

Recent Work on:

Paternalism

JESSICA BEGON

Draft. Final version published in *Analysis*. Available at:

<http://analysis.oxfordjournals.org/cgi/content/full/anw040? ijkey=01gLPmOkhfcRmE&keytype=ref>

<http://analysis.oxfordjournals.org/cgi/reprint/anw040? ijkey=01gLPmOkhfcRmE&keytype=ref>

1. Introduction: Laying the Conceptual Groundwork

It is the habit in much recent literature on paternalism to begin by setting up anti-paternalism as the status quo, and then outlining a supposedly controversial argument for some limited form of paternalism. Yet this form is so common that it is difficult to see anti-paternalism as the status quo any longer. Certainly, ‘paternalism’ is often used as a term of disapprobation in contemporary discourse, and we have many anti-paternalist intuitions, resenting institutions or individuals who interfere in our lives for what they take to be our own good. If the crux of paternalism involves treating an adult like a child – treating someone who is competent to direct their own life as if they were not so competent – then many would consider themselves anti-paternalists. Yet we also have many paternalist intuitions: many believe the state should prohibit slavery contracts, forbid the sale of heroin, and require us to wear our seatbelts.¹

There is little straightforward debate between ‘paternalists’ and ‘anti-paternalists’, then. Even those at the anti-paternalist end of the spectrum only argue that paternalism is *presumptively*, or *prima facie* impermissible. Instead, much current debate centres on how paternalism should be defined, and why it might be wrong, and it is these debates I review here. I begin by considering whether paternalism should be understood as interference with our liberty for our own good, or only as restriction of our voluntary or autonomous acts (§2). If paternalism cannot be defined by reference to the restriction imposed on the paternalised agent, it can perhaps instead be defined by reference to the unique insulting motive of the paternaliser: the distrust of another’s ability to choose well for themselves, and an assumption that they can choose better (§3). Whilst

¹ Indeed, even the supposed poster children for liberal anti-paternalism – J.S. Mill (1974) and Joel Feinberg (1986) – allow for some such, seemingly paternalist, prohibitions (most notably, of voluntary slavery). Further, there is evidence to suggest the general public does not always object to paternalist policies and institutions (Blumenthal 2013; Sunstein 2015a).

this approach seems promising, some have argued that such a motive is neither unique, nor need be insulting, given that we *are* all poor reasoners, at least some of the time (§4). Finally, then, I will consider the idea that the normative core of paternalism is not that we mistrust individuals' abilities (which may be justified), but that we fail to see their autonomy as grounding an exclusionary reason against using this distrust as a basis for our actions (§5).

First, though, given the proliferation of (sometimes contradictory) terminology in this field, it is worth beginning with a brief sketch of the conceptual landscape. Paternalism is standardly defined as “interference of a state or an individual with another person against their will, and defended or motivated by the claim that the person interfered with will be better off or protected from harm” (Dworkin 2014).² The key elements here are that an individual is prevented from, or manipulated into, performing an action or making a choice, on the basis that this will be in their own best interests, and usually the associated assumption that they are not the best judge of their interests. This is ‘hard paternalism’.

Anti-paternalists object that such interference “wrongly subordinat[es] the individual’s right to her good” (Arneson 2005: 260). The strongest form of anti-paternalism – ‘hard anti-paternalism’ – suggests that we must never interfere with *any* self-regarding choice or action performed by an adult.³ Yet this leads to the implausible conclusion that we cannot interfere even if a choice is not autonomous or voluntary: if an individual is being coerced, if they lack knowledge of the relevant empirical facts, are in a disturbed emotional state, base their choice on mistaken reasoning, or have not carefully considered their decision.⁴ The most famous case is Mill’s (1974: 166) individual about to cross a rotten bridge: surely it would not be paternalist to intervene to ensure they were

² Dworkin (2013: 29-31) provides variations on this standard.

³ As noted, few are willing to adopt either hard anti-paternalism or hard paternalism, though Wall (2009: 399–341) has argued that consistency requires that libertarians take the former view. Arneson (1980) has also defended this view but has since changed his position (Arneson 2005). We may also make a pragmatic shift to hard anti-paternalism if we think paternalist intervention will be ineffective: the punishment we impose to prevent harmful conduct may be worse than what it aims to prevent (Husak 2013: 51). Sarah Conly (2013, 2016) defends a form of hard paternalism, justified by appeals to our frequent irrationality; as does Danny Scoccia (2008), on the basis that it is the only way to justify those paternalist policies that are plausible and desirable.

⁴ These are Feinberg’s (1986: 104-106) conditions of making a voluntary choice. Given that few decisions completely fulfil these conditions, Feinberg understands voluntariness as scalar. The problems this can pose for soft paternalism will be considered below. We may think some of these problems can be avoided by having a more binary definition of voluntariness, such as Olsaretti’s, where a non-voluntary choice is one that is performed because there are no acceptable alternatives (Olsaretti 1998; Colburn 2008), though I do not discuss this position here. Further, following the soft paternalist literature I take voluntariness to be an essential component of autonomy, though Olsaretti (1998: 73) disagrees with this characterisation.

aware of the bridge's state, so their decision to endanger their life was really voluntary. Such cases led Feinberg to suggest a more moderate form of anti-paternalism – soft paternalism – according to which we should not interfere with adults' voluntary choices (regardless of their consequences), but people can be protected from the harmful consequences of their *involuntary* choices.

Soft paternalism is sometimes used to describe paternalism that intervenes to help individuals achieve goals they have set for themselves; as opposed to hard paternalism which aims to ensure individuals achieve ends that are taken to be good for them, regardless of their opinion (Haybron and Alexandrova 2013). This distinction has also been described as one between weak and strong paternalism (Dworkin 2014), loose and strict paternalism (Scoccia 2013: 76), and means and ends paternalism (Le Grand and New 2015: 101-104). I will use soft paternalism in Feinberg's sense here.⁵ Although it has its roots with Mill, the distinction between hard and soft paternalism remains one of the most central to any discussion of paternalism, and this will be my focus.

