable autonomous judgments could be justified even if it were opposed by the target's other, less reliable, autonomous judgments. To make this logical possibility begin to seem plausible, I discuss an example.

We currently lack the statistical information necessary to judge how the attitudes of drug users change over time. We do have enough evidence from cigarette smokers to be able to imagine how a drug prohibition could qualify as justified weak paternalism. To keep the example simple, I assume there is a recreational drug RD that does not directly cause harm to anyone other than those who take it. Allen is an autonomous 21-year-old. He welcomes new experiences and wants to live life to the fullest. Allen is aware of studies showing RD to be dangerously addictive, but he knows lots of people his age who use it and who strongly recommend it. They seem to have suffered no ill effects. So he judges that, on balance, it would be good for him to take RD, also.

To make the example interesting, I need to suppose not only that there are scientific studies on the effects of RD, but also that there are statistical studies on the attitudes of users and former users toward the drug. We do have some information of this kind for cigarette smoking. One survey reports 82% of those who smoke believe it would be better for them if they did not and most of them have tried many times to stop, without success.[21] This figure is almost surely a low estimate of the percentage of those who currently smoke who will later regret their decision, both because it includes many smokers who have not been smoking long enough to come to regret it and because it does not include all the former smokers who have successfully quit. It is reasonable to suppose a very large percentage of those who do successfully quit would wish they had never started. So I think it is not unreasonable to suppose that 90% or more of those who become cigarette smokers someday will regret ever having started.

The reason I must move from this real-world example to a hypothetical one is it is not reasonable to suppose that such a high percentage of those who take up smoking would eventually come to endorse a legal prohibition on smoking, even if the potential for abuse and the negative side effects of the prohibition were minimal. Most people think it is better for them to be free to make mistakes and to learn from them than to be prevented from making any. However, some mistakes are so tragic that most people would want to be prevented from making them. Suppose the recreational drug RD has such devastating effects on most people's lives that within twenty years of beginning to use it, 90% of users will not only regret making the decision to use the drug, they will also judge it would have been better for them had there a prohibition on using it, because then they would have been much less likely to use it. I also assume the evidence shows their judgments endorsing a legal prohibition are quite stable over time. Once the come to endorse a legal prohibition, they tend not to change their minds

To make the example complete, I must consider the attitudes not only of those who decide to use the drug, but also of those who do not. Suppose 90% or more of those who do not decide to use the drug will, over time, see its effects on those who do and, on that basis, will not only be glad they did not begin using the drug but would also endorse a legal prohibition on its use. These are strong assumptions. Shortly I show how to relax them.

Consider again the case of Allen, a twenty-one-year-old who plans to take the drug. Allen has heard about tragic cases involving the drug, but all his friends who use it seem glad to be taking it. He knows that over 90% of those who use the drug come to regret the decision within twenty years, but he attributes that change to their growing older. He is like the young man described by Nagel (1970), who in his youth values sex, spontaneity, frequent risks and strong emotions, but who expects that in twenty years he will value security, status, wealth, and tranquility.[22] He does not now endorse the values he expects himself to have in twenty years, so it is no surprise that he does not now endorse the judgments he expects himself to make in twenty years either.

What are we to say about this case? Would it be permissible to intervene to prevent Allen and his friends from using the drug, even though their decision to use it is an autonomous one based on all the available evidence? On the assumption the intervention would be effective in preventing use of the drug and there was no potential for abusing the law and no other potential negative side effects of the law, it seems to me that a legal prohibition could be justified. Although it is true the law would be paternalistic and it is true the law would not satisfy the hypothetical autonomous endorsement standard, I believe it could be justified on the grounds that it was overwhelmingly probable that in the future Allen and others his age would come to endorse such a law in two hypothetical cases: first, if there were no legal prohibition and they were not prevented from using the drug; second, if there were a legal prohibition and they were prevented from using it. It is useful to illustrate Allen's situation with a diagram. See Figure 1.

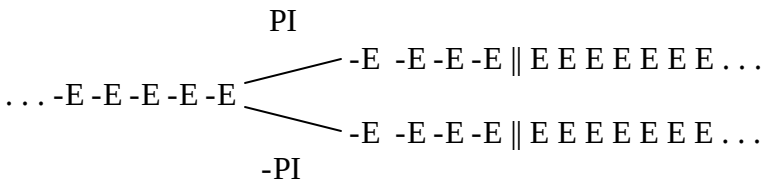

Figure 1. Allen's Decision Whether or Not to Use the Drug RD.

The diagram branches to represent two different scenarios. One branch reflects his future if there is no paternalistic intervention (-PI) and Allen starts taking the drug; the other branch reflects his future if there is paternalistic intervention (PI) and he is prevented from taking the drug. Before he reaches the division, he does not endorse paternalistic intervention (-E). No matter whether there is paternalistic intervention or not, for some time afterward he will continue to oppose such intervention (-E). In either case, eventually he will change his mind and come to endorse (E) paternalistic intervention. If there is no paternalistic intervention, he will start taking the drug and eventually come to the conclusion it would have been better for him if he had been prevented from taking it and; if there is paternalistic intervention, he will not take the drug and he will eventually come to the conclusion that it was good for him to have been prevented from taking it. Finally, in either case, it is reasonable to believe his change of mind to endorse paternalistic intervention will be *unequivocal*—that is, it is reasonable to believe he will not later change his mind. Once he comes to endorse paternalistic intervention, he will continue to do so.

Of course, there is no way to be absolutely certain what Allen's future judgments will be. In the example, I assumed there was a sound statistical basis for being able to predict his future judgments with probability of .9. Shortly I explain why any probability very much above .5 will typically be sufficient to justify paternalistic intervention.

On first impression, it may seem quite counter-intuitive for me to claim to be able to justify paternalistic intervention to prevent Allen from taking the drug. There is such a strong presumption against paternalistic interference with a person's voluntary choices it will take me some time to explain why intervention in Allen's case should be permitted. To explain why, it is useful for me to talk about the judgments of Allen's current and future hypothetical selves. This talk of temporal selves is simply a useful heuristic. I do not mean to imply Allen really is a series of different selves. After employing the heuristic, I explain how to dispense with it.

It is important to recognize the door I am opening for exceptions to the hypothetical autonomous endorsement standard is a narrow one. I am only suggesting that, in this case, Allen's *own* actual or hypothetical future judgment can justify overruling his current judgment about what is good for him, even if his current judgment is based on his settled values and preferences. I am not suggesting *other people's* judgments about what is good for him could justify overruling Allen's own judgment.

The Most Reliable Judgment Standard

To justify paternalistic intervention in Allen's case, I have to give the judgment of Allen's future hypothetical selves priority over the judgment of his current self. How can this be justified? The answer is a straightforward consequence of the consequentialist

rationale for a right against paternalism. On the consequentialist rationale, rights against paternalism are justified by the claim of first person authority, that an autonomous adult is a more reliable judge of what is good for her than other people are. The claim of first person authority leaves it open whether some of an autonomous person's own judgments about what is good for her are more reliable than others. And thus it leaves open the possibility of a narrow category of justified paternalism that involves overruling a person's *own less reliable judgment* about what is good for her, in order to give effect to *her own more reliable judgment* about what is good for her. This suggests a new standard for weak paternalism:

(Most Reliable Judgment Standard for Weak Paternalism) A's paternalistic intervention in the action of a target T is weak paternalism if and only if A is epistemically justified in believing that T's most reliable autonomous judgment or judgments endorse (or would endorse) such intervention.

How could the most reliable judgment standard support paternalistic intervention to prevent Allen from using the drug? Although there is no guarantee our future selves' judgments about what is good for us are more reliable than our past selves' judgments, for most of our lives, our judgments about what is good for us become more reliable with time. This is true because human beings learn from experience. This learning takes two forms. First, there are the facts that help us to better achieve our goals. Almost everyone can think of past decisions that they would have made differently if they had the information they now have.[23] Second, there are the changes in goals themselves. Almost everyone can think of goals that they pursued in the past that they now believe not to have been worth pursuing.[24]

It is clear that people typically regard their hypothetical future selves' judgments as more reliable than their current judgments.[25] I say "typically", because we all realize there is often a significant decline in memory and other cognitive functions late in life. Let me set aside such qualifications by focusing on future judgments made by a future self with the same level of cognitive function as the current self. It is hard to deny some sort of priority to the autonomous judgments of such a future self.

Almost everyone would have higher life prospects if they could have access to their hypothetical future selves' judgments about the consequences of their current decisions. If they could just consult hypothetical future selves ten or twenty years in the future about how well their lives happy they would be with the consequences of a current decision (e.g., a decision on marriage or on a career), people would be able to avoid making many decisions they would later regret.

The most reliable judgment standard coincides with the hypothetical autonomous endorsement standard in most cases, because in most cases a person's current autonomous judgment is the best evidence for what her future autonomous judgment will be. So, for example, both standards endorse intervention to prevent feverish Arnold from walking out his third story window. The most reliable judgment standard yields this result because it is reasonable to believe that, when his fever passes, Arnold will unequivocally autonomously endorse such intervention, which is to say he will not withdraw his endorsement of intervention at any time in the future.

The example of Allen and the recreational drug does present a conflict between earlier and later autonomous judgments. In this sort of case, the hypothetical autonomous endorsement standard and the most reliable judgment standard may give different

verdicts on the justification of paternalistic intervention. Whether they do or not will depend on the details of the policy the drug prohibition represents. It would be hard to justify a drug prohibition that included draconian punishments for possession of small amounts of the drug. However, a drug prohibition requiring that users be given treatment rather than prison sentences might well be justified by the most reliable judgment standard, though it would not be justified by the hypothetical autonomous endorsement standard.

Many people are reluctant to accept that interference with Allen's decision can be justified, due to fears about how such powers could be abused. I believe this is a legitimate concern, and one that would have to be considered in determining whether such a policy could be justified all things considered. Here I continue to set aside concerns about potential abuse and potential bad side effects, so I can focus on the most important considerations of pure theory.

One way of defending the autonomous hypothetical endorsement standard would be to insist that, in a case like Allen's, intervention *is* objectionably paternalistic. If Allen himself does not regard his future judgment as more reliable than his past and present judgment, the defender of the autonomous hypothetical endorsement standard might argue, it would be objectionably paternalistic for someone else to overrule Allen's own judgment of the reliability of his future judgment and impose *their own* judgment of the reliability of Allen's future judgment on Allen. However, this is a misleading description of the basis for intervention. Just as it is reasonable to believe Allen's future self will judge that preventing the use of drug RD would have been good for him, it is reasonable to believe Allen's future self will also judge his earlier judgment opposing the prohibition to be less reliable than his own later judgment. There is a symmetry to Allen's earlier and later judgments that is illustrated in Figure 2. For every judgment made by Allen's earlier self endorsing the judgment of his earlier self, there is a corresponding judgment made by Allen's later self endorsing the judgment of his later self.

Allen's Earlier Self:	*Allen's Later Self*:
(1a) judges that a prohibition on drug RD would be bad for him.	(1b) judges that a prohibition on drug RD would be good for him.
(2a) judges that his earlier self's judgment (1a) is more reliable than his later self's judgment (1b).	(2b) judges that his later self's judgment (1b) is more reliable than his earlier self's judgment (1a).
(3a) judges that his earlier self's judgment (2a) is more reliable than his later self's judgment (2b).	(3b) judges that his later self's judgment (2b) is more reliable than his earlier self's judgment (2a).
. . .	. . .

Figure 2. The Symmetry of the Disagreement Between Allen's Earlier and Later Selves.

