

Introduction: Paternalism – Issues and trends

Christian Coons and Michael Weber

John Stuart Mill famously decried paternalism of any kind, whether it is carried out by private individuals or the state:

[T]he sole end for which mankind are warranted, individually or collectively, in interfering with the liberty of action of any one of their number, is self-protection . . . the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant.¹

Equally famously, critics have charged that Mill's utilitarianism is inadequate to ground such an absolute prohibition. If there is a case for an absolute ban on paternalism, many have thus thought, it must be grounded instead in the fundamental importance of one of a family of considerations that includes liberty, freedom, and autonomy. However, it is unclear whether this move really helps, or so we will suggest.

This mere snippet from the historical debate is enough to show that paternalism is a topic that engages deep philosophical issues in normative ethics and political philosophy, including the significance and nature of freedom and autonomy, and the relation between individuals and the state. But interest in the topic is of course not primarily due to its theoretical depth. Instead, it is due to practical considerations. A wide variety of policies and laws in the United States and elsewhere are considered paternalistic, and are controversial for just that reason. For instance, there are laws that require motorcyclists to wear helmets and passengers in cars to wear seatbelts. Government agencies regulate both prescription and recreational drugs. Taxes are levied on cigarettes, and bans on trans-fats have been enacted. Participation in pension programs, such as Social Security in the United States, is mandatory. Mill's negative view of paternalism is reflected in some identifying policies along these lines as part of an ever-growing "nanny-state."

¹ Mill, *On Liberty*, 9.

Because paternalism in this way raises significant theoretical and practical concerns, it has been a topic of long-standing interest to moral and political philosophers as well as political actors and the lay public.

Interest has been heightened recently due to an alleged breakthrough, a way of squaring paternalism with libertarian concerns for that family of values that includes liberty, freedom, and autonomy. This “libertarian paternalism,” championed by Richard Thaler and Cass Sunstein in their recent bestseller, *Nudge: Improving Decisions About Health, Wealth, and Happiness*, raises new and interesting questions, especially in virtue of its appeal to recent research in psychology and behavioral economics.² The essays in this volume address the fundamental and long-standing issues raised by paternalism as well as the debate sparked by Thaler and Sunstein. In this introduction we will not summarize each of the contributions to the volume. Instead, more helpfully we hope, we will touch on some of the key aspects of the present-day discussion of paternalism, so as to provide a broad context for thinking about the essays herein. Our discussion of these new developments will be framed by a central question in both classic and contemporary debates: What, if anything, makes paternalism morally problematic? Where appropriate, of course, we will indicate how specific contributions to this volume figure in the larger narrative.

Normative debates about paternalism – or at least “hard” paternalism – don’t usually concern *whether* it is problematic but rather *how* problematic it is.³ Specifically, most assume that we have a *pro-tanto* reason to avoid paternalism towards competent adults. The real debate concerns if it can nevertheless ever be justified, and if so, when. Despite this, in recent years it has become surprisingly difficult to explain exactly why paternalism is even *pro-tanto* objectionable, and more difficult still to defend its absolute prohibition. It has become more difficult not because the world has somehow changed but rather because philosophical reflection has made it clearer what paternalism does and does not involve. Crucially, there’s a growing consensus that it needn’t be regarded as coercion, removal of

² Thaler and Sunstein, *Nudge*.

³ The soft paternalist maintains that interference is justified if and only if the “target” person is not acting sufficiently knowledgeably and voluntarily. Although soft paternalism is sometimes understood as paternalism that does not impose values or ends on its targets, that is not how we’re using it here. Hard paternalists allow interference to benefit agents acting knowledgeably and voluntarily. Our discussion is concerned with hard paternalism. As such, from here forward “paternalism” and its derivatives should be taken to refer to hard paternalism unless specifically noted; indeed Feinberg, *Harm to Self*, 16, famously questions whether soft paternalism qualifies as paternalism at all.

choice, or even disregard for the target's evaluative perspective. Because of this, some traditional views about why paternalism is *pro-tanto* wrong are no longer available. We make this point not to advocate paternalism, but to help illustrate why despite all the attention previously paid to it, it remains a rich and evolving topic for discussion.

Of course, one might skeptically insist that it is paternalism's very richness that makes it problematic as a focus of direct study. After all, in paternalism, perhaps more than any other topic in moral and political philosophy, deep conflicts between competing traditions emerge, not only about the limits of state authority, but in more general disputes about the relative priority of well-being, freedom, choice, and autonomy, and whether these are ideals to be promoted or respected. The skeptic about the direct study of paternalism might then insist that no progress can be made without first resolving these deeper disputes. But this assumes what we might call a "top-down" approach: General philosophical issues must be solved first, the results of which are then simply applied to ground-level disputes in political philosophy or applied ethics. The top-down approach, however, is both methodologically questionable and difficult to consistently maintain. For example, it would presumably be a *reductio* of any broad theoretical view if it entailed that we may not stop a friend from drinking the gasoline she thinks is gin. So our verdicts about paternalism needn't be a mere application of our more general commitments in moral and political philosophy. Rather, the opposite may be true: Careful reflection on paternalism's moral status may help illuminate or adjudicate debates about those deeper issues. Indeed, if we see the coercive power of the state as justified in part by its claim to benefit those subject to its power, it appears state authority itself rests on a kind of paternalistic rationale. Paternalism's relevance to wider debates is also apparent in Christopher Wellman's complaint that Rawls' Principle of Fairness is objectionably paternalistic; Richard Arneson's contribution in this volume defends Rawls on this score.⁴ In this debate, paternalism's status is treated as determining, rather than being determined by, our more general normative commitments.