2. Against Soft Paternalism: Is Autonomy All That Matters?

At first glance, soft paternalism seems clearly preferable to hard anti-paternalism, since it allows intervention to prevent non-voluntary self-harm. It is surely a concern for autonomy, rather than mere negative liberty, that motivates the anti-paternalist stance: the idea that individuals should be enabled to form and pursue their own conception of the good, without outside pressure, coercion, or manipulation. Further, that allowing someone to freely do what they want, when this undermines their real desires, is inimical to autonomy. As Feinberg (1986: 12) notes, choices arising from ignorance, coercion, or derangement may be as "alien to [us] as the choices of someone else". Yet despite this intuitive plausibility, soft paternalism and the focus on autonomy has become the subject of significant criticism in recent years, both on the basis that it allows for too much interference, and that it allows for too little.

We may not much care about intervention in our free choices in the case of the rotten bridge, but soft paternalist rationales can be used to justify wide-ranging interference. This is especially so when we note that soft paternalism is not only

⁵ Distinctions have also been drawn between pure or direct paternalism (in which the class being interfered with is identical with those whose welfare is improved), and impure or indirect paternalism (Dworkin 2014; Archard unpublished). Between legal paternalism (interference for the sake of individuals' physical welfare), moral paternalism (interference for their moral welfare), and legal moralism (interference for their moral improvement, where this is not necessarily good for the interfered-with individual) (Dworkin 2005).

used to justify preventing self-inflicted non-voluntary harm, but also to permit intervention to *check* if the harm is non-voluntary, and the general prohibition of harmful activities on the basis that they are often engaged in non-voluntarily.⁶ Further, since soft paternalists cannot be *certain* harm will occur in most cases, they will intervene in non-voluntary activities that carry some threshold level of risk of harm. As such, soft paternalism may be used to justify intervention in a variety of self-regarding activities.

Moreover, the process of defining the threshold level of autonomy itself may be paternalist. Certainly, this threshold is not obviously fixed, and where being classified as incompetent can lead to “fundamental exclusion” from society (Fateh-Moghadam and Gutmann 2014: 389) and severely curtailed freedom, we must be careful that our selection of this threshold does not perpetuate our pre-existing mistrust of certain groups, or those inclined to make certain choices. If this mistrust leads to unwarranted assumptions of non-voluntariness, then soft paternalism may be used as a pretext for limiting voluntary choices for the sake of interfered-with individuals, because we do not trust them to choose well, or indeed, trust them to choose voluntarily at all.⁷

Further, even those who are rightly judged to be incompetent might be thought to be entitled to *some* liberty of action (Fateh-Moghadam and Gutmann 2014: 391; Mullin 2014).⁸ Indeed, we are all incompetent – in the sense of not meeting all the criteria of voluntary choice – some of the time, and would not want our liberty of action interfered with in every case. Some may value the opportunity to make non-voluntary choices without interference: “a person’s settled values, attitudes, and desires may include a disposition to impetuous decision-making and an aversion to the careful consideration that renders choices voluntary” (Arneson 2005: 267). Whilst we may not wish to involuntarily harm ourselves, it is less obvious that all individuals will be averse to spontaneously exposing themselves to the risk of harm.

The concern that the focus on autonomy can lead to excessive interference in individual liberty is even more marked in approaches that aim to maximise our

⁶ For example: Feinberg (1986: 71-81; 124-127); Fateh-Moghadam and Gutmann (2014: 386); Arneson (2005: 272-273); Shafer-Landau (2005: 173; 176-181).

⁷ Similarly, I have argued elsewhere that we must avoid unwarranted assumptions that the preferences and decisions of marginalised and vulnerable groups are ‘merely adaptive’, and can be ignored on this basis (Begon 2015).

⁸ Daniel Groll (2012) offers an approach according to which the will of incompetent individuals can play a role in determining their treatment, though their will would not be considered *authoritative* in this calculation – as it should be for competent individuals. This approach will be discussed below.

autonomous choices.⁹ Steven Wall suggests such an autonomy-centred view whilst critiquing the libertarian account of paternalism. Wall starts from the observation that libertarians cannot uphold the distinction between soft and hard paternalism. A central feature of libertarianism is full self-ownership: allowing individuals the most extensive set of stringent rights over themselves and their powers that is compatible with others having such rights. Given that individuals have a fuller set of self-ownership rights if they are not subject to soft paternalistic interference, it seems libertarians must prohibit soft, as well as hard, paternalism (Wall 2009: 400-404), and so adopt the implausibly permissive hard anti-paternalism. In light of these difficulties, Wall proposes that libertarians might instead focus on full *autonomous* self-ownership: the most extensive control over one's autonomous choices.

One benefit of such an approach would be that it avoids Feinberg's reliance on harm: we do not need to make controversial judgements regarding what constitutes self-harm (does cosmetic surgery count?), nor need we determine a threshold level of risk of harm that triggers interference (riding a motorcycle; riding a motorcycle without a helmet; riding without a helmet whilst drunk?). Yet without the notion of harm to guide us, interference would be permitted in *any* non-autonomous choice, including those that "bring no, or only trivial, harm" (Wall 2009: 411).¹⁰ One solution might be to allow individuals stringent rights over autonomous choices, and non-stringent rights over non-autonomous choices. Thus, "self-ownership rights become more stringent as the autonomy of the choices they protect increases" (Wall 2009: 411). However, this will still require us to determine when choices count as autonomous *enough* to trigger stringent protection, and find some non-arbitrary way of setting a strict threshold on a scalar concept.¹¹

A contrary line of criticism of soft paternalism is that, far from allowing too much intervention, it does not allow enough. Once assured the agent is choosing voluntarily, there is no limit on the harm we are allowed to inflict on ourselves:

⁹ For example, Scoccia (2013: 77-78); and Nozick (1974) on Wall's (2009: 409) interpretation.

¹⁰ Due to the negative consequences of such interference, soft paternalists' all-things-considered judgement may not always favour intervention to prevent trivial harms, but Wall's objection is that they lack a principled reason to rule it out. This possibility is troubling given that Feinberg (1986: 94) insists "there is no such thing as a 'trivial interference' with personal sovereignty".