Resolving Conflicts Between Earlier And Later Selves

Allen's earlier self disagrees with his later self on the benefits of using the drug RD. It seems arbitrary for the hypothetical autonomous endorsement standard to favor the judgment of Allen's earlier self, simply because it is his judgment at the time of the

contemplated intervention. Perhaps it is not arbitrary. Perhaps it reflects an attitude of neutrality that we are morally obligated to take toward Allen's current and future selves. Consider again Nagel's example of the young man, call him Tom, who in his youth values sex, spontaneity, frequent risks and strong emotions, but who expects in twenty years he will value security, status, wealth, and tranquility.

I believe examples like this one strengthen the case for the most reliable judgment standard. Clearly, Tom's earlier self does not understand how it could be reasonable for him to come to value security, status, wealth, and tranquility. Tom's later self could easily understand how it could have been reasonable for his earlier self to value sex, spontaneity, frequent risks and strong emotions. Indeed, it is quite plausible to think Tom's later self would not endorse intervention to prevent his earlier self from acting on those values, because he would think it was important that his later change of values be based on his own autonomous judgment in response to his experience, not on the forcible intervention of others.

The problem is Tom's *earlier* self might well endorse a policy of intervention to prevent his *later* self from acting on his later self's values. If it were possible, his earlier self might precommit to a life of frequent risks, so as to prevent his later self from being able to avoid them. Should such precommitment strategies be legally enforceable? This is the issue raised by Parfit's example of the Russian Nobleman (1984, 327). In Parfit's example, an idealistic young nobleman wishes to be able to pre-commit to distributing his inheritance to the peasants, because he believes that by the time he receives the inheritance, his ideals may have faded and he may decide to enjoy the wealth rather than redistribute it.

The Parfit example is not directly relevant to the current discussion, because it involves a judgment about which course of action is better for everyone, rather than a judgment about what is better for the nobleman himself. To turn the example into one concerning judgments about one's own good, suppose the young nobleman believes great wealth would be bad *for him*, but he is concerned that when he receives the inheritance he will be so blinded by the self-serving reasons for enjoying it, that he will not be able to see why it is bad for him to do so. For this reason, he wishes to make an enforceable vow to turn over his inheritance to a humanitarian organization. The most reliable judgment standard can help to explain why such vows should not be enforceable. After trying out poverty for a while, the nobleman will generally be in a better position to judge how good it is for him. So his later self will be in a better position to judge whether it would be good for him to give up the inheritance. This is not to say that an earlier self should never be able to make commitments binding on a later self. There is no problem about commitments of a kind endorsed by both the earlier and later selves. However, when there is good reason to think that a later self would *not* endorse a commitment of a certain kind, that can provide a good reason for not permitting the earlier self to enter into such commitments. As I discuss below, this is the main idea needed to explain why slavery contracts, religious vows, and various other precommitment devices should not be legally enforceable, though ordinary contracts should be.

Similar reasoning applies to the example of Tom. Tom should not be able to precommit his future self to a life of frequent risks, because the fact that the more reliable judgment of his future self would not endorse such a precommitment is evidence it is not good for him.

Of course, it is only a contingent fact about human beings that their later judgments about what is good for them are generally more reliable than earlier ones. We could imagine beings who start life knowing everything there is to know about what is good for them and knowing everything they need to know to be able to make choices that would best promote their good. Over time, their cognitive capacities decline, so that their later judgments about what goals they should pursue and how they should pursue them are typically less reliable than their earlier judgments. For beings of this kind, the most reliable judgment standard would favor their earlier judgments over their later judgments. Even for beings of this kind, the most reliable judgment standard would lead to counterexamples to the necessity of the hypothetical autonomous endorsement standard, because it would justify overriding a subject's *current* autonomous judgment opposing intervention in order to give effect to an *earlier* one endorsing it.[26]

Paternalist Policies, Not Acts

It is important to emphasize that when we ask whether a future self would endorse a particular kind of paternalist intervention, we are asking about the future self's attitude toward a paternalist policy, of which the particular act is an instance. The focus on policies is important in the following kind of case: I have just bought a ticket in a lottery in which the odds of winning are 1/1,000,000. I put the ticket in my wallet. Almost surely, my future self would endorse my taking the ticket out of my wallet and throwing it away, because almost surely my future self knows that my ticket is not a winner. However, my future self would not endorse the *policy* of throwing away lottery tickets before I find out whether they have won. So my future self endorses the same policy I endorse, the policy of not throwing away a lottery ticket until I have determined whether it is a winner. For ease of exposition, if I say my future self endorses my not throwing the ticket away, I should be understood as claiming my future self endorses a policy of which the particular act is an instance. Similarly, when I say a future self endorses an act of paternalistic intervention, I should be understood as claiming the future self endorses a paternalistic policy of which the particular act of intervention is an instance.

Bilateral Future Endorsement

Paternalists often attempt to justify their intervention by saying "Someday you'll thank me for this." Call this the *future gratitude condition*. It is important to note the most reliable judgment standard's requirements for overriding a person's own hypothetical, autonomous judgment are more stringent than this. The future gratitude condition says nothing about what the target's attitude toward intervention would be if no intervention were to take place. Thus, it leaves open the sort of possibility represented in Figure 3.

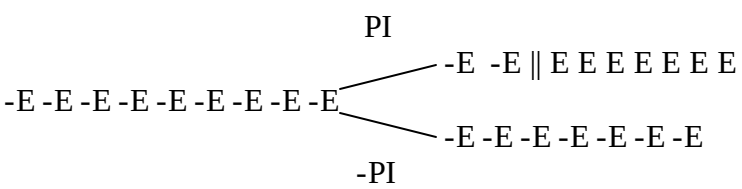

Fig. 3. An example in which the future gratitude condition is satisfied but the paternalistic intervention is not weak, because of the lack of bilateral future endorsement.

Consider the case of a proposed ban on pornography. Suppose if there is no ban on pornography, most people will never endorse a ban. Suppose also there is good evidence that if an effective ban on pornography were enacted, almost all those who currently oppose the ban would eventually become prudes who would unequivocally endorse the ban. This is the situation illustrated in Figure 3. The ban would satisfy the future gratitude condition, because it is reasonable to believe that if it were instituted, the targets of intervention would eventually come to unequivocally endorse it (represented by the top branch in Figure 3). It would not satisfy the most reliable judgment standard, because, if the ban were not enacted, those who would have been the targets of intervention never would come to endorse it (represented by the bottom branch of Figure 3).

Why does the most reliable judgment standard require both branches of the diagram to lead to unequivocal endorsement in order to be justified in overriding the earlier self's opposition to paternalistic interference? Because, in general, there is no principled basis for holding that the judgments in one branch of the diagram are more reliable than the judgments in the other branch. Each branch represents a future with different life experiences. There is no basis for claiming one branch would produce more reliable judgments than those in the other branch. So the most reliable judgment standard does not come into play unless it is reasonable to believe both branches lead to unequivocal endorsement of paternalistic intervention. When both branches lead to unequivocal endorsement, I will say the case is one of *bilateral future endorsement*. Bilateral future endorsement is represented graphically by the fact that both branches eventually lead to unequivocal endorsement of the relevant kind of paternalistic intervention. It is illustrated in Fig. 1, which I repeat here.

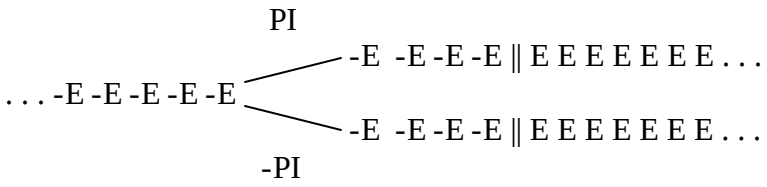

Figure 1. An Example of Bilateral Future Endorsement.

Talk of earlier and later selves is a colorful way of talking about earlier and later, actual and hypothetical, attitudes of one and the same self. The most reliable judgment standard would never diverge from the hypothetical autonomous endorsement standard unless it were possible to predict at least some future changes in attitude. To do so, we typically rely on statistical reasoning to infer what the judgment of an actual or hypothetical future self would be. In the example of Allen, information about the trajectory of the judgments of others who have used the drug and those who have not provides evidence from which we can reasonably infer Allen's hypothetical future attitudes toward intervention to prevent him from taking the drug.

If it were 100% certain Allen would come to unequivocally endorse (UE) paternalistic intervention, regardless of whether the intervention took place, that would be the strongest possible case for classifying the ban as weak paternalism. It would be unrealistic to think 100% of people would ever endorse any kind of paternalism. How

probable must it be that Allen would come to unequivocally endorse the intervention (PI) for it to qualify as weak paternalism? To answer that question, there are two probabilities that must be considered—first, the probability he would come to unequivocally endorse it, if the intervention were to take place (Prob(UE/PI), and second, the probability he would come to unequivocally endorse it, if the intervention were not to take place (Prob(UE/-PI).[27] What is the threshold value these probabilities must exceed for the intervention to qualify as weak paternalism?

Consequentialist considerations can help to answer that question. Consider this question: Which selection of a threshold value for the relevant probabilities would be the best policy for equitably promoting the life prospects of the targets of paternalistic intervention? When the question is asked in this way, it can be seen that it is of the same kind as the question about democratic decision rules discussed previously: Which democratic decision rule would be the best policy for equitably promoting the life prospects of the those bound by the rule? In a previous chapter, I suggested that although supermajorities are typically required to amend rights guarantees, simple majority rule is probably the best policy for adopting most other legislation. It seems to me that similar considerations support the conclusion that the threshold value for most weak paternalism should be .5—that is, a policy of intervention will typically qualify as weak paternalism when Prob(UE/PI) and Prob(UE/-PI) are both greater than .5.[28]

When Prob(UE/PI) and Prob(UE/-PI) are both greater than .5, it is reasonable to suppose more will benefit from the paternalism than will not. On average then, one would expect the gains from such a policy to outweigh the losses. Then it is reasonable to believe that the *policy* of enacting paternalist policies when Prob(UE/PI) and Prob(UE/-PI) are both greater than .5 would promote almost everyone's life prospects. It is important to emphasize that it is the *policy* of setting the threshold value at .5 that would be expected to promote *everyone's* (or almost everyone's) life prospects. In any particular application of the policy, some people's life prospects will be increased and some people's will be lowered. The *policy* of setting the threshold at .5 can improve everyone's life prospects, even if it is unrealistic to suppose any particular application of it would do so.

The analogy with majority rule yields another important insight about weak paternalism. Suppose a policy of legal paternalism is enacted on the grounds that Prob(UE/PI) and Prob (UE/-PI) are both equal to .6. Then we can expect that, after the policy is enacted, 60% of the target audience will come to unequivocally endorse it and 40% of the target audience will not. To the 40% who do not come to endorse it, the policy might seem like strong paternalism, but it is not. It would be strong paternalism if the justification for the law depended on the 60% majority overruling the judgment of the 40% about what was good for the 40%. In the case as I have described it, the policy is justified by the future judgment of the 60% about what is good for the 60%. The adverse effects on the other 40% are not the goal of the policy, they are simply an unfortunate *majority spillover effect*. I give other examples of the majority spillover effect below.

Unilateral Endorsement

I have identified a class of cases of weak paternalism that violate the hypothetical autonomous endorsement condition: cases where it is reasonable to believe the target's future selves would bilaterally endorse the relevant kind of intervention. Is bilateral future endorsement always necessary? Surprisingly, the answer is no. There is one

category of exceptions to the hypothetical autonomous endorsement standard where the target's future selves would *not* bilaterally endorse the relevant kind of intervention. This class of exceptions are cases in which one of the branches of the diagram terminates prematurely. See Figure 4 for a diagram illustrating such a case.

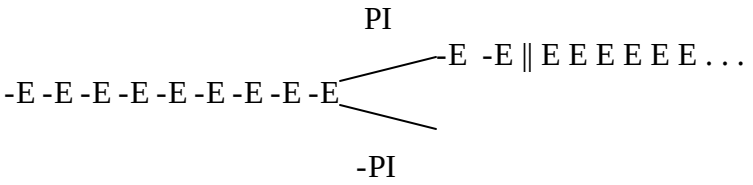

Fig. 4. Unilateral Endorsement, Because the -PI Branch is Truncated: Example of Lee the Soccer Player.