So just as consensus at the theoretical level may be useful in drawing conclusions about how we should regard paternalism, consensus about paternalism may inform our more general theoretical positions.⁵ With that

⁴ Wellman, "Samaritanism and the Duty to Obey the Law" in Wellman and Simmons, eds., *Is There a Duty to Obey the Law?*

⁵ A powerful instance of this sort of strategy can be found in Wall, "Self-Ownership and Paternalism."

in mind, it is worth returning to the widely held view that paternalism is inherently problematic – that there is always at least some *pro-tanto* reason to avoid it. But what, precisely, makes paternalism problematic? And what does this reveal about deeper moral views? At least this much initially seems obvious: The problem does not lie in its end alone – in the fact that paternalism is undertaken for *the good of* or *the sake of* its targets.⁶ Paternalism appears to be very broadly speaking benevolent. But there's nothing wrong with benevolence *per se*. What makes paternalism at least *pro-tanto* wrong, then, presumably has to do with the means – with how it goes about benefiting its target. For example, forcing you to have a second helping of broccoli casserole – a tasty and healthy choice – is objectionable to the extent that it is because I am forcing you to do it. It would likewise be objectionable, and apparently for the same reason, if I forced you to eat it for a scientific experiment I am conducting. What is *pro-tanto* wrong with paternalism in this case, apparently then, has nothing to do with the fact that I am trying to improve your health or please your palate. That it is good for you may speak in its favor. That it promotes my scientific experiment seems to count less, most likely because I am the beneficiary rather than you who are forced to eat the casserole. In both cases, however, it seems most plausible that the reasons to force you to eat the casserole are too weak to outweigh the reason not to coerce you. Generalizing from this example, we might suspect that what's wrong with paternalism is rather straightforward: Paternalism involves some form of coercion or interference, which requires special justification. Advocates of a general prohibition on paternalism are, then, simply those who think this justification cannot be met – the well-being of the target of paternalism never outweighs the morally problematic use of coercion.

But such an account will not do. As is now more widely recognized, omissions can be paternalistic, and thereby problematic, even though the omission would not otherwise require any justification. Here, fascinatingly, paternalism's benevolence appears to contribute to what makes it problematic. To illustrate, we may omit telling you about tonight's concert because we believe that we ought not tempt you. This omission may require no justification: if we had simply not felt like telling you, we would do nothing objectionable. But when we omit for your sake, it seems

⁶ For simplicity, we will generally speak of paternalism and paternalistic acts, though it is probably more accurate to describe act–reason pairs as paternalistic, as Grill, “The Normative Core of Paternalism,” emphasizes. As will become clearer below, we believe there's an important distinction between acts performed for a target's well-being and those performed for her sake.

you have greater grounds to object. Specifically, our omission seems to involve reasoning *for you* – we’ve weighed the options for you, and decided not to tell you partly because we believe you might reach a different and “imprudent” conclusion. This feature of paternalism – a sort of reasoning for another – has not always been explicitly highlighted, but takes center stage in Seana Shiffrin’s recent but already highly influential characterization of paternalism. On Shiffrin’s view, paternalism by A towards B is behavior (including omissions) meeting the following conditions:

- (a) aimed to have (or to avoid) an effect on B or her sphere of legitimate agency
- (b) involving the substitution of A’s judgment or agency for B’s
- (c) directed at B’s own interests or matters that legitimately lie within B’s control
- (d) undertaken such that, compared to B’s judgment or agency with respect to those interests or other matters, A regards her judgment or agency to be (or as likely to be), in some respect, superior to B’s.⁷

This view in fact highlights *two* features of paternalism that have historically been overlooked or under-emphasized: i) paternalism may not involve coercion or active interference; ii) it needn’t involve a specific concern for the target’s *well-being*. The definition is not merely a potential theoretical advancement, it also sheds light on types of paternalistic practices and policies that, historically, have gone unappreciated. Daniel Haybron and Anna Alexandrova’s contribution to this volume is an excellent illustration of the point, as it uses some of Shiffrin’s insights to turn the tables on those who argue happiness-driven economics is objectionably paternalistic in a way that more traditional “minimalist” methodology in economics is not. And Sigal Ben-Porath’s contribution uses some of these same insights to shed light on the paternalism that is at the heart of contemporary debates about school choice.

Shiffrin’s account also offers us a direct answer regarding paternalism’s morally problematic nature. There’s no need to *explain* why paternalism is presumptively problematic, as all one needs to do is possess the concept to see that it is. For on her view “paternalism” is a kind of *thick* negative ethical term: Paternalistic acts by definition intrude on areas the target legitimately controls. Put in laymen’s terms, the paternalist intrudes in areas that are “none of her business” – it is an invasion of *personal* sovereignty. Return to our omitting to tell you about the concert. When

⁷ Shiffrin, “Paternalism, Unconscionability Doctrine, and Accommodation.”

we omit because we don't feel like telling you, that's "our business," but given that we omit out of exclusive concern for you, we've effectively made your business – questions about what is best for you – our business.

Though there's much to learn from Shiffrin's account, it also seems dubious that paternalism's problematic nature is part of its conceptual content. In the first place, some, like Thaler and Sunstein, think that paternalism is sometimes not problematic at all – that the salient negative features are not just outweighed, but eliminated. And as Peter de Marneffe emphasizes in his contribution to this volume, paternalism need not violate an individual's "sovereignty" – matters over which one should have legitimate control. Of course, we might contest these conclusions. But Shiffrin's view entails that they are conceptually incoherent, which doesn't seem right. More strikingly, notice that if Shiffrin were right, error theorists and moral skeptics should deny paternalism even exists. And although there have been some quibbles about which act-types are genuinely paternalistic, we tend to agree about which particular acts and policies qualify. This agreement would seem miraculous on Shiffrin's view, as there's great variation in our views about which matters are *legitimately ours to control*, and so there should be equally great variation in our views about which acts are paternalistic. Indeed, a hardline, old-fashioned act-consequentialist can regard some measures as paternalistic, and have reservations about them as such. But such theorists may also deny that there are *any* matters that *legitimately* lie within an agent's control because legitimate control implicitly appeals to something akin to an entitlement or right – normative categories that such a consequentialist may not countenance. Presumably, we should not characterize paternalism in a way that entails these consequentialists must deny its existence.

We don't intend to provide our own definition of paternalism here. Rather we wish only to raise a concern about normative or morally loaded characterizations.⁸ If we are to *explain* what is *pro-tanto* wrong with paternalism we may be better off identifying a non-normative characteristic of such acts that is morally problematic. Shiffrin's "substitution of judgment" in some non-normatively specified domain(s) – perhaps the very ones Shiffrin believes are "within our legitimate control" – is an intriguing and promising possibility, not merely to better understand paternalism but to cast light on what is now emerging as a central but neglected morally relevant feature – a feature whose moral status will be

⁸ For a very nice survey of recent definitions of paternalism including his own normative characterization, see Grill, "Anti-Paternalism and Public Health Policy."

touched on again below. Nevertheless, we should be further wary of canonizing Shiffrin's influential characterization. For Daniel Haybron and Anna Alexandrova's contribution offers some compelling reasons why it needs to be at least modified, and deeper, more developed concerns are a focus of Gerald Dworkin's opening section.