¹¹ Michael Cholbi suggests that we can interfere in non-autonomous choices (in which our "deliberation is distorted" (Cholbi 2013: 121)), only when they prevent us from achieving "our rationally chosen ends or conceptions of the good" (Cholbi 2013: 125). This allows a restriction on the non-autonomous choices that are subject to intervention without appeal to harm, though it is not clear that identifying which of an agent's goals are central to their conception of the good will be any less controversial.

suicide, slavery, amputation, heroin use, and lobotomy might all be permitted if we can be assured that the individual passes the threshold of a procedurally voluntary choice. These cases may seem especially troubling when we remember that to be immune to interference these choices need only be voluntary enough, where this determination “depends on a variety of factors, each of which is realisable in degrees and none of which is commensurable with one another” (Shafer-Landau 2005: 190).

Further, for those who believe that we have duties to ourselves – in virtue of our dignity, not to waste the opportunity provided by having a life (Arneson 2005: 279-281) – the fact that a choice is voluntary (enough) may be deemed insufficient to change the permissibility of such acts.¹² The argument is not that the above acts are never permissible, only that they cannot be performed for “frivolous reasons”, simply because we have met the procedural conditions of voluntariness. For others, the concern is not that soft paternalism allows us to violate our duties to ourselves, but that it seems contradictory to allow our respect for individuals’ autonomous capacities to lead us to permit them to sacrifice them.¹³ If our concern is autonomy, where this involves having control over one’s life, and being able to act on the basis of deliberation and reflection on goals worth pursuing, then this should surely give us reason not only to protect choices, but also to *limit* them, when our choices undermine our capacity for control and deliberation (de Marneffe 2006: 72; 81-82; 2013: 70-73).¹⁴

Indeed, the central thread in much of the critical response to soft paternalism is that it leads us to place far too much weight on a line between voluntary and non-voluntary choice. As Arneson (2005: 268) puts it, the soft paternalist “has to urge the enormous overriding moral importance of the line between self-harming choice that is not quite voluntary enough and choice that just passes the threshold of being voluntary enough”. Many contend that this line is difficult to draw non-arbitrarily, and that it may not capture the motive underlying soft paternalism: a concern for individual autonomy, and a distrust of action that fails to respect this autonomy. Further, it may seem that “actions by themselves cannot be paternalistic”: paternalism depends on the reason why the action was carried out (Grill 2007: 445). Thus, a number of influential accounts have begun

¹² Also, Tadros (2011); Velleman (1999); Kant (2011: IV 429-430; 1996: VI: 421-447); Arneson (2013: 153-155).

¹³ On some interpretations, it is this that led Mill to his prohibition of voluntary slavery (Dworkin 1997: 73).

¹⁴ Similarly, Ben-Porath (2013) insists that paternalist policies may be necessary for democratic equality and meaningful choice. See also Quong (2011: 98); Wall (2003: 236; 2013: 105); (Sjöstrand *et. al.* 2013).

to focus not just on the nature of the choice interfered with (is it voluntary enough?), but on the *motive* of the putatively paternalist individual or institution (it is appropriately respectful?). Motivational accounts, to which I turn next, aim to identify paternalist acts on the basis of the uniquely insulting negative judgement that is taken to motivate all instances of paternalism.¹⁵

3. The Paternalist Motive

I will here consider some of the most influential and interesting motivational accounts of paternalism, beginning with Seana Shiffrin's (2000), which contains a number of important features, some of which have been widely incorporated into subsequent accounts, and some of which have not. Shiffrin (2000: 218) argues that an action is paternalist when A substitutes her judgement for B's, on the assumption that her judgement is superior to B's, and it aims to have (or avoid) an effect on "B's own interests or matters that legitimately lie within B's control". The characteristically paternalist motive, then, is distrust of an individual's agency regarding matters in her legitimate control, and an assumption that the paternaliser's judgement is superior.

Being motive-based means both that not all coercion is paternalist (depending on its motive), and that not all paternalism is coercive. For example, if we refuse to help someone build some shelves because we think it would be good for them to learn carpentry, and do not trust them to learn without this push (and not for some self-regarding reasons, such as the inconvenience of providing help) we treat them paternalistically, but without coercion (Shiffrin 2000: 213). Paternalism need not involve either the restriction of an agent's liberty, or the violation of an autonomy right. Otherwise legitimate (non-invasive or coercive) acts can become insulting, paternalist, and potentially impermissible on the basis of their motive.¹⁶ Indeed, Shiffrin (2000: 213-214) suggests that even *freedom-enhancing* behaviour may be considered paternalist: if we are given options we have declared we would rather not have (perhaps because we find too much choice overwhelming, or worry about yielding to temptation) then to provide these options anyway (perhaps because they think we would have a better

¹⁵ It should be noted that motivational accounts need not be incompatible with soft paternalism. Soft paternalists can, and do, agree that the problem with interference in voluntary self-harm is that it expresses a negative judgement of the agent's capacity. Motivational accounts *can* agree that the negative judgement is only expressed when the interference is with agent's voluntary self-harming choices – though most expand the domain of problematic interventions beyond this, and take it as a benefit of their approach that acts that neither violate an individual's autonomy rights, nor restrict their liberty, can still be considered paternalist, as we will see.

¹⁶ Also, Archard (1990: 37).

character if tested, or that the option we are trying to avoid is a good one) can be paternalist.

Shiffrin also insists that paternalism need not be motivated by a distrust of an agent's assessment of their interests. She argues that we can act paternalistically towards someone if we agree on what is in their best interests (and trust their ability to identify them), yet distrust their ability to effectively advance these interests. In other words, we trust their judgement, but distrust their will (Tsai 2014: 86). Thus, taking and hiding someone's cigarettes without their permission is paternalist, even if they have agreed they want to stop smoking.¹⁷ Moreover, paternalism need not involve acting *against* someone's will. If we act on an agent's behalf before they have had a chance to consider a matter and establish their intention, because we do not trust them to select the right outcome, then we fail to respect their autonomy, even if they would have favoured the outcome chosen for them. For example, if a doctor is concerned that their patient will refuse surgery, and so performs it before they have gained consciousness, they act paternalistically, even if the patient would have opted for the surgery (Groll 2012: 697-8).