Figure 4 represents a situation in which, if there is paternalistic intervention, the target will come to unequivocally endorse it, but if there is not, the target will lose the ability to make autonomous judgments, which is represented by the premature termination of the lower -PI branch. Typically, the loss of the ability to make autonomous judgments is due to death or serious mental impairment.

Consider an example of the kind of case illustrated in Figure 4:
Lee is a 25-year-old soccer player. He fell in love with soccer when he was in elementary school. He has built his life plan around playing and eventually coaching soccer. For years he has organized his life around playing for the U.S. World Cup soccer team. He is intensely competitive. He has often told his friends that if something happened to him to prevent him from playing soccer, he would rather not go on living.

Lee was recently selected to play on the U.S. World Cup soccer team. Shortly after, on his way home from practice, he was involved in an auto collision that left him permanently paralyzed from the waist down. Having lost all hope of realizing his dreams, Lee has lost all motivation to live. He wants to commit suicide. Can involuntary commitment to prevent him from committing suicide be justified?

This case is different from the example of feverish Arnold who must be restrained from stepping out of a third story window, because, in this case, the hypothetical autonomous endorsement standard is satisfied. It is clear Lee's decision to commit suicide is an expression of his settled values and preferences and his autonomous judgment would not endorse intervention to prevent him from committing suicide in this sort of case.

Nonetheless, the case is not so clear-cut as the hypothetical autonomous endorsement standard would imply. To see why not, suppose if he is prevented from committing suicide, it is almost certain, through the autonomous exercise of his judgment, he will be able to make a life for himself that he regards as worthwhile and that he will eventually judge autonomously that intervention to prevent him from committing suicide after the accident was good for him. Suppose statistical studies have shown that, if they are prevented from committing suicide, the overwhelming majority of athletes like Lee who suffer career-ending injuries will go through a period of depression for one or two years, after which time they will put together a new life they regard as worthwhile and they will come to unequivocally endorse the intervention necessary to prevent them

from committing suicide during the temporary period of depression. To add to the force of the example, suppose if he does not commit suicide, it is reasonable to expect Lee will have forty or more years of autonomy in his new life, and during the entire forty years he will consistently endorse the intervention that prevented him from committing suicide.

This case is not one in which Lee's future selves would bilaterally endorse the relevant kind of intervention, because if Lee commits suicide, there will be no future self to endorse or fail to endorse anything. This fact is represented in Figure 4 by the fact that the -PI branch of the diagram terminates prematurely.

Because it does not involve bilateral future endorsement, the paternalism in this case seems to me to be more difficult to justify than the paternalism to prevent Allen from taking the drug RD. Ultimately I do believe suitably humane intervention to prevent Lee from committing suicide can also be justified, and justified for the same kind of reason it is justified in Allen's case. It seems to me it is reasonable to take Lee's future judgment when he is prevented from committing suicide (the PI branch of the diagram in Figure 4) to be more reliable than his current judgment in favor of suicide. His current judgment may be that it is not worth the pain of rebuilding his life to get to the point where he has new values to pursue. However, his later self will have lived through the transformation and will be in a better position to evaluate whether it was worth enduring. If so, then the import of the missing second branch in Figure 4 is that there is no other equally reliable judgment that conflicts with the hypothetical later judgment endorsing the paternalistic intervention, so the intervention satisfies the most reliable judgment standard as weak paternalism.

A second complicating factor in the Lee example is that, if the intervention occurs, Lee will undergo a transformation in values so radical it might seem misleading to think of the later Lee as the same person as the earlier one. Perhaps it is more accurate to think of the earlier and later Lee as so different that allowing the later Lee's judgments to overrule those of the earlier Lee about what is good for him would be objectionable in the same way that allowing the judgments of another person to overrule Lee's own judgments about what is good for him would be.

Advocates of continuity theories of personal identity (e.g., Parfit 1984) would find this sort of response particularly compelling. This seems to me to be a mistake. Even large changes in values can be appropriate, when the occasion warrants a large change. Part of being a person is to be capable of such transformations when appropriate. It seems to me that Lee's situation is just such an occasion. So it seems to me that Lee's situation is one in which it would be reasonable to expect one and the same person to undergo a large transformation.

If I am right that the most reliable judgment standard would categorize a suitably humane policy of intervention to prevent Lee from committing suicide as weak paternalism, then it is not necessary for the target's future selves to bilaterally endorse intervention that violates the hypothetical autonomous endorsement standard for it to be justifiable. Bilateral future endorsement is required when there are two equally reliable hypothetical future selves. However, when the –PI branch terminates prematurely, then the lack of a second branch makes the judgment on the remaining branch authoritative.

Earlier I suggested that a right against legal paternalism should include some kind of right to assisted suicide. The example of Lee the soccer player shows this right should not be unconditional. A prohibition on assisting Lee in committing suicide qualifies as

only weak paternalism, because it is reasonable to believe that, if Lee is prevented from committing suicide, he will come to unequivocally endorse the policy of preventing suicides in such cases. Prohibitions on assisted suicide qualify as strong paternalism when it is not reasonable to expect that the paternalistic intervention would later come to be unequivocally endorsed by the target of the intervention. For example, if a person facing a slow and painful death asks to have her death hastened, intervention to prevent the hastening of her death cannot generally be justified by any reasonable prospect of her coming to unequivocally endorse it. So such intervention would qualify as strong paternalism, and a right against strong legal paternalism would provide protection against such intervention. A right to assisted suicide would have to be fashioned so as not to apply to cases such as Lee's.

Against Euthanasia

In Figure 4, the -PI branch of the diagram is truncated. Symmetry considerations suggest there is another kind of exception to the hypothetical autonomous endorsement standard where bilateral future endorsement is not required, when the PI branch is truncated. See Figure 5.

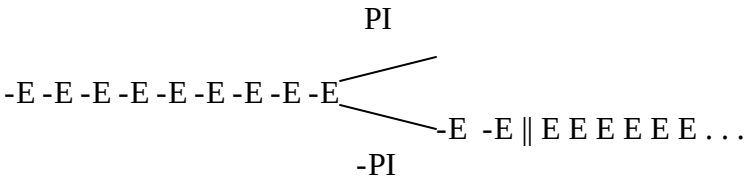

Fig. 5. The PI branch is truncated. Paternalistic Intervention is not Weak.

Figure 5 represents a case in which paternalistic intervention renders the target incapable of autonomous judgment. Thus it represents paternalistic intervention that kills or severely mentally impairs the target of the intervention, which the target does not endorse at the time of the intervention. Although Figure 5 may represent a logical possibility, I do not include it as a category of weak paternalism, because I do not believe there are any examples that fit it.

To explain why not, I will try to construct one. Consider, for example, euthanasia. I want to limit the example to cases of euthanasia where the potential target autonomously judges at the time of the intervention that being killed would not be good for her and on that basis does not endorse being euthanized. To fit the diagram in Figure 5, there would have to be an example in which it is reasonable to believe that a potential target of euthanasia might autonomously object to being euthanized now, but if euthanasia were not performed, would later come to autonomously endorse having been euthanized earlier. How could there be such an example? It is always possible that sometime in the future the target might change her mind about being killed and ask to be killed. Perhaps she finally becomes convinced there is no other alternative to suffering a prolonged, painful death. Even then she would only be endorsing assisted suicide. She would not be endorsing euthanasia, because she would not be endorsing being killed against her will. Since I cannot think of any plausible scenario on which reasonable people would autonomously endorse being killed (or being severely mentally impaired)

for their own good against their will, I do not believe there are any exceptions to the hypothetical autonomous endorsement standard that fit the diagram in Figure 5.

I should add that I do not mean to suggest the considerations raised here are the only reasons to oppose euthanasia. There are many reason to oppose it. It is important to note that one of the reasons for opposing euthanasia is euthanasia always constitutes strong paternalism, and autonomous adults should be free of strong paternalism.

The cases discussed so far are exceptions to the claim that the hypothetical autonomous endorsement standard is a necessary condition for weak paternalism. I turn now to a briefer discussion of exceptions to its sufficiency.

The Hypothetical Autonomous Endorsement Standard is not a Sufficient Condition for Weak Paternalism

Exceptions to the sufficiency of the hypothetical autonomous endorsement standard will be cases in which the target of paternalistic intervention autonomously endorses the relevant kind of intervention, but the intervention is not justifiable. The previous discussion of cases of bilateral future endorsement (e.g., the example of Allen and the drug RD) immediately suggests a parallel class of cases of bilateral future non-endorsement, as illustrated in Figure 6.

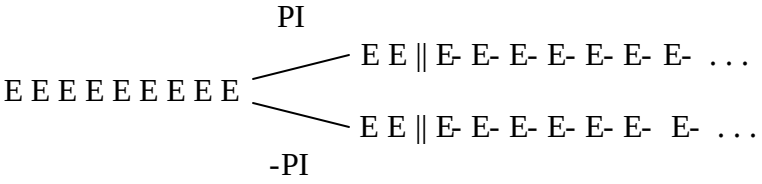

Fig. 6. Bilateral Future Non-Endorsement of Paternalistic Intervention That the Earlier Self Endorses. The target of the potential intervention would autonomously endorse it (E) at the time of the potential intervention, but it is reasonable to believe the target would come to endorse non-intervention (E-), both if the intervention were to take place and if it were not to take place.

Figure 6 illustrates the dual of the example of Allen, illustrated in Figure 1 above. Suppose Bill autonomously endorses paternalistic intervention to prevent him from taking drug HD. It would seem that such intervention would be weak paternalism, because it would satisfy the hypothetical autonomous endorsement standard. However, suppose, in addition, good statistical evidence makes it reasonable to believe that if the intervention takes place, Bill's future self will come to unequivocally endorse non-intervention and that if the intervention does not take place, Bill's future self will also come to unequivocally endorse non-intervention. As before, talk of future selves is only a heuristic for talking about Bill's actual and hypothetical attitudes. The heuristic can be replaced by setting a threshold for the probability that Bill would come to autonomously and unequivocally endorse non-intervention, if the intervention were to take place (Prob(UE-/PI), and the probability that Bill would come to autonomously and unequivocally endorse non-intervention, if the intervention were not to take place (Prob(UE-/-PI). For reasons discussed above, I believe the threshold value for these probabilities should be .5.

Continuing the symmetry with the earlier discussion of necessity, there are even some exceptions to the sufficiency of the hypothetical autonomous endorsement standard involving only unilateral future non-endorsement. They are the cases illustrated in Figure 7.

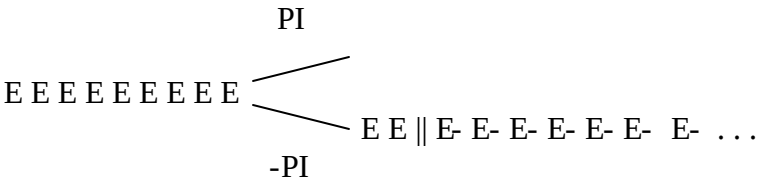

Fig. 7. Unilateral Future Non-Endorsement, Because the PI Branch is Truncated: The Example of Albert.

Figure 7 illustrates a kind of case very similar to the example of Lee the soccer player, illustrated in Figure 3 above. Albert wants to commit suicide but lacks the nerve to do it. Albert autonomously endorses your assisting his suicide. Suppose you have good reason to believe that if you do not assist his suicide, Albert will come to unequivocally endorse not intervening in such cases. This case is more clear-cut than the example of Lee the soccer player. It would be wrong to assist Albert to commit suicide.

There is one further logical possibility of a case with a truncated branch. This kind of possibility is illustrated in Figure 8.