If we set aside morally "loaded" characterizations of paternalism, we again face the difficulty of explaining why paternalism is at least *pro-tanto* objectionable, and why some have thought it never permissible. Ironically, paternalism's most famous critic, Mill, appears to be of little help. Mill is traditionally seen as an exponent of a tradition – act-utilitarianism – which has a particularly difficult time explaining why paternalistic acts are presumptively problematic.⁹ The trouble, of course, is that whatever the correct characterization of paternalism, there can't be a guarantee that such acts never maximize utility. Paternalistic acts also need not involve any (perhaps outweighed) loss of utility. Accordingly, this tradition cannot explain either why paternalism is always wrong or even why there's always some reason to avoid it.

Utilitarian anti-paternalists might insist that this criticism involves a too-narrow view of what it is to make a presumptive case against paternalism. Rather than trying to locate a problem with what paternalists *do*, they might say, we should look at the relation between the paternalist and what she does. More specifically, one might argue that given our epistemic position, paternalism can always be *expected* to yield suboptimal outcomes. Following Mill, perhaps the proposed target of paternalism is always better placed to know where her happiness lies and how to best achieve it. And although we might quibble with this presumption – e.g., if I'm depressed, a wise friend may be a better judge than I about where my self-interest lies – Mill's claim is quite plausible, and it is especially powerful against *state* paternalism. State executives, assemblies, and state bureaucrats do not know us well – are not wise friends – and so are especially ill-placed to know (better than you) what's good for you. Furthermore, because state paternalism manifests in policies and statutes that apply to the citizenry at large, they are too blunt to be sensitive to the differing interests of the citizens bound by the policy or statute. Paternalist *penal* statutes suffer from still further justificatory problems, as Douglas

⁹ Mill's actual view is almost certainly more rich and nuanced than the traditional hedonistic act-consequentialism he's often portrayed as holding. See, among many others, Dworkin, *Mill's On Liberty*; Arneson, "Mill Versus Paternalism"; and more recently Jacobson, "Utilitarianism Without Consequentialism."

Husak explains in his contribution to this volume. The utilitarian anti-paternalist, then, could argue that the problem is that paternalism rarely promotes the good, and the paternalist is never in a position to know when it will. Indeed, these facts would be sufficient to make paternalism always wrong on a rule-utilitarian view and versions that appeal to *expected* rather than actual consequences.

The worry about such a reply is that even if the host of sociological, psychological, and economic data required to make such a case were available, it would seem useful only as further rationalization and not a reflection of what's seemingly wrong with paternalism. After all, we tend to be anti-paternalists independent of any grasp of the empirical data required to make this utilitarian case for it. Notice that the case focuses on an appeal to paternalism's inefficacy or "backfiring." But this focus seems misplaced, because anti-paternalists do not generally lament the ineffectiveness of these policies; they wouldn't wish or seek a "breakthrough" in socio-psychological engineering that would allow us to craft successful paternalistic policies. In fact, as Jeremy Blumenthal's essay in this volume points out, new empirical work is revealing that paternalism is effective in some contexts. But even so, its effectiveness is presumably rightly regarded as merely a necessary and not sufficient condition for its justification. Apparently then, conventional anti-paternalism, and the nearly uncontroversial sentiment that paternalism is in need of justification, reflects some non-utilitarian elements of our moral thought.

Utilitarianism's inability to capture what is seemingly wrong with paternalism might not be a feature of consequentialism generally. Perhaps a consequentialist will fare better by looking beyond happiness and towards other values such as freedom and autonomy. However, it is unclear this helps. For even if the target of a paternalistic act necessarily thereby loses some freedom or autonomy, such limitations can ultimately produce a net *increase* in freedom or autonomy. Mill's lone exception to his anti-paternalism illustrates the point well: Preventing one from contracting into slavery increases one's freedom overall.¹⁰ So merely appealing to freedom and autonomy as values to be promoted cannot capture what is wrong with paternalism, but instead will yield exceptions. In this volume, Larry Alexander discusses voluntary slavery contracts in great detail, ultimately offering, contra Mill, a provocative defense of such contracts.

¹⁰ Arneson, "Mill Versus Paternalism," has a particularly interesting discussion of Mill's case of contracting into slavery.

In general, it seems that no matter what value the consequentialist takes on board, the problem will be the same. Consider, for instance, certain values that seem perfectly suited for anti-paternalism, values such as non-interference and non-paternalism itself. Each of these could be promoted by paternalism. For instance, the State Department might place travel restrictions to various countries on citizens by appeal to the value of non-interference: Interfering with your travel will save you from even greater interference abroad. Similarly with non-paternalism: Acting paternalistically towards you now could save you from much greater paternalism later. For instance, if you live in the world's least paternalistic state, a prohibition on emigration may maximally promote non-paternalism.

It seems, then, that the tension between consequentialism and deep anti-paternalism is rooted in consequentialism's view of all values, including liberty, freedom, and autonomy, as things to be promoted, where promotion involves maximizing. Deontologists, in contrast, sometimes suggest that such values are to be respected, or that people are to be respected as free and autonomous. But what does it mean to respect someone as free and autonomous? What does such respect require? And most importantly, how might paternalism fail to fully respect persons as free and autonomous? Perhaps if we want to understand ordinary anti-paternalist sentiment, it make sense to turn to ordinary expressions of it. Typically, we say things like "you/the government don't know what's good for me," "you/the government can't tell me what's good for me," or "it's none of your/the government's business." Some of these expressions won't help, as they are simply ways of expressing anti-paternalism or reasons why it may be ineffective. However, some may be fruitful. So, for instance, the expression "you/the government can't tell me what's good for me," suggests that paternalism fails to respect persons as free and autonomous by imposing values on them. For instance, consider laws requiring motorcyclists to wear helmets. Such laws apparently assume that the risk of injury or death outweighs the thrill of riding without a helmet. But some motorcycle enthusiasts might well value the thrill over added safety. Consider also travel bans to dangerous countries such as Somalia, or Syria during the 2012 uprising. It may be true that such travel isn't what is best for *you*, even by your own lights. However, the ban assumes that you attach greater significance to your own well-being than you do to, for instance, documenting the events in such war-torn countries. The ban, then, imposes at least a weighting, attaching greater significance to personal well-being than other ends. But surely some, e.g., brave war-correspondents, do not weigh their personal well-being quite

so highly. The problematic feature of paternalism, then, might be “replacing” the values of the target with the values of the paternalist.