All these acts are considered paternalist because they all involve ignoring or bypassing an individual's agency because we distrust it (and assume our judgement is superior). By contrast, if we were to instead discuss the matter with the subject of paternalism, and attempt a process of reasoned persuasion, then (according to Shiffrin) this would show an appropriate respect for their agency, and so avoid paternalism. Yet George Tsai (2014) has recently argued that it is possible for rational persuasion to be paternalist too. He insists that not all rational persuasion is on a moral par. In some cases, the provision of reasons may pre-empt someone's deliberative activities, and undermine their ability to be self-directing (Tsai 2014: 92-95). Thus, as in the above cases, an act may be paternalist not because it opposes someone's autonomy, but because it prevents them from exercising it. Whilst Shiffrin does not consider rational persuasion in this light, Tsai's arguments are nonetheless consistent with her account given that he, too, identifies paternalism on the basis of motive – one of “distrust in the other's capacity to adequately recognize or weigh reasons that bear on her good” (Tsai 2014: 111).

¹⁷ This is means paternalism, on the above terminology. Means paternalism is defended as permissible by some on the basis that it simply allows individuals to more effectively achieve their own ends (Conly 2013: 100-125; Le Grand and New 2015: 79-104); though means and ends may not be easy to distinguish (Archard 2013: 398-399).

However, it should be noted that there is one way in which Tsai's account differs from Shiffrin's: for Tsai the paternalist distrusts the individual to weigh reasons that bear *on her good*. Indeed, Shiffrin's account differs from the vast majority of paternalist accounts in insisting that a paternalist judgement need not be directed at improving an agent's welfare, or protecting them from harm. Paternalist interventions may also be directed at matters that lie within their legitimate sphere of agency, and whilst we may think that this sphere includes control over whether we harm ourselves, it is surely broader than this.

Shiffrin (2000: 217) suggests two sorts of cases that seem to involve the unique paternalist insult, but are not directed at the interfered-with agent's interests. In the first, the paternalist motivation is directed at the welfare of a third party. For example, a park ranger prevents someone from making a difficult climb because they fear their spouse would be grief stricken if the worst happened. In the second, the interference does not involve a concern for anyone's welfare. For example, someone asking a question at a conference is interrupted by another participant, who believes they could frame the point more eloquently, and proceeds to do so. In both cases, although the paternalist is not acting for the agent's good, the act is no less insulting: it 'treats us like children' for someone else to assume command over any domain of our life that is legitimately in our control, even if they are not doing it for our own good. In either case, Shiffrin (2000: 220) argues, "the paternalist's attitude shows significant disrespect for those core capacities or powers of the agent that underwrite and characterise his autonomous agency".

Few share Shiffrin's intuitions about these cases, though some arguments against this position seem to miss the mark. For example, Jonathan Quong argues that Shiffrin's focus on legitimate spheres of agency, rather than individual welfare, will lead to the misclassification of key cases. He gives an example in which he refuses to lend someone £50 out of fear they will use it to fund their heroin habit (Quong 2011: 79-80). He insists that since it is *his* £50, it is under his sphere of control, so Shiffrin cannot deem the refusal paternalist. Yet, surely, when we refuse to lend the money because we distrust how it will be spent (and not because we do not have £50 to spare) we aim to influence our friend's legitimate domain of control: something like, 'control over what they do to their body'. The focus on legitimate domains of control, then, need not exclude omissions or refusals to aid as paternalist. Nonetheless, the inclusion of interference in our duties to third parties (such as our spouse or children) and our right to pursue our projects (such as question-asking) within the scope of paternalism will

continue to strike many as counter-intuitive. Further, Quong's critique does bring to light the difficulty of determining what falls properly within the authority of the agent, and the absence of such an account in Shiffrin's approach.¹⁸

Quong's own motivational account aims to avoid these implications. This is 'the judgemental definition', where paternalism is defined as any act where:

1. Agent A attempts to improve the welfare, good, happiness, needs, interests or values of agent B with regard to a particular decision or situation that B faces
2. A's act is motivated by a *negative judgement* about B's ability (assuming B has the relevant information) to make the right decision or manage the particular situation in a way that will effectively advance B's welfare, good, happiness, needs, interests, or values. (Quong 2011: 80)

In particular, Quong (2011: 81-82) argues that paternalism involves a negative judgement about agents' practical reasoning, willpower, and emotion management. He insists that if someone blamelessly lacks some relevant information, then this does not entail a negative judgement nor ground paternalist actions (Quong 2011: 82). Presumably, though, if this lack is not blameless, a negative judgement to this effect may ground paternalism.

Quong (2011: 100-106) argues that this approach allows him to ground the wrong of paternalist acts not on the basis of the overriding importance of autonomy, but because to treat someone paternalistically, in his sense, will involve treating that individual as if they lack the capacity to plan, revise, and rationally pursue their own conception of the good – Rawls's second moral power.¹⁹ Being motivated by a negative judgement about B's capacities implies that the paternaliser treats B as having an inferior status: they trust their own capacities, but do not accord B's the same respect.²⁰ Further, paternalistic acts insultingly "involve treating an adult as if he or she (at least temporarily) lacks the ability to rationally pursue his or her own good" (Quong 2011: 101). This, Quong insists, gives us reason to believe that paternalist policies are presumptively wrong.

¹⁸ I have suggested how these legitimate domains of control might be identified in Begon (2013).

¹⁹ It may be objected that Quong still relies on autonomy insofar as disrespect for the second moral power *just is* disrespect for our capacity for autonomy (Mills 2013a). Alternatively, we may think he has in mind the failure to respect the *equal status* of citizens (Fowler 2014: 310-311).

²⁰ It has been argued that a negative judgment alone does not constitute a motivation: alongside a factual belief that the target of the prospective intervention is lacking in some relevant capacity, the paternalist must act for some end to which she ascribes value (Fox unpublished). Indeed, Nicholas Cornell (2015) has argued that paternalism should be identified solely by what is expressed by the act, and not by its underlying motivation or intention.

4. Is the Paternalist Motive Uniquely Insulting?

Motivational accounts do seem to get at the core of why we worry about paternalistic interference: acting because we believe that someone is incapable of choosing well for themselves, or because we do not trust them to make decisions about matters that are legitimately within their control, does seem to involve treating adults as though they were children in a problematic way. However, it has been argued that the 'paternalist motive', however it is identified, is neither so unique, nor so insulting, as it is often assumed. Thus, we should not worry about all instances of paternalism, since the insult that it supposedly represents occurs on other occasions that we do not object to. Rather, we should object only when it restricts particular valuable liberties, or undermines central capacities (de Marneffe 2006; 2013; Shafer-Landau 2005).