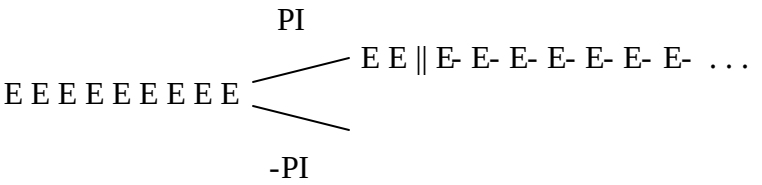

Fig. 8. Unilateral Future Non-Endorsement Because The -PI Branch is Truncated: Example of Florence

I do not believe there are any realistic examples fitting Figure 8. To see why not, I will try to construct one: Florence will die unless there is paternalistic intervention to keep her alive. Her autonomous judgment endorses the intervention. However, later in her life she will become miserable and she will autonomously and unequivocally judge she would be better off dead. Could she come to endorse a policy of allowing her to die at the earlier time, when she wanted to be kept alive? I don't see how she could reasonably do so. The only policy it would be reasonable for her to endorse is the policy of keeping her alive when she wants to be kept alive and not keeping her alive when she does not. So I do not see how Figure 8 could generate in realistic examples of weak paternalism.

Varieties of Weak Paternalism

On the most reliable judgment standard, paternalism is weak when it is reasonable to believe it is part of a policy that would be endorsed by the target's most reliable autonomous judgments. On the assumption that a person's later autonomous judgments

are generally more reliable than her earlier ones, the most reliable judgment standard explains why the hypothetical autonomous endorsement standard is neither a necessary condition nor a sufficient condition for weak paternalism. The previous discussion of exceptions to the hypothetical autonomous endorsement standard provides a useful framework for classifying the types of weak paternalism. I begin with an account in terms of what it is reasonable to believe about the target's future endorsement or non-endorsement and then I translate it into an account in terms of what it is reasonable to believe the relevant conditional probabilities to be.

(Weak Paternalism—Preliminary Account) An act of paternalistic intervention is *weak* just in case:

(a) The target would autonomously endorse the relevant policy of intervention and neither of the following two exceptions holds: (i) bilateral future non-endorsement (e.g., Fig. 6) or (ii) unilateral future endorsement because the –PI branch is truncated (Fig. 7); or

(b) The target would not autonomously endorse the relevant policy of intervention, but one of the following two exceptions holds: (i) bilateral future endorsement (e.g., Fig. 1) or (ii) unilateral future non-endorsement because the –PI branch is truncated (Fig. 4).

The preliminary account can be generalized by employing the relevant conditional probabilities:

(Weak Paternalism—Final Account) An act of paternalistic intervention is *weak* just in case:

(a) The target would autonomously endorse the relevant policy of intervention and neither of the following two exceptions holds: (i) it is reasonable to believe both Prob(UE-/PI) and Prob(UE-/-PI) are greater than the threshold value for weak paternalism (typically .5) or (ii) it is reasonable to believe the –PI branch will be truncated and Prob (UE-/PI) is greater than the threshold value for weak paternalism (typically .5); or

(b) The target would not autonomously endorse the relevant policy of intervention, but one of the following two exceptions holds: (i) it is reasonable to believe both Prob(UE/PI) and Prob(UE/-PI) are greater than the threshold value for weak paternalism (typically .5); or (ii) it is reasonable to believe the –PI branch is truncated and Prob(UE/PI) is greater than the threshold value for weak paternalism (typically .5).

Ideally Reliable Judgments

It is useful to consider some potential misunderstandings and objections. The most reliable judgment standard justifies overruling a person's less reliable earlier judgments by her more reliable later hypothetical ones. Once hypothetical judgments can be entertained, why not entertain even more reliable hypothetical judgments than the judgments people can ever make? For example, undoubtedly people's judgments about what is good for them would be more reliable if they were omniscient. Omniscience would be the highest ideal of reliability. In evaluating a policy of paternalistic intervention, we could ask: Would the target endorse the policy if she were omniscient? If so, because this hypothetical endorsement would be more reliable than her current or future judgment, the most reliable judgment standard would justify using it to overrule the target's current and future judgments.

There are two problems with this sort of suggestion. The first is the obvious one that we are not omniscient so we have no way to know what the target's judgment would be if she were omniscient. This objection is not decisive, because it might have been the case that fallibly trying to apply this standard would have been a good way of promoting people's life prospects. Thus, the second objection is the most important one. The second objection applies to all ideally reliable judgment theories. The track record of such theories is abysmal. Berlin (1969) has reminded us of the awful things that have been done to people on the grounds of an idealized theory about what is good for them. Idealized theories are not good ways of promoting the life prospects of the targets of intervention, because they are vehicles for overriding judgments the target herself makes or is expected to make with hypothetical judgments the target never could make, because she is not omniscient or otherwise ideally rational. Because the policy of basing paternalistic intervention on such theories is such a disastrous one, it cannot be given a consequentialist rationale. For this reason, ideally rational judgment theories cannot be used to justify legal paternalism. The only hypothetical judgments that can justify paternalistic intervention are the autonomous judgments the target herself, with all her cognitive and other limitations, would make, if she had the opportunity to do so.

Most Legal Paternalism is not Weak

Because my main focus in this paper has been on the categories of weak paternalism that cannot be accounted for by Feinberg's voluntariness standard or by the hypothetical autonomous judgment standard, I may have given the mistaken impression that the most reliable judgment standard classifies lots of paternalism as weak. Actually, very little paternalism qualifies as weak paternalism under the most reliable judgment standard. There are two reasons for this. First, our later selves are typically very conservative about endorsing policies of paternalistic intervention in our choices. Even when they rue what they take to have been a mistaken earlier decision, most people judge it is better that they be allowed *to make mistakes and to learn from them* than that outsiders interfere to prevent them from making what they will later judge to have been a mistake. It is usually only in cases where mistakes can be expected to cause severe, unavoidable losses that most people would endorse intervention to be prevented from making them.

The second reason very little paternalism qualifies as weak paternalism under the most reliable judgment standard is that, in most cases, a person's current autonomous judgment about a policy of paternalistic intervention is the best evidence we have of what her future autonomous judgment will be. As the example of Allen and the recreational drug RD illustrates, typically the only way to be epistemically justified in believing a person's future self will overrule her current judgment is on the basis of statistical reasoning from evidence about how the judgments of other people in relevantly similar situations have changed over time. We simply do not now have much evidence of this kind. I believe it would be useful to obtain more.

Why A *Right* Against Strong Legal Paternalism?

In a democracy, a paternalistic law can be enacted by a majority. This is as it should be where a majority endorses the law because of the benefits of paternalistic intervention in their own case, and not because the majority thinks the law will be good for a minority who do not endorse it themselves. In the former case, a minority may be

bound by a law it does not endorse, due to the majority spillover effect. I illustrate this possibility shortly. Such laws do not violate any right of the minority.

In the latter case, a majority uses the legal system to give effect to what it believes is good for *other people*. Because in a democracy majorities have the power to enact laws, so long as they are constitutional, only a constitutional right to protection against strong legal paternalism can effectively protect minorities against strong legal paternalism favored by a majority. For this reason, there is no other way to protect a minority against strong legal paternalism than by incorporating a right against strong legal paternalism into the constitution.

It is important to note that the consequentialist considerations favoring a right against strong legal paternalism do not imply weak legal paternalism is always justified, all things considered. To focus the discussion on the pure theory of paternalism, I temporarily set aside concerns about government abuse of or negative side effects of paternalistic policies. Even if a law qualifies as weak legal paternalism, if the potential for abuse is great enough or if its negative side effects are bad enough, the law will not be justifiable on consequentialist grounds. Thus, being weakly paternalistic is only a necessary, not a sufficient, condition for being justifiable on consequentialist grounds.

To the credit of the U.S. Supreme Court, it has identified a constitutional right to a sphere of autonomy free of some legal paternalism. The most reliable judgment standard provides the dividing line between weak and strong paternalism necessary to be able to articulate a right against strong legal paternalism. Probably a constitutional amendment would be required to establish the full right. In the next part, I consider the consequences of adopting such a constitutional amendment on existing laws.

PART III. LEGAL APPLICATIONS OF THE MOST RELIABLE JUDGMENT STANDARD

Suppose there were a constitutional right against strong legal paternalism, as defined by the most reliable judgment standard. What difference would it make to existing laws? I begin by considering cases where the most reliable judgment standard might reasonably yield different results than the hypothetical autonomous endorsement standard. Then I take up cases where they yield the same results.

Suicide And Assisted Suicide

In the previous section, I regretted the fact that the U.S. Supreme Court had declined the invitation in the Philosopher's Brief to declare such a right and to extend it to some cases of suicide and assisted suicide. The most reliable judgment standard provides a framework for explaining which cases of suicide and assisted suicide should be protected.

When a person has a terminal illness and has no reasonable prospect of any future free of pain or greatly diminished consciousness or when a person has Alzheimer's and the disease has progressed to the point that she is on the verge of losing her memory of her past and her family and friends, she should be permitted to decide to end her life free from paternalistic interference and she should be able to obtain assistance in ending her life painlessly and with dignity. Because there is no reasonable prospect of a transformation into an autonomous future self who would find her life worthwhile and would endorse her having been prevented from committing suicide, these cases are not cases in which intervention would satisfy the most reliable judgment standard. I am not advocating euthanasia in such cases; I am only advocating giving effect to the person's own most reliable judgment about what is best for her.

As the example of Lee the soccer player illustrates, the case is very different if a suicide would cut off a life the person living it would come to regard as worthwhile. Where it is reasonable to believe intervention to prevent suicide would give a person years or even decades of a life her hypothetical future self would regard as worthwhile, intervention to prevent the person from committing suicide can qualify as weak paternalism under the most reliable judgment standard. If the intervention is done in a humane way, it is reasonable to think the hypothetical future self would endorse it. Such intervention could be justified on consequentialist grounds, all things considered, if the policy of intervening did not risk being abused and did not have other negative side effects that outweighed the beneficial effects.

Slavery Contracts And Religious Vows

Slavery contracts raise different issues from typical cases of suicide or assisted suicide. A closer analogy would be entering into a contract to commit suicide. Consider, first, slavery contracts, where a person who would not willingly choose to be a slave contracts to become one in order to obtain subsistence. This example raises a puzzle that stumped Mill. Here is what he says about it: "[B]y selling himself for a slave, he abdicates his liberty. He foregoes any future use of it, beyond that single act. He therefore defeats, in his own case, the very purpose which is the justification of allowing him to dispose of himself. . . . The principle of freedom cannot require that he should be free not to be free"(1859, 116) Mill's argument is quite eloquent, but we should suspect it is unsound because of an implication that Mill himself draws from it: "These reasons, the

force of which is so conspicuous in this particular case, are evidently of far wider application; yet a limit is everywhere set to them by the necessities of life, which continually require, not indeed that we should resign our freedom, but that we should consent to this and the other limitation of it."(116)

There is something seriously wrong with Mill's premises if they imply that *any* contractual limitation on liberty is objectionable, because joint contractual commitments play an essential role in making possible joint improvements in well-being. Even if all of the necessities of life were provided for us, contractual commitments would still make possible great improvements in well-being. The consequentialist benefits of ordinary contracts are so obvious that Feinberg (1971) reasonably reinterpreted Mill's own argument as a *modus tollens* argument for the rejection of one or more of the premises of his argument against slavery contracts.

The most reliable judgment standard provides a basis for reconstructing Mill's argument against slavery contracts without calling into question the justifiability of most other contracts. To make the argument, it is useful to consider separately two different kinds of slavery contracts:

(1) Antecedently unwilling slaves. The first kind covers slavery contracts entered into by those who have no antecedent intention to be a slave. Someone who had no antecedent intention to be a slave might well be motivated to enter into a slavery contract to avoid starvation or great deprivation. Governments should refuse to enforce such contracts, but the grounds are not paternalistic. Refusing to enforce a slavery contract of this kind is the first step toward a legal minimum wage. As I discussed in a previous chapter, minimum wage laws solve a CAP for unskilled workers.