But paternalism need not involve the imposition of values. This becomes apparent if we remind ourselves that people often fail to do what best reflects their values. Consider two ways this might happen. First, a person might choose inadequate (or suboptimal) means to their ends, e.g., optimistically leave only twenty minutes to travel from home to the airport when in all likelihood it will take longer. Someone – perhaps the prospective traveler’s spouse – might coercively interfere in such cases so as to ensure that the person (efficiently) realizes his ends, e.g., by removing his breakfast plate before he is done. This seems paternalistic, as it is done for the traveler’s sake. However, it needn’t involve imposing values.

More strikingly, our potential traveler might be akratic. It is of paramount importance to him to get to the airport to make his flight. He knows that a leisurely breakfast would be nice but is not as important as making his flight. And yet he finds himself lounging over breakfast despite the whispers of his conscience. In this case, too, someone might coercively intervene for his sake. And here it is only the paternalizer who seems to duly respect the potential traveler’s values. Though it is less explicitly discussed, this may in fact be paternalism’s most common form. It is very much like the Kantian paternalism in Michael Cholbi’s contribution to this volume, or the “loose” paternalism described by Dan Scoccia in his. The paternalist, of course, is meant to see himself as acting for the target’s sake, and it’s difficult to sincerely believe you’re doing so while knowingly imposing alien values on the target. If that’s correct, it’s not merely that paternalism does not necessarily “impose” values – the paternalist is also, in some sense at least, committed to avoiding doing so. If paternalism is even *pro-tanto* wrong, it does not appear to be because it involves failure to respect its target’s values.

Alternatively, one might think that respecting someone as free and autonomous has nothing to do with anything as rarefied as “respecting her values.” Paternalism is not suspect for this reason. Instead, what is suspect about paternalism is more on the ground-level: Provided I’m competent and acting voluntarily, any choice or decision I make, assuming it doesn’t threaten others or violate others’ rights, is mine to make. What requires respect is not my “values” but my *decisions* – whether or not they reflect my values.

This position certainly has a resonance and more than its share of popular currency. But it has at least three odd features. First, it’s hard to see how a practically rational person can really insist on such a thing. A practically rational person must assess their choices in terms of their values. Consider the first-person point of view. From this point of view we

do not prioritize our decisions over our values: When someone recognizes that she has failed to do what's best from her perspective, either due to *akrasia* or because she now detects an error she made in applying her values, she must also see herself as having made a *mistake*. Presumably this is because one cannot genuinely endorse choices that are not reflective of one's values – for endorsement conceptually requires appealing to these very same values. It's puzzling, then, why the priority should be reversed in *second-* or *third-personal* cases – that others should prioritize your decisions over your values. It seems strange that the would-be paternalizer must respect the choices that you (the target) do not. To put it differently, if there's some choice your evaluative perspective demands, and you're about to fail to make that choice, how can it be that that very same perspective must demand that someone else not help you make it?

Second, it's not at all clear that any norm to respect choice really is thoroughly anti-paternalistic. After all, we actually do sometimes choose paternalism. Some of the most paternalistic states are democratic, and occasionally paternalistic measures are chosen by majorities via referendum. Indeed, it's at least possible that such measures could be the object of unanimous consent. This leads to paradox for the anti-paternalist who champions respecting decisions, as people have decided to institute policies that will constrain their decisions. Prohibiting us from adopting such measures even if we all want them seems objectionably paternalistic. This might be called the paradox of anti-paternalism.

Of course, the anti-paternalist might reply that consenting to a measure effectively rules it out as paternalistic. But this response has an unacceptable implication: It entails that those societies that are most tolerant of and endorse paternalism will thereby be the *least* paternalistic. The anti-paternalist might here respond by distinguishing between respecting a person's current decisions and respecting a person's decisions across time. Paternalism, they might note, requires third parties to do something not *presently* welcomed by its target – and it's the agent's present choices or deliberations that we must respect. So, though we voted for the seatbelt law, we're still targets of paternalism when we're issued the unwanted citation – and it is our decision *now* not to be treated this way, not in the way we decided in the voting booth three years ago, that's owed respect. But why – why is it that it is my decisions *now* that deserve respect?¹¹ After

¹¹ The relative significance of *present* vs. *past* decisions/consent is explored in various places, including Feinberg, *Harm to Self*, 81–87; Parfit, *On What Matters*, vol. 1, chap. 8; and Michael Cholbi's contribution to this volume.

all, it is plausibly held that the decision in the voting booth is the product of my more considered view, made in a cool hour.¹² One traditional (libertarian) answer appeals to self-ownership, according to which persons have the most extensive set of stringent rights over themselves and their powers that is logically compatible with all other persons having an equally extensive set of stringent rights over themselves and their powers. However, strong views such as this seem to prove too much, as they seem to rule out even soft paternalism. For the most extensive set of stringent rights that are compatible with all other persons having an equally extensive set of rights over themselves and their powers includes a right against soft-paternalistic interference.¹³

In the cases above we've tried to identify a feature of paternalistic acts that can explain why paternalism is at least presumptively problematic. And yet in each case we find that paternalism may not always possess the problematic feature, or that the proposed explanation is otherwise problematic. So whether we look at Mill's classic defense, broader consequentialist variants of it, or even the most natural articulations of our "folk" intuitions about respecting persons, what makes paternalism, in general, presumptively wrong eludes us. If we're looking for a wrong-making feature common to all paternalistic acts, perhaps our best bet is to return again to what is characteristic or definitive of paternalism. So perhaps returning again to Shiffrin's characterization will offer some insight. When we look carefully at her characterization, one approach immediately suggests itself – an approach suggested by the very term "paternalism": Paternalism objectionably treats its targets like children.¹⁴ Roughly following Shiffrin's lead, it seems that paternalism (by Y towards X) always involves some degree of *deciding for X* on the basis of X-regarding reasons (e.g., X's well-being, ends, values, desires, health, happiness, or even her freedom or autonomy). But why is this akin to Y treating X as his child, and what exactly is wrong with that? An initial answer is that Y, like a parent, supposes an authority to decide for X – an authority Y does not have. But this explanation is at best incomplete, and at worst begs our question. We cannot simply assume that third parties don't have this authority, as this assumption would presuppose paternalism's problematic nature, rather than explain it.