First, de Marneffe points out that a motivational approach might undermine the 'project of reconciliation' that many anti-paternalists undertake. Given that various seemingly paternalist policies appear highly plausible, those who wish to maintain an anti-paternalist stance might engage in a process of providing a non-paternalist justification for these policies.²¹ For example, Shiffrin mounts such a defence of the unconscionability doctrine, which allows courts to decline to enforce contracts whose terms are one-sided, exploitative, or unfair. She argues that the contractors do not have a right to unqualified assistance in upholding their contracts, and that the reasons for refusing this assistance might not be grounded in distrust of the abilities of the exploited party, but for the self-regarding reason that the state does not want to be involved in such exploitation (Shiffrin 2000: 221-224). However, if our only concern is motive, then we do not need to provide a good non-paternalist justification: the mere psychological possibility of a non-paternalist motive, however feeble the justification, suffices. Further, providing a good justification fails to demonstrate that the policy is *not* paternalist, if this is not, in fact, what motivates the state (de Marneffe 2006: 71).

Second, and more damagingly, de Marneffe (2006: 78-79) points out that the government commonly substitutes its judgement for that of its citizens, meaning that such distrust and substitution is not limited to cases of paternalism. For example, when imposing speed limits the government substitutes its judgement of what is best for citizens' judgements. If paternalism does not generate a *special* insult, then its supposed uniqueness may not explain why we think

²¹ For example Feinberg (1986: 79-81); Anderson (1999: 329-332).

paternalism matters (Fox unpublished). Further, such substitutions need not imply an insulting belief that individuals are “too stupid to run their own lives” (Anderson 1999: 330). Rather, such judgements might only imply “that his rationality is imperfect, and so is open to the kinds of error we all are” (de Marneffe 2006: 80).

Indeed, there has been a recent proliferation of evidence that we *are* all frequently subject to errors in reasoning.²² This has given rise to the ‘nudge’ literature, which contends that default rules, framing effects and starting points influence our preferences such that it is often unclear what our ‘true’ or authentic preferences are. Indeed, if our choices are determined by the way in which we are presented with the options, then it may be that we do not have true preferences at all in some cases. Many examples are discussed in the literature, including the changes in children’s choice of food depending on where items are placed in a cafeteria, the rate and level of uptake of pension plans depending on the default option, and willingness to undergo an operation depending on whether it is presented as having a 90% survival rate over five years, or a 10% death rate.

Sunstein and Thaler have argued that these influences mean that paternalism is inevitable. If policy cannot *but* shape our choices and preferences, surely they should be shaped in the best interests of those affected. Yet they also insist that such paternalism can be libertarian: though individuals are nudged into doing what is in their best interests, they are not compelled. Thus, their liberty and free choice is not restricted, so libertarians could have no grounds for complaint. Paternalism of this sort may not, in fact, be insulting, then, contrary to motive-based accounts.

There has been a considerable critical response to Sunstein and Thaler’s proposal, and I cannot do justice to all of it here.²³ Some have argued that this libertarian paternalism is not really *libertarian*, insofar as it is manipulative, or

²² For example, Thaler and Sunstein (2008); Le Grand and New (2015); Conly (2013); Kahneman and Tversky (1973; 1974; 2000); Kahneman (2011); Trout (2005); Glaeser (2005). Indeed, Kristoffer Ahlstrom-Vij (2013) has argued that paternalism may be justified for our *epistemic* good: an approach he calls epistemic paternalism.

²³ See, for example, Goodwin (2012); Guala and Mittone (2015); Hanna (2011); Hausman and Welch (2010); Klick and Mitchell (2006); Mills (2013b; 2015); Mitchell (2005); Rachlinski (2003); Rebonato (2012); Salvat (2008); Schnellenbach (2012); Smith and MacPherson (2009); Sugden (2008); Wilkinson (2013). For a partial response, see Sunstein (2014a).

bypasses our capacity for rational agency as opposed to engaging with it.²⁴ Some have argued that it is not really *paternalist* insofar as the nudges may not always be in the interests of those who are subject to them. Individuals are nudged towards what will make the standard individual better-off, but choices that are good for a majority may not be good, and can even be very bad (possibly deadly), for a minority.²⁵

Further, it has been objected that, contrary to Sunstein and Thaler's (2003) initially stated position, paternalist interference is not *inevitable* (Kelly 2013; de Marneffe 2006: 71). Whilst some influences are impossible to avoid, it is not true that this influence need be paternalist. Sunstein and Thaler suggest that the alternatives to paternalism are morally reprehensible, or unfeasible. For example, in the cafeteria case, the alternatives they suggest to choosing an order that will make the children better off include choosing an order to maximise profits or to maximise sales from suppliers willing to offer the largest bribe; or alternatively, choosing an order that will lead to choices that correspond with the children's 'true' preferences (though it is unclear how these could be identified), or a random order (where the consequences may be very bad). Thus, paternalism is presumably seen as 'inevitable' in the sense of being the only permissible use of the ability to influence decisions. Yet surely this is not true. Jamie Kelly (2013: 222-225) has outlined systems of utilitarian and Rawlsian nudges, according to which we influence choices to maximise the general good, or to promote Rawlsian justice. Indeed, Sunstein and Thaler (2003: 1193; 2008: 184-192) allow for non-paternalistic nudges themselves: encouraging people to be organ donors is surely done for the general good, not for the sake of those who are subject to the influence of an opt-out scheme of donor registration.²⁶

Given this, we may allow Sunstein and Thaler the weaker conclusion that the agents setting background conditions – perhaps especially when these agents are states – can *sometimes* permissibly nudge individuals for the sake of their own best interests. Returning to our objections to motive-based accounts of paternalism, then, we may question whether distrust is always insulting if we acknowledge that we are all sometimes mistaken in our reasoning; and further,

²⁴ Le Grand and New (2015: 141-142); Conly (2013: 8); Bovens (2009: 209); Blumenthal-Barby (2013: 188); Mills (2015); Grüne-Yanoff (2012). Sunstein (2014b; 2015b) has argued that requiring choice may still be paternalist.

²⁵ Coons and Weber (2013: 18; 21); Salvat (2008: 8); Blumenthal-Barby (2013: 183).