Minimum wage laws assure that workers do not have to accept slavery contracts in order to get a job. What about those who are unable to find a job? As I discussed in a previous chapter, consequentialist considerations will require a government to make sure that none of its citizens suffer starvation or other severe deprivation because they are unable to find work. No one will have to agree to be a slave to avoid starvation or other severe deprivation.[29]

Thus, the first difference between slavery contracts and most other contracts is that prohibiting slavery contracts would be expected to solve a CAP for those who would otherwise enter into them. To say the prohibition solves a CAP is to say it makes those people better off than they would be without the prohibition. This is obviously not true of most other contracts. Prohibitions on most other kinds of contracts would make the potential contractors worse off, not better off.

(2) Willing slaves. The preceding analysis applies to people who have no antecedent intention to be slaves. What about those who want to be slaves and need no monetary or other inducement to motivate them to enter into a slavery contract? Prohibiting this sort of slavery contract is not a solution to a CAP. It is this sort of case that persuaded Feinberg (1971) and, ultimately, G. Dworkin (1983) that there is no principled objection to enforcing slavery contracts. It is easy to see why Feinberg would reach this result. If we apply the hypothetical autonomous endorsement standard to the policy of refusing to enforce slavery contracts, we easily see that those who autonomously choose to enter into slavery contracts would not endorse a policy of non-enforcement at the time of entering into the contract. Such a policy would thwart their

intention. So no such policy could be justified by the hypothetical autonomous endorsement standard.

However, the most reliable judgment standard yields a different result. We know enough about the life of a slave to know there is a very high probability that willing slaves who learned more about what life as a slave was like would eventually change their mind about wanting to be a slave. It is reasonable to think that their later selves would want to be able to terminate the arrangement. For these reasons, it is reasonable to think it very probable their later selves would endorse a policy of refusing to enforce slavery contracts. For the same reasons, if such contracts were enforceable, it is reasonable to project that those willing slaves who entered into slavery contracts would also come to endorse refusing to enforce them. It would be a mistake for the government to enforce agreements made by their earlier, less reliable, selves when it is reasonably foreseeable that their later, more reliable, selves would endorse not enforcing them.

The most reliable judgment standard would thus favor a policy of refusing to enforce such contracts. Note that no such argument can be made for the enforcement of contracts generally. Almost everyone at one time or another enters into a contract s/he later regrets. However, practically no one would ever endorse a general policy of refusing to enforce contracts. The results would be disastrous.

The most reliable judgment standard draws our attention to a small category of contracts it is reasonable to expect a later self will want to be unenforceable. If slavery contracts were the only such contracts, the category would be of little practical interest, because there are not very many people who want to be slaves. However, there are examples of people who autonomously choose to give up a large amount of freedom. Members of some religious orders often take vows of celibacy, poverty, and obedience. The combination of these vows amounts to a great loss of freedom.

As in the case of slavery, there are two reasons for not making such vows enforceable. The first is that doing so solves a collective action problem, because it is better for members of religious orders to be able to obtain the benefits of membership without giving up their option to leave the order. The second is that their judgment of the value of a life of celibacy, poverty, and obedience is likely to be more reliable after they have lived it for a while. It would be a mistake to enforce the agreements of their less reliable, earlier selves when it is reasonably foreseeable that their more reliable, later selves would oppose enforcement. Thus, a government should allow people to enter into a religious order, but it should not permit them to enter into enforceable promises of perpetual celibacy, poverty, and obedience that will prevent their more reliable later selves from reconsidering their decision.

Although the law permits those who make vows of chastity and obedience to break their vows if they later change their mind, vows of poverty are not so easily reversible. A person who changes her mind about a vow of poverty is free to acquire property in the future, but she will not generally be able to recover the property she surrendered at the time she took the vow. Because a system of property rights that permits owners to sell or give away their property so greatly enhances everyone's life prospects, it can be justified on consequentialist grounds even though it provides no recourse for donors who later regret their donations. What could not be justified on consequentialist grounds is a system that enforced a vow not to acquire property after the person who made the vow no longer wanted to be bound by it.

Thus, though Mill himself failed to articulate it, there is an important distinction between slavery contracts and religious vows, on the one hand, and most other contracts, on the other. The most reliable judgment standard provides a principled reason for enforcing the latter, but not the former.

Addictive Drugs

The question of whether bans on addictive drugs can be justified is a complicated one. The harms from drugs are not solely harms to those who use them. Drug users often neglect their children. Under the influence of certain drugs—for example, alcohol—drivers are much more likely to kill and injure others in auto accidents. On the other hand, even if a ban were justifiable in theory, experience has shown that there are serious practical problems in implementing a ban. I wish to temporarily set aside all such issues, to focus on what seems to me to be the most interesting theoretical issue raised by the existence of addictive drugs. Consider the case of drug users without children or others to whom they owe duties of support. Suppose their drug use would not lead to increases in auto accidents or other sources of death and injury to others. Under these circumstances, could a ban on some or all addictive drugs qualify as weak legal paternalism?

Feinberg's (1971) account would yield a negative answer, so long as the decision to use the drugs was fully voluntary in the sense of expressing the agent's settled values and preferences. Those who would justify paternalistic intervention typically point to the addictive characteristics of the drugs, and argue that a person should not be free to compromise her future autonomy in the way an addict's autonomy is compromised.[30]

However, the mere fact that drugs are addictive could not justify paternalistic intervention to prevent people from experimenting with them. If the people experimenting with a drug know it is addictive, then when they judge that experimenting with it is a good thing to do, they are implicitly judging that the addiction is not so bad as to outweigh the good. It would be an objectionable form of paternalism to insist all addictions are bad. On the contrary, it is plausible to think some addictions, for example, an addiction to exercise, might be good for some people. Perhaps they know enough about themselves to know they would not get enough exercise if they were not addicted to it.

If addiction per se is not always bad, what could be the basis, at least in theory, for paternalistic intervention to prevent people from experimenting with addictive drugs? The answer is given by the most reliable judgment standard. There might be good reason to think the experimenters' hypothetical future selves would endorse a policy of intervention.

In the previous section, I discussed the example of Allen and the recreational drug RD. Suppose it is reasonable to expect the future potential users of RD such as Allen would come to unequivocally endorse a drug prohibition, both in the case in which there were such a prohibition and in the case in which there were not. Then the prohibition would satisfy the most reliable judgment standard. It would qualify as weak paternalism.

It is difficult to apply the most reliable judgment standard to actual drug prohibitions, because of the lack of information about the retrospective attitudes of drug users to their drug use. I believe it would be very useful for the government to gather and to disseminate this sort of information on each type of drug, giving the percentage of users who come to regret ever having used it. It would be necessary to categorize their

reasons for regret, so that it was possible to distinguish between the effects of the drug itself and the effects of its being illegal. For example, convicted drug users would surely regret their time in prison. This would only be a reason to regret the drug's being illegal, not to regret the use of the drug itself. If this sort of information were available, it might lead to more informed decisions by potential drug users and perhaps, ultimately, to revisions in the drug laws to eliminate prohibitions on drugs users don't regret using. Unfortunately, the U.S. government would never publish such information, because it would be interpreted as an implicit endorsement of some drugs over others.

Because of the lack of accurate statistical information concerning the retrospective attitudes of drug users, there is no way to know if any existing drug prohibitions could be justified as weak paternalism under the most reliable judgment standard. It is important to emphasize, however, that even if such a prohibition qualified as weak paternalism, that does not imply it could be justified on consequentialist grounds, all things considered. The all things considered justification would have to consider the potential for abuse and the potential for negative side effects. It is hard to find a policy that has had worse side effects than drug prohibitions have had in the U.S.

Thus, when all the relevant factors are considered, it is probably true that the explicit autonomous endorsement standard, Feinberg's voluntariness standard, the hypothetical autonomous endorsement standard, and a consequentialist account based on the most reliable judgment standard would all agree on the conclusion that the existing drug prohibitions in the U.S. cannot be justified. But there is a big difference between my account and the other three. On the other accounts, that people should be free to make choices that, judged by their own future selves, will ruin their lives is one of the imperatives of human freedom. To me, it is sad, regrettable, and dismaying that we have no effective means to save people from making choices that they themselves will very probably come to judge to have ruined their lives.

Other Vice Laws

The most important examples of unjustified strong paternalism in the U.S. today are the various vice laws, including laws against gambling, prostitution, and pornography. I focus on laws that prohibit private behavior, not public behavior that might cause offense to others. Of course, even private "vices" have some negative impacts on others, but I temporarily set aside justifications that appeal to effects on others. The question to be addressed is whether such laws could be justified paternalistically, on the grounds that they prevent the people who engage in such behavior from harming themselves.

My discussion of addictive drugs provides a framework for discussing these other vice laws. In each case, a prohibition would only qualify as weak legal paternalism if enough of those who engaged in the vices would eventually come to endorse it. No plausible case of this kind could ever be made for bans on any kind of pornography. Could such a case be made for a ban on gambling?

In many ways, gambling resembles an addictive drug. For some people, it has much the same effect on their lives as alcohol has on the life of an alcoholic. Nonetheless, there would never be anything like a majority of gamblers who would come to judge that their lives would have been better if they had never gambled. So I see no reasonable prospect that a ban on gambling would qualify as weak legal paternalism under the most reliable judgment standard.

What about prostitution? It would be useful to know what percentage of prostitutes and what percentage of their clients later come to regret their decision.[31] I doubt a large percentage of prostitutes' clients later regret their decisions, so I focus on the prostitutes themselves. The case is complicated by the fact that many of the bad things that befall prostitutes are due to the fact that it is illegal. For example, prostitutes typically are exploited by their pimps and subject to physical abuse, because they cannot seek protection from the police without implicating themselves in illegal activity. They need their pimps to provide the protection they cannot obtain from the police. Also, many of the health problems of prostitutes could be eliminated or greatly attenuated by legalization combined with strict regulatory supervision. The evidence from places where prostitution is legal indicates that those who engage in it can come to regard it as simply one among many occupations. There is no evidence that significant numbers of those who choose it later regret doing so.

This makes prostitution a good test case for views on paternalism. For many people, the fact that, where prostitution is legal, prostitutes can come to view their job as simply one occupation among many is itself just one more bad effect of making prostitution legal. Of course, in a rights-respecting society, those people would be free to try to persuade young women and young men not to become prostitutes. I myself would regard it as a good thing if all women and all men had enough other occupational choices that no one chose to become a prostitute. However, I see no acceptable paternalist justification for banning prostitution.

Recognizing the difficulty of justifying vice laws on paternalistic grounds, prosecutors and judges defending such laws tend to defend them by appeal to the "secondary effects" of gambling, prostitution, etc. that involve genuine harm to others. These secondary effects include other crimes committed by gamblers and prostitutes and pimps. Most of these secondary effects are effects of the laws prohibiting these activities, not the activities they prohibit. If "secondary effects" were really the problem vice laws were aimed at addressing, prohibitions would be replaced with reasonable regulation.

Misogynist Pornography

I considered and rejected non-paternalistic justifications for prohibiting misogynist pornography in a previous chapter. For me, misogynist pornography represents the most compelling challenge to my defense of a right against strong paternalism, because a good case can be made that misogynist pornography promotes a bad way of thinking about women. But even here, it seems to me, persuasion is a better policy than prohibition. The case against bans on misogynist pornography is strengthened by the fact that, as I discussed in a previous chapter, the power to ban pornography is inevitably abused. In addition, laws against pornography are another example of laws that define a crime without victims. Enforcing such prohibitions often does more harm than good.