¹² The related conflict between our "consumer" preferences and our "political" preferences is famously addressed in Sagoff, "At the Shrine of Our Lady of Fatima."

¹³ This point is much more thoroughly made by Wall, "Self-Ownership and Paternalism."

¹⁴ An influential piece that reveals a connection between paternalism and the relationship between parent and child is Schapiro, "What Is a Child?"

An improved answer might be reached by noting a second way that Y's act treats X like a child. Specifically, the paternalist treats Y as though, like a child, she cannot effectively deliberate for herself. After all, under the description above, Y's act may actually take on board X's *inputs* into practical deliberation – her ends, values, and/or interests – and so the rationale for Y's paternalism must be his superiority in determining which courses of action those inputs recommend. But we might well wonder why this is wrong, or even *pro-tanto* problematic, for surely some people, especially in special domains and circumstances, *are* better placed to deliberate for me.

Nonetheless, there are at least two candidates for what is wrong with such an assumption of superiority. First, if Y's paternalistic actions make sense only if Y believes that the target, X, is less capable of deciding for herself, then we can say at least this much about paternalistic acts in general: They *express* superiority over the target.¹⁵ Of course, the relevant superiority expressed is not anything as grandiose as, say, superior moral standing. Instead, it appears to be an expression of one's having superior technical expertise – an expertise in deliberation – which is rather like expressing one's superior intellectual powers. That this may be morally problematic is not hard to see. For even though we may well be more intelligent than about half the population, it's at least *pro-tanto* wrong for either of us to express our belief that we're your intellectual superior. Indeed, public endorsement of such acts, in the form of state-sanctioned paternalism, may seem to thereby objectionably express the view that the citizenry isn't equally suited to make political decisions – for if a person does not deliberate well enough to be sovereign of her own affairs, she's even less likely to deliberate as well as third parties in public affairs.

While this line of argument may be worth exploring, it's also dissatisfying in a number of ways. First, it seems that what's wrong with paternalism is more than what is wrong with an inappropriate expression of superiority, which seems no more than an insult. More importantly, the superiority expressed by paternalistic acts may express no *intrinsic* superiority over the target. One of us may act paternalistically towards you because we recognize that you're currently in a context or state where people (including us) happen to make frequent mistakes in instrumental reasoning. And so the relevant superiority expressed by paternalism in such cases may be local,

¹⁵ Elizabeth Anderson, *Value in Ethics and Economics*, is best known for emphasizing the importance of expressive considerations.

temporary, and circumstantial, and thus consistent with a firm belief that the target is all-things-considered one's equal, even one's superior.

Second, let us briefly consider a very different kind of explanation why X may not have the authority to paternalize Y. This explanation appeals not to the fact that Y's act would express Y's superior deliberative powers, but focuses instead on another assumption that makes Y's act intelligible – his assumption that he knows well enough what X's ends, values, or interests are. Here the worry is not Mill's complaint that Y probably does not really know what's best for X, or what matters to her. We might even assume that Y has a good sense of what's most important to X, perhaps even a better sense than X herself. The problem, however, is that this information might be thought to be in some important way private. For this reason, others are not entitled to know it – to know which personal values shape my vision of what makes my life worth living – never mind to employ it in paternalistic fashion. So while you might have only true beliefs about what I care most about, and these beliefs may be justified, you are nevertheless not entitled to these contents, just as, perhaps, you are not entitled to the contents of one of our diaries. The paternalist, then – especially those who are not close to us, or the state – might be accused of objectionably using private information – information one is not entitled to use to make a public decision. What is wrong with paternalism, then, is that its effective exercise would require a kind of invasion of privacy.¹⁶ One might reply that effective paternalism does not require intimate knowledge of one's personal values and value weightings. This is true. But, even so, when we paternalize *without* claiming a sensitivity to a target's particular evaluative perspective, then we act as though we know what's best for the target, regardless of the content of her particular conception of the good. The paternalizer thereby reveals a willingness to impose values on the target – something that may be objectionable in its own right, or problematic as a violation of liberal neutrality. Though an approach like this certainly needs development, it offers a kind of disjunctive account of one thing that may be, at least, *pro-tanto* wrong with paternalism: It either constitutes a violation of one's privacy or disregards one's evaluative perspective.

We think this is an interesting proposal. But the point here is not so much to push for a particular answer to the central question of what is presumptively problematic about paternalism as to illustrate just how

¹⁶ This may again be part of treating one like a child – since children don't have the same privacy rights as adults.

difficult answering the question is, and, thus, how fundamental questions about paternalism remain open.

As noted at the outset, interest in paternalism has been heightened recently by the publication of Thaler and Sunstein's *Nudge*. They tout an approach to policy-making that is at once paternalistic but will not offend its traditional (libertarian) critics. Their approach – “libertarian paternalism” – advocates policies and practices that push, or “nudge,” people to make better choices without foreclosing options or significantly increasing their costs. Such policies are paternalistic insofar as they nudge people; it is libertarian insofar as it preserves freedom of choice.¹⁷ For instance, in their now famous school cafeteria example, food is arranged in such a way that common cognitive biases will lead people to make healthier choices. Less healthy options are still available, and selecting them requires no special effort or cost. However, strategic placement of the healthy food nudges people towards healthier choices. Thaler and Sunstein's approach is worth mentioning here not merely because of its implications for (and burgeoning impact on) public policy, but because it bears on what, if anything, makes paternalistic policies inherently problematic. If even the libertarian – the traditional arch-nemesis of paternalism – can happily endorse their approach, perhaps we should confidently say there's sometimes nothing wrong with paternalism at all.