²⁶ Indeed, Sunstein and Thaler (2003: 1193) seem to acknowledge this point, given that they call this libertarian benevolence. See also Korobkin (2009); Smith and McPherson (2009); Guala and Mittone (2015).

when influence is unavoidable, whether picking the paternalist option is necessarily more insulting than the alternatives.

Finally, de Marneffe (2006: 83) argues that “the purpose of government is to establish justice, and to promote the general welfare, consistent with respecting individuals’ rights and other principles of fairness”. If this is so, then we may wonder whether it is really presumptively impermissible for a government to pursue non-rights violating acts, motivated by the best interest of their citizens. Whilst we may have good reason to be concerned about the infringement of certain core freedoms, de Marneffe (2006: 84) insists that our “status as an independent adult” is not undermined by every violation of liberty: being required to wear our seatbelt, for example.²⁷

Such concerns connect to those raised against soft paternalism, and are indeed a general refrain against any iteration of anti-paternalism: whether paternalism involves interference with voluntary self-harm, or actions motivated by a distrust of agents’ judgements about their interests, surely it is fanaticism to insist that these policies are always *pro tanto* impermissible, regardless of how insignificant the restriction to our liberty or autonomy, or how great the contribution to our interests.²⁸ De Marneffe concludes, then, that we should not always distrust paternalist motives: sometimes paternalist arguments are the strongest ones on offer, in which case we should not demand that they are not used to justify particular actions.

5. Paternalism, Respect, and Exclusionary Reasons

However, we might think that at least some of these criticisms miss the mark. Perhaps the problem with paternalism is not that individuals are insulted by a judgement that they cannot to be trusted to choose well about their welfare or interests, but that we use these judgements to ground actions, where this involves overriding individuals’ wills, and interfering with their actions. David Enoch (2016) has recently argued that the wrong of paternalism does not lie in having a negative belief about others’ future behaviour. Rather, it lies in acting on this belief.

²⁷ In response we may think that having our will ignored by the state – either as an individual, or as a collective – may be importantly disrespectful, even over apparently insignificant matters. The relevant problem with the undemocratic introduction a policy may not be the loss of some valuable opportunity, or ‘core freedom’, but that citizens are not respected as agents capable of choice (Begon 2016: 65-68).

²⁸ Arneson (2005: 264; 278); Sunstein and Thaler (2003: 1165); Archard (unpublished: 10).

Enoch (2016: 10) begins by arguing that if the evidence indicates that the potential subject of our paternalism will “make poor decisions, or fail to follow through on her good ones”, then we do not wrong her or fail to treat her with appropriate respect if we believe, as we should, “what the evidence sufficiently strongly indicates”. The same point can also be made when we do not merely make judgements about their likely behaviour, but distrust their very ability to deliberate and act rationally: “whether or not someone has the relevant abilities is a factual question, and whether or not we should believe this is, arguably, fully determined by the evidence” (Enoch 2016: 12). However, Enoch (2016: 26-30) suggests that we have exclusionary reasons, based on personal autonomy, not to act on this judgement. Thus, the paternalist does not err because they judge that we will make a poor decision, but because they allow this judgement to have weight in their deliberation about how to act.²⁹

The idea of exclusionary reasons is also central to Daniel Groll’s (2012) account. Groll argues that we can act paternalistically even when we act in accordance with someone’s will because it is their will. His central distinction is between treating someone’s will as authoritative, and treating it as part of an all-things-considered assessment of their good. A surrogate, deciding on a course of treatment for a non-competent patient, might find that the fact that the patient does not want a procedure weighs most heavily in their deliberation on whether they should receive it (Groll 2012: 700). It is physically and psychologically distressing to be compelled to undergo a medical procedure; and, even when an individual lacks the competence to make their own medical decisions, part of what it is for their life to go well may be that their will plays a central role in determining the shape of their life (Groll 2012: 704). Such considerations may lead the surrogate to decide against treatment, despite the likely benefits to the patient’s health. In Groll’s terminology, their will is treated as substantially decisive.

However, Groll argues that a competent patient would justifiably resent their doctor merely weighing their will in the balance of reasons in this way, even if they ultimately decide to act in accordance with his will. If we are competent, our will should be authoritative, or structurally decisive: “it is meant to supplant the reason-giving force of other considerations not because it outweighs those other considerations but because it is meant to silence or exclude those other

²⁹ As Tsai (2014: 87-88) notes, having a paternalist motive involves not just having an attitude of distrust and concern, but also endorsing these attitudes, such that they “issue in action”. It is paternalist action that troubles Tsai – actions motivated by, and conveying, this distrust – rather than merely having these attitudes.

considerations” (Groll 2012: 701). On Groll’s account, treating someone’s will as merely substantially decisive is always paternalist. This paternalism is impermissible when we treat someone “whose will is intact” in this way, because we believe they will “fail to exercise a capacity for sound judgement in the situation at hand” (Groll 2012: 718).

Although Groll does not define paternalism in terms of this motive, like Shiffrin and Quong, he sees the *wrong* of paternalism as deriving from the motive and attitude of the paternaliser. However, we do not avoid this wrong by always trusting individuals on matters relating to their own good or legitimate domains of control. We do not need to believe that individuals choose well, but we must nonetheless treat their will as authoritative because it is their will.

6. Conclusion

We may remain unconvinced by these arguments. If paternalist interference can be good for us by our own lights – indeed, may improve our autonomy – then why insist that it is always *prima facie* impermissible? Alternatively, we might be persuaded that respecting individuals requires treating their will as decisive – as an exclusionary reason in the domains over which they have legitimate control – regardless of how much good could be done by ignoring it. There may no longer be a straightforward debate between paternalism and anti-paternalism, but this is not to say there is anything like a consensus on how we should understand paternalism, nor what our attitude to paternalistic acts and agents should be.³⁰

Nuffield College, University of Oxford
Nuffield College, New Road, Oxford, OX1 1NF
jessica.begon@nuffield.ox.ac.uk

References

- Ahlstrom-Vij, K. 2013. *Epistemic Paternalism: A Defence*. Basingstoke: Palgrave Macmillan.
- Anderson, E. 1999. What is the Point of Equality? *Ethics* 109: 287-237.
- Archard, D. 2013. Review: Against Autonomy: Justifying Coercive Paternalism by Sarah Conly. *Journal of Applied Philosophy* 2013: 397-400.
- Archard, D. 1990. Paternalism Defined. *Analysis* 50: 36-42.