One of the mistakes sometimes made by opponents of misogynist pornography is to think that someone who favors tolerance must be committed to thinking there is nothing wrong with it.[32] Nothing could be further from the truth. The advocate of tolerance may think there are very good reasons for people not to be prevented from reading misogynist pornography, if they want to, while thinking they ought not to want to. The advocate of tolerance simply insists that those who believe people ought not to

read misogynist pornography ought to try to persuade others not to read it, not use coercion to prevent them from doing so.

Forced Medical Care

Another important issue of legal paternalism is the question of whether it is justifiable to force a life saving medical procedure on someone who rejects it. The classic example of a case of this kind is forcing a blood transfusion on a normal adult Jehovah's Witness or Christian Scientist against her will. The hypothetical autonomous judgment standard would require us to defer to the patient's wishes, and this has been the rule enforced by the Courts.[33] I believe it is the rule that, in our present state of knowledge, should be enforced. I do not believe intervention could be justified simply because those who are not Jehovah's Witnesses or Christian Scientists thought it would be good for the patient to have the blood transfusion.

At least hypothetically, we can identify evidence that I believe might qualify such intervention as weak paternalism. Suppose there had been many cases in which normal adult Jehovah's Witnesses or Christian Scientists had been given life saving blood transfusions. It would be useful to know how they evaluated their lives after the transfusion. Consider the two most extreme outcomes. In the first, those who received the transfusion all felt their bodies had been polluted. They became despondent and lost their desire to go on living. The most reliable judgment standard would not be satisfied in such a case, and to continue to force them to undergo such transfusions would be strong legal paternalism.

Suppose the evidence were different. Suppose those who were forced to undergo such transfusions went on to lead happy lives they themselves regarded as worthwhile. The case would be complicated if, nonetheless, they still insisted it would have been better for them if they had never received the transfusion. Suppose they do not. Suppose they come to judge it was good for them to have been forced to receive the transfusion. Then this example would resemble the example of Lee the soccer player, illustrated in Figure 4 above. The most reliable judgment standard would be satisfied and forced blood transfusions would qualify as weak legal paternalism. Whether or not laws requiring such transfusions would be justified on consequentialist grounds, all things considered, would depend on the potential for abuse and the potential negative side effects.

Mandatory Motorcycle Helmets And Seatbelt Use

It is hard to know what to say about mandatory seatbelt and motorcycle helmet laws. Motorcycle helmet laws, like rules requiring hockey players to wear helmets, seem to be a solution to a macho status CAP. However, not every macho status CAP should be solved by laws. For the reasons discussed in a previous chapter, at the very least, those who ride motorcycles without helmets should be required to carry disability insurance to pay for their care if they suffer a severe brain injury, because if they have no resources to pay for their care, society's obligation to provide baseline rights will require society to foot the bill. Perhaps requiring helmets is easier to enforce than requiring disability insurance.[34]

Mandatory seatbelt laws also may be a solution to a macho status CAP. I suspect many people who would not otherwise fasten their seatbelts would endorse the laws giving them the extra impetus to do so. For these people, seatbelt laws qualify as weak legal paternalism. What about those who do not wish to be forced to fasten their seatbelts? Here there seems to me to be a greater potential for the most reliable judgment

standard to be useful. It would be useful to track how the opinions of those who do not use seatbelts change over time. Those who die or are seriously disabled due to lack of a seatbelt can probably be counted as hypothetically favoring the law. If a substantial majority of those who do not suffer death or serious disability eventually came to think seatbelts were worthwhile and to endorse the law, then the most reliable judgment standard would be satisfied, and the law would qualify as weak legal paternalism. What if the most reliable judgment standard is not satisfied? It does not seem that forcing them to fasten their seatbelts could be justified by the majority spillover effect, because it would be possible to make an exception for those who do not endorse the law. Perhaps those who objected to the law could register as conscientious objectors and receive a special sticker to put on their car to indicate their objection to it.

There is one additional reason for making seatbelt use mandatory. Drivers of cars with airbags might think the seatbelt is unnecessary. This is a mistake. Research shows seatbelts are much more effective than airbags in saving lives.[35] Thus, making seatbelts mandatory could be a way of giving effect to the judgment of those who mistakenly think seatbelts are unnecessary in a car equipped with airbags.

Minor's Rights And Paternalism Toward Children

It is easy to justify paternalism toward children on consequentialist grounds. For example, truancy laws are paternalistic toward children. The most reliable judgment standard explains how such paternalism would qualify as weak legal paternalism. Suppose we are epistemically justified in believing that if children are guaranteed such rights, their hypothetical future autonomous selves will endorse the enforcement of those very rights. Then enforcing those rights, even against the will of the child, would qualify as weak legal paternalism under the most reliable judgment standard, because a child's judgment about what is good for him is generally much less reliable than the judgment of her autonomous future self. Almost all adults would endorse their having been forced to go to school. So the most reliable judgment standard explains why such paternalism is only weak legal paternalism.

Legally Established Authorities

So far, I have focused on the differences between the most reliable judgment standard and the hypothetical autonomous judgment standard. There are many cases where their verdicts overlap. In this section I consider one important category of such cases, those involving deference to legally established authorities.

Thus, for example, I do not regard myself as a reliable judge of hazardous products (e.g., tires). I am willing to defer to the judgment of others who are authorities on these questions to keep off the market products the authorities deem to be hazardous. Typically, I do not explicitly designate them to be authorities for me. I may vote for a candidate for president, expecting him/her to make responsible appointments to the Consumer Product Safety Commission (CPSC), who in turn will hire appropriate staff personnel to make reliable safety determinations on potentially hazardous products. In such a case, I do not believe my vote for president constitutes an explicit delegation of authority to the CPSC. The case is even clearer if my candidate for President loses. If the winning candidate makes reasonable appointments, I may regard the CPSC as an authority on product safety and I may judge that it is good for me that the CPSC keeps off the market products deemed unsafe, even though I voted against the candidate who appointed the members of the CPSC, and thus did not in any way explicitly delegate the

CPSC this authority. If I regard the CPSC as an authority on product safety, regardless of whether I explicitly authorized them to do so, the CPSC could be limiting my freedom (my freedom to buy that very product) in a way that does not *override* my judgment about what is good for me, but rather *gives effect* to my judgment of what is good for me (viz., that it is good for me that CPSC decide what is safe and what is not). Even if I did not explicitly authorize the CPSC to act in this way, it is plausible to think I have implicitly delegated them to make safety determinations on my behalf. The most reliable judgment standard enables us to remove the mystery surrounding the idea of an implicit delegation of authority. To implicitly delegate authority to the CPSC to make safety determinations, all that is necessary is that a majority of those affected by CPSC determinations autonomously endorse their making such determinations (where there is no reason to expect them to withdraw their endorsement in the future).

The CPSC's actions may be only weakly paternalistic when directed toward those who regard them as an authority on product safety. What about those who do not regard the CPSC as an authority? Suppose Mike regards himself as an authority on the hazards of tires and does not regard the CPSC as an authority on those hazards. Mike would prefer not to have any tires taken off the market, so he can make his own judgment about their merits.

When the CPSC takes a tire off the market, it prevents Mike from being able to make his own decision about whether or not to buy it. Is it acting paternalistically toward Mike (whom I assume to be an autonomous adult)? If so, its action would be strong paternalism, and a right against strong paternalism would give Mike the right to block CPSC actions.

Mike should not have the right to block the CPSC actions. It is important to see that the CPSC's action in taking a tire off the market is not strongly paternalistic, even if it does prevent Mike from being able to buy a tire he would like to buy. The reason is that the justification for the law is not to overrule Mike's judgment about what is good for *him*. It is to give effect to the majority's opinion about what is good for *them*. The limitation on Mike's freedom is a byproduct rather than the aim of CPSC action.

One test for whether preventing others from doing things they believe would promote their own good is a byproduct rather than an aim of the relevant legislation is to consider whether there could have been alternative legislation that would have achieved the non-paternalistic goals without preventing others from doing things they believe would promote their own good. For example, it might be argued that CPSC action to take a product off the market is paternalistic, because the non-paternalistic goals of the action could be achieved without taking the product off the market. If the CPSC determined a tire to be hazardous, they could simply order that the tire carry a warning label of the hazards, with a note that the CPSC recommends that consumers not buy it. This alternative would enable me, someone who regards the CPSC as an authority, to avoid products the CPSC deems hazardous and for Mike, someone who does not regard the CPSC as an authority, to make his own determination of whether or not to buy a product the CPSC deems hazardous.

The problem with this suggestion is that if products deemed hazardous by the CPSC are put on the market and purchased, anyone, even someone who did not purchase the product, may be injured by it. If Mike decides to buy tires deemed hazardous by the CPSC and they fail, his vehicle may collide with mine. Also, it is impossible to make

sure the warnings are communicated to subsequent buyers. Mike may buy the tires deemed hazardous by the CPSC and put them on his vehicle. Later, Mike may sell his vehicle to a buyer who is not aware the tires were deemed hazardous by the CPSC. Suppose I am the buyer. If the tires are defective, they may fail and cause me to have a collision. If the tires had not been allowed on the market, that collision would not have occurred. It is easy to see how a majority of citizens might want to be protected from such hazards.

So there really is no other alternative that provides the same level of protection against hazardous products provided by giving the CPSC the power to keep hazardous products off the market. Where a majority wants to be protected from hazardous products in that way, the law is not paternalistic, even if, as a byproduct, it prevents Mike and others like him from being able to buy some products they would regard as having benefits that justify whatever risks there might be.

There are a large number of laws and government regulations that have the feature that in order to give effect to a legitimate aim of a majority, they have as a byproduct, not a goal, that they prevent some people from being able to do what they think would be best for them. This is the *majority spillover effect*. The majority spillover effect is most likely to be confused with paternalism when the majority preference is to delegate authority to others to make decisions for their own good. Examples include a great variety of food, drug, product, and occupation safety laws and regulations, as well as laws requiring prescriptions for some medications, laws requiring fluoridation of water, and laws requiring the enrichment of flour with certain nutrients. It would be surprising if a majority endorsed these laws with no reservations. However, it is almost certain that a majority of U.S. voters would endorse these laws, even with their imperfections, over the alternative of eliminating them. Indeed, each time there is a report of injuries due to an unsafe product or an unsafe drug or unsafe food—for example, in the recent discovery of "Mad cow" disease—there are widespread calls to strengthen the relevant safety laws. So it is quite plausible that a majority would endorse these laws. If so, those who do not endorse them cannot object to the laws as paternalistic. They are simply an example of the majority spillover effect.

Delegations Of Authority That Impair Future Autonomy

The most reliable judgment standard draws our attention to the special problems raised by one kind of delegation of authority. These are delegations of authority so broad that they impair the autonomy of one's future judgments. These situations have the structure of the example of Lee the soccer player, except that they bring about not the death of the target, but only the end of the target's autonomy. Such delegations of authority should only be permitted in the extreme case in which it is reasonable to believe that the endorsement is unequivocal—that is, the subject would be expected to continue to unequivocally, autonomously endorse the surrender of autonomy if the surrender were not accepted. In a subsequent chapter I provide a hypothetical example of this kind.

Such extreme examples are not realistic. Most people would not endorse a delegation of authority that impaired their autonomy. This provides a theoretical limit on justifiable policies by legal authorities.

Scanlon (1972) raises the question of what would be necessary for a delegation of authority not to compromise one's autonomy. Scanlon's answer is this: For a delegation of authority to be autonomous, the decision to delegate such authority must be an

effective exercise of the agents' judgment about what is good for them and that the delegation of authority cannot be so great as to prevent them from being able to continue to effectively exercise their judgment on the question of whether they should continue to delegate that authority. He points out that for citizens to delegate to the state the power to determine what is true and what is false on all matters would not be something they could do without sacrificing their autonomy. The previous discussion shows it would be possible to delegate to the state the power to determine *some* truths, for example, about which products are safe and which are not, so long as the delegation of authority did not compromise the autonomy of their judgment that the delegation of authority itself was a good idea.