At the heart of Thaler and Sunstein's defense of libertarian paternalism is a body of psychological research showing that people's choices are influenced by small and apparently insignificant aspects of the choice environment, or what they call the “choice architecture.” For example, there is what has come to be known as the status-quo bias: People have bias towards the current state (rather than change). Thus, choice is greatly influenced by a default option, if there is one. Another example is the tendency to treat gains and losses differently: People tend to be risk-averse when it comes to potential losses, and risk-seeking when it comes to potential gains. Thus how a choice is presented or “framed” – whether it is in terms of gains or losses – can influence choice. Famously, people choose differently if they are facing a medical decision and a certain course of treatment is described as offering (a) a seventy-percent chance of survival, versus (b) a thirty-percent chance of death. Thaler and Sunstein recommend that policy-makers, or “choice architects,” in both the public and private sector take these kinds of psychological tendencies into account and design choice situations in such a way that people are more

¹⁷ Thaler and Sunstein, *Nudge*, 5.

likely to choose what is in their best interest. The cafeteria is their central example: Place the food in such a way that people are more likely to choose healthier food. Many other examples appeal to the status-quo bias, and call for setting default options so that if the person does nothing she ends up with an option that (maximally) promotes her well-being, e.g., a retirement program structured so as to make automatic contributions and decrease the level of risk of the investments as retirement approaches.

Opponents of paternalism might insist at this point that nudging, even paternalistic nudging, should be avoided – better not to nudge at all. People should be not just free to choose, but free to choose without such influence. Thaler and Sunstein’s response to this is that the minor factors that influence choice in this way are ubiquitous. Therefore, the choice-situation will influence choice no matter what:

[I]t is [not] possible to avoid influencing people’s choices. In many situations, some organization or agent *must* make a choice that will affect the behavior of some other people. There is, in those situations, no way of avoiding nudging in some direction, and whether intended or not, these nudges will affect what people choose . . . In this respect, the antinudge position is unhelpful – a literal nonstarter.¹⁸

Given this inevitability Thaler and Sunstein think that choice architects are justified in nudging people towards what is in their best interest. If influence on choice is inevitable, then why not structure the choice environment so as to benefit choosers?

There are a number of problems with this reply. First, if influence is indeed pervasive in this way, then it seems to make little sense to talk in terms of what people would choose without such influence – what they would choose “for themselves.” There is simply no fact of the matter. This is a problem for Thaler and Sunstein to the extent that they maintain that libertarian paternalism nudges people towards options better *from their own point of view*, and, as such, does not impose values on the targets of nudges. The values imposed may be very general, e.g., health and wealth. They are, nonetheless, values ascribed to those who are nudged. Thaler and Sunstein are not blind to this. For they include amongst the options for the choice architect in the cafeteria example for her to try to arrange the food to get the kids to pick the same food they would choose “on their own.” However, they note, this is difficult to implement because there is no clear sense of what the children would choose on their own, as there is

¹⁸ Ibid., 10–11.

no such thing as “neutral architecture.”¹⁹ Nonetheless, they continue to describe libertarian paternalism as making choosers better off by their own lights, or “*as judged by themselves*.”²⁰

Second, even if it is not possible to avoid influencing people’s choices, it may be that intentional influence and unintentional influence are not morally equivalent. Though there are exceptions, most people think that there is an important difference here – even if what is not intended can be foreseen. And if there is an important difference here, then choice architects have available to them the option of eliminating *intentional* influence altogether. In the cafeteria example, choice architects might, for instance, use aesthetic considerations to determine how to arrange the various items, or considerations of efficiency. These will have an influence on choice. However, the influence – the direction of the nudges – will be, at most, a foreseen consequence of their decisions. Thaler and Sunstein do suggest an approach that doesn’t aim to influence choice in the cafeteria example: Place the different food items at random. They rule this out on the grounds that this will lead to different arrangements of the food at different schools, with the result that children at some schools will have less healthy diets than those at other schools. But it is unclear that this is sufficient. First, libertarian or not, one might think that nudging people is worse than any inequality that might result from neutral approaches such as randomly ordering the food. Second, since school cafeterias serve hundreds of meals per year, we should expect no significant differences between schools and the influence on children’s food choices over the course of the entire year.

Even for those suspicious of a morally significant difference between intended and unintended consequences, there is a third problem with Thaler and Sunstein’s reply that influence is pervasive and therefore unavoidable – that the anti-nudge position is a “nonstarter.” As Jamie Kelly argues in his contribution to this volume, the fact that the choice architecture inevitably influences does not settle whether we ought to exert *paternalistic* influence. For there are alternatives to nudging people towards what is most in their individual interest. For instance, Kelly suggests, people might be nudged towards choices that are best from a utilitarian point of view – towards what is best for the population as a whole. Alternatively, they might be nudged in the direction of choices that fit some other social ideal such as a Rawlsian conception of justice.

¹⁹ Ibid., 2–3.

²⁰ Ibid., 5. This issue is discussed further in J. S. Blumenthal-Barby’s contribution to this volume.

Thaler and Sunstein do not consider these among the options available to the choice architect in the cafeteria example. Amongst the options they do consider are arranging the food so as to maximize profits, or so as to maximize the sales of items from suppliers willing to offer the largest kickbacks.²¹ It is relatively easy to see why nudging people in the direction of healthier food is justified over these alternatives (at least in a public school cafeteria). However, it is not nearly so clear why libertarian paternalism is justified over what Kelly calls libertarian utilitarianism or a Rawlsian nudge.

Of course there is a well-known objection to utilitarianism: Insofar as it aims to maximize good in the aggregate, individuals can be sacrificed for the greater good. As such, John Rawls famously objects that utilitarianism does not recognize the “separateness of persons.”²² The same charge, it seems, could be leveled against libertarian paternalism, due to “non-standard actors.” While spinach is a healthy food for most people, it may not be for some, who are for this reason non-standard. It might be rather bad for them – positively harmful. And of course this phenomenon is pervasive: What is good for a majority is not good, can even be very bad (possibly deadly), for a minority. Paternalistic nudges, therefore, risk harming a portion of the population. Perhaps, then, choice architects should instead be driven by a “no harm” principle. Of course, since spinach is healthy for most people, and may only harm a few, the “public” or “common-good” is promoted by nudging in this generally healthy direction. However, what examples like this show is that the common-good can come at the expense of the minority, who are, in a sense, being used, or at least sacrificed, as a means to the production of well-being for others. One needn’t be a libertarian to be suspicious of policies that work this way. One might suggest a policy of implementing only nudges that are most likely to benefit and least likely to harm. Indeed, Thaler and Sunstein suggest something along these lines. However, it seems that few if any nudges can satisfy this demanding criteria, as minimizing harm will likely significantly reduce or eliminate the benefits. Moreover, surely libertarians will prioritize avoiding harm over benefiting society as a whole. Indeed, this is one of the hallmarks of libertarians, and classical liberals in general. John Hospers, in one of his memorable statements describing libertarianism, asserts that “[n]o human being should be a nonvoluntary mortgage on the life of another.”²³ The non-standard actor,

²¹ Thaler and Sunstein, *Nudge*, 2. ²² Rawls, *A Theory of Justice*.