³⁰ For valuable comments, suggestions, and discussion on earlier drafts of this paper I would like to thank Carl Fox, Chris Mills, Joe Saunders, and Daniel Viehoff.

- Archard, D. Unpublished. What is Wrong with Paternalism?
- Arneson, R.J. 1980. Mill versus Paternalism. *Ethics* 90: 470-489.
- Arneson, R.J. 2005. Joel Feinberg and the Justification of Hard Paternalism. *Legal Theory* 11: 259-284.
- Arneson, R.J. 2013. Paternalism and the principle of fairness. In *Paternalism: Theory and Practice*, ed. C. Coons and M. Weber, 134-156. Cambridge: Cambridge University Press.
- Begon, J. 2013. *Policy without Paternalism: A Capability Approach to Legitimate State Action*. Thesis submitted for PhD, University of Sheffield. Available at: <http://etheses.whiterose.ac.uk/id/eprint/5412>
- Begon, J. 2015. What are Adaptive Preferences? Exclusion and Disability in the Capability Approach. *Journal of Applied Philosophy* 32: 241-257.
- Begon, J. 2016. Athletic Policy, Passive Well-Being: Defending Freedom in the Capability Approach. *Economics and Philosophy* 32: 51-73.
- Ben-Porath, S. 2013. Paternalism, (school) choice, and opportunity. In *Paternalism: Theory and Practice*, ed. C. Coons and M. Weber, 247-265. Cambridge: Cambridge University Press.
- Blumenthal, J.A. 2013. A psychological defence of paternalism. In *Paternalism: Theory and Practice*, ed. C. Coons and M. Weber, 197-215. Cambridge: Cambridge University Press.
- Blumenthal-Barby, J.S. 2013. Choice Architecture: A mechanism for improving decisions while preserving liberty? In *Paternalism: Theory and Practice*, ed. C. Coons and M. Weber, 178-196. Cambridge: Cambridge University Press.
- Bovens, L. 2009. The Ethics of Nudge. In *Preference Change: Approaches from Philosophy, Economics and Psychology*, ed. T. Grüne-Yanoff and S.O. Hansson, 207-219. Berlin: Springer.
- Cholbi, M. 2013. Kantian paternalism and suicide intervention. In *Paternalism: Theory and Practice*, ed. C. Coons and M. Weber, 115-133. Cambridge: Cambridge University Press.
- Colburn, B. 2008. Debate: The Concept of Voluntariness. *The Journal of Political Philosophy* 16: 101-111.
- Conly, S. 2013. *Against Autonomy: Justifying Coercive Paternalism*. Cambridge: Cambridge University Press.
- Conly, S. 2016. Review: *Government Paternalism: Nanny State or Helpful Friend?* Julian Le Grand and Bill New. *Economics and Philosophy* 32: 156-162.

- Coons, C. and M. Weber. 2013. Introduction: Paternalism – Issues and trends. In *Paternalism: Theory and Practice*, ed. C. Coons and M. Weber, 1-24. Cambridge: Cambridge University Press.
- Cornell, N. 2015. A Third Theory of Paternalism. *Michigan Law Review* 113: 1295-1336.
- De Marneffe, P. 2006. Avoiding Paternalism. *Philosophy & Public Affairs* 34: 68-94.
- De Marneffe, P. 2013. Self-sovereignty and paternalism. In *Paternalism: Theory and Practice*, ed. C. Coons and M. Weber, 56-73. Cambridge: Cambridge University Press.
- Dworkin, G. 1997. Paternalism. In *Mill's On Liberty: Critical Essays*, ed. G. Dworkin, 61-82. Oxford: Rowman and Littlefield.
- Dworkin, G. 2005. Moral Paternalism. *Law and Philosophy* 24: 305-319.
- Dworkin, G. 2013. Defining Paternalism. In *Paternalism: Theory and Practice*, ed. C. Coons and M. Weber, 25-38. Cambridge: Cambridge University Press.
- Dworkin, G. 2014. Paternalism. In *The Stanford Encyclopedia of Philosophy*, ed. E.N. Zalta. Available at: <http://plato.stanford.edu/archives/sum2014/entries/paternalism/>
- Enoch, D. 2016. What's Wrong with Paternalism: Autonomy, Belief and Action. *Proceedings of the Aristotelian Society CXVI*. Draft available at: https://issuu.com/aristotelian.society/docs/enoch_220a2e53144ba9/1?e=3458282/30639094
- Fateh-Moghadam, B. and T. Gutmann. 2014. Governing [through] Autonomy. The Moral and Legal Limits of “Soft Paternalism”. *Ethical Theory and Moral Practice* 17: 383-397.
- Feinberg, J. 1986. *The Moral Limits of the Criminal Law. Volume Three: Harm to Self*. New York: Oxford University Press.
- Fowler, T. 2014. Perfectionism for children, anti-perfectionism for adults. *Canadian Journal of Philosophy* 44: 305-323.
- Fox, C. unpublished. What's Special About the Insult of Paternalism?
- Glaeser, E.L. 2005. Paternalism and Psychology. *The University of Chicago Law Review* 73: 133-156.
- Goodwin, T. 2012. Why We Should Reject ‘Nudge’. *Politics* 32: 85-92.
- Grill, K. 2007. The Normative Core of Paternalism. *Res Publica* 13: 441-458.