Scanlon focuses on the importance of freedom of thought and expression to the effective exercise of one's judgment, but the other civil liberties (freedom of the press and freedom of association) are implied. I would add that it is also necessary that the decisions of the authority and the reasons for them be available to the public, so they can be reviewed and evaluated by knowledgeable commentators.

Thus, for example, for my delegation of authority to the CPSC on product dangers not to result in a surrender of my autonomy, there must be freedom of expression, of the press, and of association to make possible private and public discussions of the successes and failures of the CPSC. In addition, there must be publicly available information on the decisions of the CPSC and the rationales for them. Even if I am not an authority on product safety, open debate among commentators more knowledgeable than I am can give me the information I need to exercise my judgment on whether it is good for me to have the CPSC continue to play its role as an authority for me on product safety.

> Scanlon considers a particularly extreme form of this delegation of authority:
>
> It is quite conceivable that a person who recognized in himself a fatal weakness for certain kinds of bad argument might conclude that everyone would be better off if he were to rely entirely on the judgment of his friends in certain crucial matters. Acting on this conclusion, he might enter into an agreement, subject to periodic review by him, empowering them to shield him from any sources of information likely to divert him from their counsel on the matters in question. Such an agreement is not obviously irrational, nor if it is entered into voluntarily, for a limited time, and on the basis of the person's own knowledge of himself and those he proposes to trust, does it appear to be inconsistent with his autonomy. The same would be true if the proposed trustees were in fact the authorities of the state.(1972, 219)

Call this the *Scanlon Condition on Autonomous Delegations of Authority.* The idea of delegating authority to protect oneself from a weakness for bad arguments or bad reasoning is not merely a hypothetical one. I give some examples shortly. Because the delegation of authority is itself based on the person's autonomous judgment about what is good for her, it could qualify as weak paternalism.

However, if the delegation of authority is to qualify as weak paternalism, the delegation of authority must not undermine the autonomy of the delegator's future judgments about whether it would be good to continue or to revoke the delegation of authority. Implicit in the Scanlon Condition is the requirement that the delegator of authority be able to make an autonomous judgment in the future about whether or not it

would be good for him to revoke it. For example, a person who came to the conclusion that he had a weakness for bad arguments about what was good for him and decided to rely on the state to make such determinations for him would not be able to delegate that authority to the state without surrendering his future autonomy. Presumably, if the state shielded him from arguments about what was good for him, it could easily give him only arguments that supported his continuing the agreement and not revoking it. Then the fact that he later continued to believe the agreement was good for him and did not revoke it would not make his delegation of authority autonomous.

So the Scanlon Condition should be understood to require that the agreement to delegate authority not compromise the autonomy of the delegator's future judgments about whether or not to revoke the agreement. Understood in this way, the Scanlon Condition seems quite plausible. Nonetheless, it seems to me it needs one slight modification, which I motivate by way of an example.

Pyramid Schemes and Related Phenomena

Consider, for example, pyramid schemes. What makes these schemes almost irresistible is that they take advantage of what is probably the most trusted form of human reasoning and that they combine it with one of the most powerful sources of human motivation. A pyramid scheme can be transparent or non-transparent. I discuss the transparent scheme first. A *transparent* pyramid scheme is one that everyone recognizes to be a pyramid scheme. It operates by way of recruitment. In the simplest case, a new recruit pays a certain initiation fee (e.g., $10,000). This initiation fee is divided among the person who recruited her, and perhaps the person who recruited the person who recruited her, etc., all the way back to the person started the whole scheme. Each new recruit can fully recover her initiation fee and turn a profit, by attracting enough new recruits, who may themselves attract additional new recruits, etc.

Consider the recruits to the pyramid scheme by generation. There is no guarantee a pyramid scheme will last any given number of generations, but there is a powerful snowball effect for pyramid schemes that survive more than two or three generations. In each generation, the members of previous generations get richer. One of the most powerful of human inferences is to project the future on the basis of what has happened in the past. Each time a pyramid scheme survives a single generation, all those who invested in the past become richer. As a result, it is almost irresistible to infer that those who invest now will become richer in the future. Call this sort of reasoning the *reasoning based on past performance*.

Of course, there is also a line of reasoning that competes with this conclusion. This is the line of reasoning that guarantees that in some generation, without previous warning, the scheme will collapse for lack of new investors. For example, if in each generation, each new recruit brings in only two new recruits, the number of investors at the end of the nth generation will be 2^n. By the 20^{th} generation, the total number of recruits would be over 1,000,000. Not very many generations later it would exceed the number of people on earth who would be able to pay the initiation fee. By the 33^{rd} generation it would exceed the entire population of the earth. Call this reasoning the *reasoning to inevitable collapse*.

Note that the reasoning to inevitable collapse does not completely undermine the reasoning based on past performance. It only implies there is *one* generation in which past performance will not be a good guide to future performance. Ironically, because that

one generation is the last, it can *increase* the motivation to invest early, so as to get rich before the inevitable collapse.

There is a second powerful source of motivation to invest that is not completely undermined by the reasoning to inevitable collapse. This second source of motivation to invest comes from the fact that most people evaluate their level of well-being in part on the basis of comparisons with others. Call this motivation the *motivation based on comparative well-being*. If those who have not invested in the scheme compare themselves with others who have, the fact that the others are getting richer each generation will make those who have not invested regard themselves as worse off, even though they have not actually lost anything. When combined with the reasoning to inevitable collapse, this motivation also favors investing as early as possible, so as to keep up with those who have already been enriched by the scheme.

The difference between a transparent and a non-transparent pyramid scheme is that in a non-transparent pyramid scheme the investors are led to believe that they are investing in something other than a pyramid scheme. In a non-transparent pyramid scheme, the investors have no basis for recognizing the inevitable collapse, so the motivation to invest is typically even more irresistible.[36] Non-transparent pyramid schemes are typically prohibited by laws against fraud.

What about transparent pyramid schemes? Transparent pyramid schemes involve no force or fraud, so prohibiting them would not be acceptable to a libertarian. A Scanlon-type rationale can justify a prohibition.

In transparent pyramid schemes the reasoning to inevitable collapse is available, but there are still strong motivations to invest, the sooner the better. When these pyramid schemes collapse, there are always more losers than winners, because the average gain of the winners (some of whom may acquire very large sums) is greater than the average loss of the losers. Knowing all this, it might be quite reasonable for a majority to favor banning pyramid schemes altogether. So long as they were able to effectively exercise their judgment on the question of whether the ban should be removed—that is, so long as the issue could be freely discussed in the press and among citizens and so long as there information on the effects of the ban was readily available, such a ban would not be strong legal paternalism.

It would be somewhat optimistic to think most people in the U.S. understand pyramid schemes well enough to understand why most people have a "fatal weakness" for investing in them. So laws against transparent pyramid schemes cannot be justified on the grounds that most people understand the "fatal weakness" in human reasoning that make pyramid schemes so appealing and authorize the government to protect them from this weakness.

To understand the justification for laws against transparent pyramid schemes, it is necessary to make a slight modification the Scanlon Condition discussed above:

(Extended Scanlon Condition) It is quite conceivable that a person might recognize that he was susceptible to weak reasoning, without knowing exactly what kinds of weak reasoning he was susceptible to. Concerned about the possibility that his unknown susceptibilities might cause him significant losses, he might enter into an agreement delegating to someone he regards as an authority on such susceptibilities the power to shield him from influences that might trigger his susceptibilities to weak reasoning in cases in which there would be a potential for significant losses. Such an

agreement is not obviously irrational if it is entered into voluntarily, for a limited time and on the basis of reasonable beliefs about the reliability of the judgment of those he proposes to trust and if it is not so extensive as to prevent him from making autonomous judgments in the future about whether to continue the agreement. The same would be true if the proposed trustees were government authorities.

I do not know if Scanlon would accept this modification of his principle, but it is one that seems reasonable to me. A person who delegated to the government the authority to shield her from her own unknown susceptibilities to weak reasoning would be very close to surrendering her autonomy. To avoid surrendering her autonomy, the delegation of authority agreement would have to protect the autonomy of her future decisions on whether or not to continue to entrust the authority with the power to protect her against unknown weaknesses in reasoning. This would not be impossible. For example, the authority could be required to make public its own reasons for any limitation it imposes and there should be freedom of expression and of the press for discussion of the reasons for and against continuing such limitations.

It might seem that this requires a person to actually come to understand the weaknesses in her reasoning the authority is protecting her against. That is not correct. An authority might justify a ban on transparent pyramid schemes, not by explaining to me the susceptibility to weak reasoning it is protecting me against, but simply by pointing to examples in which people like me suffered significant losses due to the weakness of reasoning, whatever it might be. If people are free to dispute the reasons given by the authority and no one does so, I may reasonably conclude that the ban on pyramid schemes is protecting me from losses I want to avoid, without understanding the weakness in my reasoning that makes me susceptible to those losses.

If this is correct, then in a democracy with rights to civil liberties and rights to relevant information about the rationale for government legislation, a majority could delegate to the government the power to protect them from unknown as well as known susceptibilities to weak reasoning. Of course, if our reasoning were so weak that we had to be protected from almost all of it, then there would not be very much for a right against legal paternalism to protect. In the actual world, cases where the government is justified in protecting us from weak reasoning are relatively rare. So long as we can learn from experience how to make allowances for our own limitations and weaknesses in reasoning, there is no need or justification for the government to intervene. Government intervention is always a blunt instrument, only to be resorted to when individuals cannot reliably correct for their own weaknesses. What is surprising is that there are some cases in which a good case can be made for government intervention. Transparent pyramid schemes seem to me to be such a case.

Understanding transparent pyramid schemes helps to understand a related case in which government intervention can be justified. These are cases of *speculative bubbles*, cases that share the motivational structure and logic of pyramid schemes, even though they may not be *schemes*, because they are not planned by anyone. The most famous speculative bubble was the tulip craze in Holland in the 17th century.[37] Stock markets are particularly susceptible to them. Any investment evaluated by what people are willing to pay for it can generate a speculative bubble, where the prices people are willing to pay for an investment climb simply on the basis of expectations based on the fact that they have been climbing recently. This is another case where reasoning based on past

performance can generate the very behavior it predicts, until an inevitable collapse. In addition, motivation based on comparative well-being can make it almost irresistible to invest, if the people one compares oneself to have invested and become richer. As a result, the motivation to invest in speculative bubbles is irresistible for most people. Call these factors that motivate investment in speculative bubbles, *pyramid phenomenon motivation*.

Speculative bubbles are more like non-transparent than transparent pyramid schemes, because it is never possible to discern exactly how much of the price of an investment is due to pyramid phenomenon motivation and how much is due to other factors. Like pyramid schemes, speculative bubbles can have disastrous consequences. Not only can they result in great monetary losses for most investors, but they can also undermine investor confidence and thus discourage productive investment. After the great stock market crash of 1929-30, it took decades to restore investor confidence in the U.S. stock market. Much the same justification that can be given for banning transparent pyramid schemes can be given for regulating stock markets and other investments to protect against speculative bubbles. Examples of such regulations are circuit breaker requirements that suspend trading in a stock or an entire market if there is too much gain or loss in a single session. Using the Extended Scanlon Condition we can explain why these regulations need not be paternalistic, even if most investors do not understand the factors in their reasoning that generate speculative bubbles. An awareness of the disastrous effects of speculative bubbles can justify a majority in authorizing its government to protect them from their own susceptibility to them.