²³ Hospers, “What Libertarianism Is.”

who is intentionally nudged to make harmful choices just so that others' lives can go better, seems to be just such a mortgage.

A different criticism has been leveled against Thaler and Sunstein's libertarian paternalism by Daniel Hausman and Brynn Welch.²⁴ They argue that the nudges Thaler and Sunstein recommend fall into one of two categories: either (i) they are not paternalistic at all, or (ii) they should not be acceptable to libertarians. Some fall into the first category because they simply provide information, give advice, and engage in rational persuasion. As such, they claim, these practices treat decision-makers as fully competent, and are thus not paternalistic.²⁵ The remainder, they argue, are paternalistic but do not adequately allay libertarian concerns because simply preserving freedom of choice is not sufficient to preserve autonomy:

[T]hose who have been worried about the ways in which government action and social pressure limit liberty have been concerned about liberty in a wider sense than closing off alternatives or rendering them more costly . . . call . . . this wider sense of liberty "autonomy" – the control an individual has over his or her own evaluations and choices. If one is concerned with autonomy as well as freedom, narrowly conceived, then there does seem to be something paternalistic, not merely beneficent, in designing policies so as to take advantage of people's psychological foibles [even if] for their own benefit.²⁶

In other words, insofar as such nudges take advantage of flaws in human decision-making, they diminish the control people have over their evaluations and choices because their decision-making "reflects the tactics of the choice architect rather than their own evaluation of alternatives." As such, the choices and local evaluations of those nudged are "shaped" by others – and this certainly appears to be the "substitution in judgment" thought to be characteristic of paternalism. Such shaping, which some might call manipulation, seems incompatible with a libertarian concern for liberty construed in this broader way.²⁷

Transparency might be put forward as a solution to both the problem of non-standard actors and Hausman and Welch's charge that the shaping of evaluation and choice is objectionable (to the libertarian) even if freedom of choice is preserved: These problems are avoided if the use of nudges is a matter of public record. Indeed, Thaler and Sunstein themselves call for a publicity condition.²⁸ And in this volume J. S. Blumenthal-Barby argues that transparency should quell libertarian qualms. Thaler and Sunstein do not make it clear whether transparency must be explicit – requiring

²⁴ Hausman and Welch, "Debate." ²⁵ *Ibid.*, 127. ²⁶ *Ibid.*, 128. ²⁷ *Ibid.*, 127–129.

²⁸ Thaler and Sunstein, *Nudge*, 246–249.

actually informing those subject to influence that they are – or whether it requires merely a *willingness* to defend such policies publicly. But either way it is unclear that it really helps. First of all, nudges often rely on deeply embedded psychological tendencies, ones that likely influence evaluation and choice whether we like them to or not. They are, to a disturbing degree, automatic, part of the (largely sub-conscious) “fast thinking” system described by Daniel Kahneman in his most recent book.²⁹ As such, it seems to make it no better, and possibly worse, either if subjects are explicitly informed or warned about such nudges or if the choice architects would be willing to defend such policies publicly. If one of us has a love potion that will make you fall in love with one of us despite your current loathing, we are not morally off the hook if we tell you about it before we inject it into your veins. Indeed, arguably, this only makes it worse – converting it from illegitimately controlling a person to positively torturing them. Helplessly watching yourself fall victim to such influence – your agency being snatched away – seems particularly horrific in much the way being eaten alive (by a shark, for instance, *à la* Captain Quint in *Jaws*) seems a particularly horrific way to die.³⁰ Of course injecting you with a love potion against your will is not merely a nudge. But the point remains: Transparency doesn’t mitigate untoward influence if the influence remains effective. Consider, alternatively, distorting mirrors that make people look fatter than they actually are, an example discussed by Thaler and Sunstein, and by Blumenthal-Barby in this volume. These can nudge people towards eating less, or more healthily. But a sign announcing that a mirror is distorting seems to make little difference morally if people will inevitably be influenced by the image in the mirror. Indeed, Thaler and Sunstein seem to recognize the general point, as they think it is not enough to disclose the use of subliminal advertising because knowing of its use is not enough to counteract its effect.³¹

Perhaps we have underestimated our capacity to resist the deeply embedded psychological tendencies that nudges take advantage of. If so, then perhaps transparency can make a moral difference because, by alerting us to fact that we are being nudged, it gives us the opportunity to resist. But this doesn’t obviously help. For we either resist or we do not: Either

²⁹ Kahneman, *Thinking, Fast and Slow*.

³⁰ There is some connection here, we think, with David Sussman’s interesting account of what makes torture the particularly deep moral transgression that it is. See his “What’s Wrong with Torture?”

³¹ Thaler and Sunstein, *Nudge*, 248–249. They express the concern by simply saying that such manipulation is objectionable because it is invisible and thus impossible to monitor. They also note that knowing about the “planning fallacy” (*ibid.*, 7) does not help with resisting it.

we choose the less healthy items in the cafeteria despite being nudged towards the healthy ones or we indeed choose the more healthy ones. If we resist and choose the less-healthy items then the nudge has been unsuccessful. As such, this would not constitute an instance of morally acceptable paternalism; it is only *attempted* paternalism. If instead we do in fact choose the more healthy options, it is unclear whether this counts as paternalism either. For now it seems more like the cafeteria patron has been simply reminded by the sign to think (explicitly) about eating more healthily. What leads him to the healthy choice is that he has been reminded of the importance of eating healthily, and not because of the psychological mechanisms nudges target. In other words, transparent choice architecture appears to be more akin to providing information, giving advice, or engaging in rational persuasion. And as Hausman and Welch pointed out, this is not paternalism at all.