Of course, there will always be some people who see transparent pyramid schemes or speculative bubbles as a way to get rich. They would not want to be protected from them, because they would like to try to profit from them. Nonetheless, laws banning pyramid schemes or regulating investments to try to prevent bubbles are not paternalistic, because the fact that they limit people who would like to try to profit from them is simply a majority preference spillover effect, a byproduct of legislation aimed at satisfying a legitimate majority preference.

Conclusion

From a Kantian or other nonconsequentialist perspective, the distinction between weak and strong paternalism would almost have to be based on the autonomy or lack thereof of the target's choice at the time of intervention. The hypothetical autonomous endorsement standard is one promising way for the nonconsequentialist to make the distinction between weak and strong paternalism.

However, from a consequentialist perspective, the focus shifts from the autonomy of the target's choice to the reliability of the target's judgment. From the consequentialist perspective, the hypothetical autonomous judgment standard is approximately correct because the judgments of autonomous adults about what is good for them are generally reliable, and generally more reliable than other people's judgments about what is good for them (the claim of first person authority). From the consequentialist perspective, it is in theory possible to justify overruling a person's current autonomous judgment of what is good for her to give effect to her own (typically later) more reliable judgment about what is good for her. Thus, the most reliable judgment standard explains how there can be exceptions to the hypothetical autonomous judgment standard. This standard provides a new way of understanding how, in theory, some prohibitions on addictive drugs or on

suicide could be justified. It also provides a new way of reconstructing Mill's argument against the enforcement of slavery contracts, to show why the argument extends to religious vows, but does not extend to most other contracts. Finally, the most reliable judgment standard provides the basis for a right against strong legal paternalism that should be a constitutional right of all autonomous adults.

REFERENCES

Berlin, Isaiah. 1969. Two Concepts of Liberty. In *Four Essays on Liberty* (Oxford: Oxford Univ. Press): 118-172.

CNN. 2001. Gallup Poll: 82 Percent of U.S. Smokers Say They Would Like to Quit. (1/5/2001). URL: http://www.tobaccofreedom.org/issues/cessation/survey_2001.html.

Cummings, Peter, Barbara McKnight, Frederick P. Rivara, and David C. Grossman. 2002. Association of driver air bags with driver fatality: a matched cohort study. *British Medical Journal* 324 (11 May): 1119-1122.

Dworkin, Gerald. 1972. Paternalism. *The Monist* 56: 64-84. reprinted in Sartorius (1983): 19-34.

Gerald Dworkin. 1983. Paternalism: Some Second Thoughts. in R. Sartorius (1983): 105-111.

Dworkin, Ronald, Thomas Nagel, Robert Nozick, John Rawls, Thomas Scanlon, and Judith Jarvis Thompson. 1997. Assisted Suicide: The Philosopher's Brief. *New York Review of Books* (March 27): 41-47.

Feinberg, Joel. Legal Paternalism. *Canadian Journal of Philosophy* 1: 106-124. reprinted in Sartorius (1983): 3-18.

Gibbard, Allan and William L. Harper. 1978. Counterfactuals and Two Kinds of Expected Utility. In *Foundations and Applications of Decision Theory*, vol. I (Dordrecht: D. Reidel): 125-162.

Good, I.J. 1967. On the Principle of Total Evidence. *British Journal for the Philosophy of Science* 17: pp. 319-321.

Mackie, Gerry. 1996. Ending Footbinding and Infibulation: A Convention Account. *American Sociological Review* 61: 999–1017.

MacKay, Charles. [1841]. *Extraordinary Popular Delusions And The Madness Of Crowds* (New York: Noonday Press; 1969).

Mill, John Stuart. [1859]. *On Liberty* (New York: Prometheus; 1986).

Millgram, Elijah. 1997. *Practical Induction* (Cambridge: Harvard Univ. Press).

Nagel, Thomas. 1970. *The Possibility of Altruism* (Oxford: Oxford Univ. Press).

O'Rourke, P.J. 1998. *Eat the Rich* (New York: Atlantic Monthly Press).

Parfit, Derek. 1984. *Reasons and Persons* (Oxford: Clarendon Press).

Scanlon, T.M. 1972. A Theory of Freedom of Expression. Philosophy and Public Affairs 1: 204-226.

Sartorius, Rolf, ed. 1983. *Paternalism* (Minneapolis: University of Minnesota Press).

Schelling, Thomas C. 1978. *Micromotives and Macrobehavior* (New York: W.W. Norton).

Waldron, Jeremy. 1993. *Liberal Rights* (Cambridge: Cambridge Univ. Press).

NOTES

[1] This paper has benefited from comments on an earlier draft from Jacques Corriveau, Karen Houle, Katherine Kim, Jamie Mayerfeld, Liam Murphy, Tom Nagel, John Talbott, and Martin Tweedale.
[2] I set aside here the discussion of other non-paternalistic justifications that might be given for the legal establishment of religion.
[3] *Griswold v. Connecticut*, 381 U.S. 479 (1965).
[4] The Court extended this right to unmarried people in *Eisenstadt v. Baird*, 405 U. S. 438 (1972).
[5] *Loving v. Virginia*, 388 U. S. 1, 87 S.Ct. 1817 (1967).
[6] 87 S.Ct. at 1824.
[7] *Roe v. Wade*, 410 US 113 (1973).
[8] *Planned Parenthood v. Casey*, 505 U.S. 833, 851 (1992).
[9] *Bowers v. Hardwick*, 478 US 186 (1986).
[10] *Powell v. State,* 510 SE 2d 18 (1998).
[11] *Lawrence v. Texas,* 123.S. Ct. 2472 (2003).
[12] 123 S. Ct. at 2481. By defining the right as a liberty right, the Court moved it from the penumbra and gave it a home in the Due Process Clause.
[13] The United States is no longer in the forefront of the development of a right against legal paternalism. While the U.S. Congress debates a constitutional amendment to legally enshrine heterosexual marriage, a Canadian court has invalidated legal prohibitions on homosexual marriage and the Legislature has acquiesced in the decision. *Halpern* v. *Toronto*, 2003 WL 34950 (Ontario Ct. App.). In Europe, most countries recognize same sex unions, and Belgium and the Netherlands have legalized same sex marriage.
[14] *In re Quinlan***,** 70 N.J. 10, 355 A. 2d 647, cert. denied, (1976).
[15] *Cruzan v. Director, Missouri Dep't of Health*, 497 U.S. 261 (1990).
[16] The Supreme Court's ruling came in the companion cases *Washington v. Glucksberg*, 117 S.Ct. 2258 (1997) and *Vacco v. Quill*, 117 S.Ct. 2293 (1997). For the Philosophers' Brief, see Dworkin, Nagel, Nozick, Rawls, Scanlon, and Thomson (1997).
[17] Of course, they could regard themselves as better off because other people are not engaging in the prohibited activity, if they would no longer worry that those other people were harming themselves. Here I ignore any increase in well-being simply due to relief that others are not harming themselves. I follow Mill [1859] in believing that this sort of loss is more than outweighed by the gains from allowing people freedom to act on their autonomous judgments of what is good for them.
[18] Before the National Hockey League adopted a rule requiring non-goal-keepers to wear helmets, Schelling quoted one player who explained why the players themselves did not voluntarily choose to wear them: "It's foolish not to wear a helmet. But I don't because the other guys don't. I know that's silly, but most of the players feel the same way. If the league made us do it, though, we'd all wear them and nobody would mind." (1978, 213)
[19] In my citations of Feinberg (1971), page references are to the reprinted edition in Sartorius (1983).
[20] Mill ([1859], 109), quoted with approval by Feinberg (1971, 8). The names in brackets are my additions to Mill's example.

[21] CNN (2001). Because this survey also shows that an overwhelming majority of smokers began smoking before age 18 (indeed, a majority began smoking before age 16), there is no basis for thinking their decision to begin smoking was an autonomous one. I set aside that issue here, because I am interested in the theoretical possibility that intervention could be justified even in cases in which a person makes an autonomous decision to begin smoking.

[22] Although Nagel does not make explicit the assumption that both the earlier and later selves regard the other's judgment as less reliable than their own, I think this assumption is implicit in his discussion (1970, 74).

[23] This is an informal version of a theorem in Bayesian decision theory that more information is always better than less. See Good (1967).

[24] See Millgram (1997) for a much more detailed account of the extent to which practical reasoning involves learning, or in Millgram's terms, *practical induction*.

[25] Compare G. Dworkin's suggestion that paternalism can be justified by "future-oriented consent"(1972, 28). My page citations to Dworkin (1972) are to the reprinted version in Sartorius (1983).

[26] It should be clear that my claim that people's later autonomous judgments about what is good for them are more reliable than their earlier ones is not an implicit argument for raising the voting age. Even if my later autonomous judgments about what is good for me will be more reliable than my current ones, the claim of first person authority implies that my current autonomous judgments are still generally more reliable than *other people's* judgments about what is good for me, even the judgments of other people who are older and more experienced than I am.

[27] Because this is not a situation in which it is important to distinguish conditional probability from the probability of the corresponding subjunctive, I use the familiar symbol for conditional probability '/' in the text. For more on the distinction, see Gibbard and Harper (1978).

[28] The argument in the text depends on the simplifying assumption that the relevant probabilities, Prob(UE/PI) and Prob (UE/-PI), are the same for everyone. In the more general case when they are not all equal, the requirement becomes that their average be greater than .5. As I mentioned in the discussion of democratic decision rules, there is no requirement that all legislation be governed by the same rule. The same can be said for policies of paternalistic interference. Where a paternalistic policy has the potential to produce very bad results, it might be reasonable to require that Prob(UE/PI) and Prob (UE/-PI) both be significantly greater than .5. I am not sure that this would be necessary, because presumably the potential for very bad results would be considered by each individual in deciding whether or not to endorse the intervention. In any case, I believe that .5 should be the default threshold value for weak paternalism, in the absence of special factors that would justify raising it.

[29] Note that a similar analysis applies to contacts to commit suicide made by someone who antecedently had not intention to commit suicide.

[30] G. Dworkin advocates a formula of this kind: "Paternalism is justified only to preserve a wider range of freedom for the individual in question."(1972, 28) I am not sure how Dworkin would apply this formula to addictions, because he does not directly discuss the topic.

[31] Thanks to Katherine Kim for reminding me that the regrets of those who use prostitutes' services should also be considered.
[32] On this distinction, see Waldron (1993, Chap. 3).
[33] The rule accepted in almost all jurisdictions is that a competent adult has the "basic right . . . to refuse treatment even when the treatment may be necessary to preserve the person's life." *In re Fosmire v. Nicoleau*, 551 N.E.2d 77, 81 (N.Y. 1990).
[34] There is one further relevant consideration. Motorcycle helmet laws might be justified on the grounds that human beings don't reason well with small probabilities. Perhaps motorcycle laws are protections against the illusion that the probability of death or serious injury is so small as to be negligible. For such a justification to be weakly paternalistic, there would have to be evidence that when people were made aware of this tendency to misevaluate events with small probabilities, they would come to endorse intervention to protect them from misevaluation in cases involving a small probability of death or serious disability. In the absence of this sort of evidence, this sort of rationale cannot be used to justify a helmet law as weak legal paternalism.
[35] According to one study, seatbelts alone increase chances of surviving a collision by 65%, much higher than the 8% increase from airbags alone (Cummings, McKnight, Rivara, and Grossman 2002).
[36] In Albania in the 1990's, a large fraction of the nation's savings was lost to non-transparent pyramid schemes. See O'Rourke (1998, Chap. 3).
[37] For more on the history of this and other speculative bubbles, see MacKay [1841], who reports that in Holland at the height of the tulip craze, a single tulip bulb was sold for two lasts of wheat, four lasts of rye, four oxen, eight swine, twelve sheep, two hogsheads of wine, four tuns (approximately 1000 gallons) of beer, two tons of butter, one thousand pounds of cheese, a bed, a suit of clothes, and a silver drinking cup.