All that being said, it is hard to avoid thinking that many of the nudges Thaler and Sunstein recommend are rather benign. Do we really need to get all up in arms about setting defaults that enroll people in retirement programs? Perhaps not. But there are things to be worried about here. First, there remains the problem of non-standard actors: Someone with a very short life expectancy could well be harmed by being enrolled in a standard retirement plan if it involves contributing a portion of his income, as he will never enjoy the benefits in his retirement. Similarly for someone set to inherit a large sum upon retirement (at age sixty-five). And again, although every choice architecture will have its “victims,” when a nudge is *intended* the unfortunate have special grounds for complaint: The state tried to push them towards the disastrous option, knowing that for some this push will be successful and disastrous. And though one might reply that “push” is too strong a word – because individual “nudges” (by stipulation) are easy to resist – this reply is little comfort given that Thaler and Sunstein never address the concern that nudges might be “stacked” within and across choice architectures, e.g., the food could be ordered a particular way, presented a particular way, listed on menus in most attractive fonts, etc. This possibility clearly illustrates that there can be more or less “nudging” in an environment, and again raises the question whether trying to reduce or avoid these influences is the best option, even if some influence is inevitable.

Thaler and Sunstein would presumably object to “stacking” nudges because collectively they might constitute a “shove.” Thus proper use of the power to nudge will require choice architects, and those who monitor them, to be self-conscious and careful about the manner and degree to

which they influence our preferences. Put plainly and unflatteringly, they need to adopt their role as a manipulator of preference and be vigilant about not misusing it. But this leads us to our final worry about the libertarian paternalist program. Specifically, there's something unsettling, maybe even something somewhat incoherent, about viewing oneself as one who manipulates preference because it is less objectionable than coercion. Indeed, though many of their proposed nudges seem innocuous, libertarian paternalism, when considered at an abstract distance, can seem even more sinister than its coercive or option-removing cousin. To get a sense of our worry, first notice that debates about paternalism can be fundamentally framed in terms of conflicts between an individual's choices and her well-being. We will not always choose what's best for ourselves, or even what is best by our own lights. When we won't the paternalist finds it sometimes justified to act to prevent these imprudent or subjectively "sub-optimal" choices. Traditional paternalism typically pushes us away from these choices by dramatically changing the cost of making such decisions. State paternalism often imposes fines or penal sanctions for making such choices. Less controversially, the state can attach incentives to choices deemed best for its citizens, thereby making putative suboptimal choices "costly" in terms of benefits foregone. These measures rarely involve actual physical force; instead they make choosing in a way deemed "imprudent" costly – costly relative to the target's own evaluative perspective. The traditional paternalist thus takes efforts to artificially ensure it is irrational (by the subject's own lights) to make choices deemed imprudent, but he does not alter the target's fundamental evaluations. Rather he changes the external environment in a way that makes those fundamental evaluations now favor the "prudent" options. This is clearly in some sense disrespectful to the person and *his perspective* about what is to be done, because he might have otherwise wanted to choose something else. In rigging the external cost structure "for his sake" the paternalist effectively disregards his ability to make prudent decisions, and disregards his own, would-be view about what choice to make.

Libertarian paternalism might initially be thought of as operating in exactly the same way, the only difference being that the external costs imposed are somewhat smaller. If that is so, then we have a difference in degree and not kind: The traditional anti-paternalist will view the program as unjust but presumably better than more traditional forms of paternalism. However, that is probably the wrong way to view many nudges. The biases of our automatic systems do not reflect the weightings of options relative to our evaluative perspective – we do not judge that we have more

reason to choose items that we are offered first, or that we have more reason to choose a default. We do, however, judge ourselves as having reason to seek financial incentives and avoid fines. In this way, the traditional paternalist at least treats us as rational by appealing to what we value, and what we ourselves take ourselves as having reason to do (even if only like the gunman who says “your money or your life”).

The libertarian paternalist, as manifested in at least some of Thaler and Sunstein’s proposed nudges, takes measures to make us act on features we do not take to be reason-giving. Effectively, these nudges change our (relative) evaluation of options without appealing to our current evaluations or view about what provides reasons. Seen this way, libertarian paternalism involves a much deeper disrespect of its target. The traditional paternalist may be accused of *ignoring* the target’s own view of what is to be done, or worse, *turning that own view against her* by changing the external environment so that the “favored option” now becomes rational relative to her own point of view. The libertarian paternalist, on the other hand, does not ignore the target’s point of view, or use it against her; rather he sees her perspective itself as something to be fashioned or altered. But if we think of the agent as partly constituted by her evaluative perspective, libertarian paternalism is apparently a threat to the agent herself, and not just the unimpeded expression of her evaluative outlook. Our targets are no longer beings whose choices, but not evaluations, are to be guided. Rather, we’re engaged in a project that more directly shapes the target’s preferences and evaluations themselves. If traditional and libertarian paternalism really are distinct in this way, libertarian paternalism – though non-coercive – may seem worse and less justifiable than more traditional forms. It may seem worse because the target’s view of *what is to be done* is no longer merely ignored or thwarted, but is rather viewed as something one may attempt to extinguish, supplant, or “manage.” There may be a deep tension in thinking our viewpoints are subject to management, but our actions should not be. One might alternatively put our preliminary worry this way: If traditional paternalism is somehow guilty of “violating” our rights to autonomy or self-sovereignty, libertarian paternalism treats us as beings that could not even be owed such rights.³²

³² Many of the points being made here are consistent with Hausman and Welch’s criticism, and could plausibly be seen as explicating and expanding upon the worry they only begin to develop in their review article. They suggest that choice architects shaping people’s evaluation and choice is incompatible with autonomy. This seems right. However, no explanation of the incompatibility is provided. The comments here, though themselves not fully worked out or conclusive, might plausibly be seen as a beginning of providing just such an explanation.

Ultimately, it remains unclear what precisely, if anything, is wrong with paternalism. And answering this question has become even more difficult as we've become more aware of the fact that it needn't be accompanied with the dubious features typically associated with it: coercion, removal of choice, and the imposition of values on its targets. Though this suggests that paternalism is benign in the absence of these features, as we've just seen, it may be possible to argue that some paternalism without these features may be worse. Of course, not much rides on whether paternalism always has a bad-making feature or not. Our trouble in identifying such a feature simply reveals the complexity of the issue. There are no easy answers. It appears unreasonable to reject paternalistic policies because they are paternalistic; but equally unreasonable to welcome paternalism when it lacks the evils typically associated with it. Further investigation is needed, but its promise needn't merely be a clearer view about if, when, and how to paternalize. It promises to further illuminate more general moral issues about the relationship between persons, the state, and its citizens, and the substantive content of central rights and values. We view all the works in this volume as valuable contributions to this investigation.