- Groll, D. 2012. Paternalism, Respect, and the Will. *Ethics* 122: 692-720.
- Grüne-Yanoff, T. 2012. Old wine in new casks: libertarian paternalism still violates liberal principles. *Social Choice and Welfare* 38: 635-645.
- Guala, F. and L. Mittone. 2015. A Political Justification of Nudging. *Review of Philosophy and Psychology* 6: 385-395.
- Hanna, J. 2011. Consent and the Problem of Framing Effects. *Ethical Theory and Moral Practice* 14: 517-531.
- Hausman, D.M. and B. Welch. 2010. Debate: To Nudge or Not to Nudge. *The Journal of Political Philosophy* 18: 123-136.
- Haybron, D.M. and A. Alexandrova. 2013. Paternalism in economics. In *Paternalism: Theory and Practice*, ed. C. Coons and M. Weber, 157-177. Cambridge: Cambridge University Press.
- Husak, D. 2013. Penal Paternalism. In *Paternalism: Theory and Practice*, ed. C. Coons and M. Weber, 39-55. Cambridge: Cambridge University Press.
- Kahneman, D. 2011. *Thinking Fast and Slow*. London: Allen Lane.
- Kahneman, D. and A. Tversky. 1973. On the Psychology of Prediction. *Psychological Review* 80: 237-251.
- Kahneman, D. and A. Tversky (eds.). 2000. *Choices, Values, and Frames*. Cambridge: Cambridge University Press.
- Kant, I. 2011. *Groundwork of the Metaphysics of Morals*, trans. M. Gregor and J. Timmermann. Cambridge: Cambridge University Press.
- Kant, I. 1996. *Metaphysics of Morals*, trans. M. Gregor. Cambridge: Cambridge University Press.
- Kelly, J. 2013. Libertarian paternalism, utilitarianism, and justice. In *Paternalism: Theory and Practice*, ed. C. Coons and M. Weber, 216-230. Cambridge: Cambridge University Press.
- Klick, J. and G. Mitchell. 2006. Government Regulation of Irrationality: Moral and Cognitive Hazards. *Minnesota Law Review* 90: 1620-1663.
- Korobkin, R.B. 2009. Libertarian Welfarism. *California Law Review* 97: 1651-1685.
- Le Grand, J. and B. New. 2015 *Government Paternalism: Nanny State or Helpful Friend?* Princeton: Princeton University Press.
- Mill, J.S. 1974. *On Liberty*, ed. G. Himmelfarb. London: Penguin.
- Mills, C. 2013a. The Problem of Paternal Motives. *Utilitas* 25: 446-462.

- Mills, C. 2013b. Why Nudges Matter: A Reply to Goodwin. *Politics* 33: 28-36.
- Mills, C. 2015. The Heteronomy of Choice Architecture. *Review of Philosophy and Psychology* 6: 495-509.
- Mitchell, G. 2005. Libertarian Paternalism Is an Oxymoron. *Northwestern University Law Review* 99: 1245-1277.
- Mullin, A. 2014. Children, Paternalism and the Development of Autonomy. *Ethical Theory and Moral Practice* 17: 413-426.
- Nozick, R. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.
- Olsaretti, S. 1998. Freedom, Force, and Choice: Against a Rights-Based Definition of Voluntariness. *The Journal of Political Philosophy* 6: 53-78.
- Quong, J. 2011. *Liberalism without Perfection*. Oxford: Oxford University Press.
- Rachlinski, J.J. 2003. The Uncertain Psychological Case for Paternalism. *Northwestern University Law Review* 97: 1165-1125.
- Rebonato, R. *Taking Liberties: A Critical Examination of Libertarian Paternalism*. Basingstoke: Palgrave Macmillan.
- Salvat, C. 2008. Is Libertarian Paternalism an Oxymoron? A Comment on Sunstein and Thaler. *HAL Working Paper Series*. Available at: http://www.greqam.fr/sites/default/files/_dt/greqam/2008-42.pdf
- Schnellenbach, J. 2012. Nudges and norms: On the political economy of soft paternalism. *European Journal of Political Economy* 28: 266-277.
- Scoccia, D. 2008. In Defence of Hard Paternalism. *Law and Philosophy* 27: 351-381.
- Scoccia, D. 2013. The right to autonomy and the justification of hard paternalism. In *Paternalism: Theory and Practice*, ed. C. Coons and M. Weber, 74-92. Cambridge: Cambridge University Press.
- Sjöstrand, M., S. Eriksson, N. Juth, and G. Helgesson. 2013. Paternalism in the Name of Autonomy. *Journal of Medicine and Philosophy* 38: 710-724.
- Smith, M.A. and M.S. McPherson. 2009. Nudging for Equality: Values in Libertarian Paternalism. *Administrative Law Review* 61: 323-342.
- Shafer-Landau, R. 2005. Liberalism and Paternalism. *Legal Theory* 11: 169-191.
- Shiffrin, S.V. 2000. Paternalism, Unconscionability Doctrine, and Accommodation. *Philosophy & Public Affairs* 29: 205-250.
- Sugden, R. 2008. Why incoherent preferences do not justify paternalism. *Constitutional Political Economy* 19: 226-248.

- Sunstein, C.R. 2014a. *Why Nudge? The Politics of Libertarian Paternalism*. New Haven: Yale University Press.
- Sunstein, C.R. 2014b. Requiring Active Choosing is a Form of Paternalism. *Harvard John M. Olin Discussion Paper Series 782*. Available at: <http://nrs.harvard.edu/urn-3:HUL.InstRepos:12921733>
- Sunstein, C.R. 2015a. Do People Like Nudges? Draft available on the *Social Sciences Research Network* at: <http://ssrn.com/abstract=2604084>
- Sunstein, C.R. 2015b. Behaviorally Informed Health Policy? Patient Autonomy, Active Choosing, and Paternalism. *Nudging Health: Health Law and Behavioral Economics*. Available at: <http://nrs.harvard.edu/urn-3:HUL.InstRepos:16162450>
- Sunstein, C.R. and R.H. Thaler. 2003. Libertarian Paternalism is Not an Oxymoron. *The University of Chicago Law Review* 70: 1159-1202.
- Tadros, V. 2011. Consent to Harm. *Current Legal Problems* 64: 23-49.
- Thaler, R.H. and C.R. Sunstein. 2008. *Nudge: Improving decisions about health, wealth and happiness*. London: Penguin.
- Trout, J.D. 2005. Paternalism and Cognitive Bias. *Law and Philosophy* 24: 393-434.
- Tsai, G. 2014. Rational Persuasion as Paternalism. *Philosophy & Public Affairs* 42: 78-112.
- Tversky, A. and D. Kahneman. 1974. Judgement under Uncertainty. Heuristics and Biases. *Science* 185: 1124-1131.
- Velleman, D.J. 1999. A Right to Self-Termination? *Ethics* 109: 606-628.
- Wall, S. 2003. The Structure of Perfectionist Toleration. In *Perfectionism and Neutrality: Essays in Liberal Theory*, ed. S. Wall and G. Klosko, 231-256. Oxford: Rowman and Littlefield.
- Wall, S. 2009. Self-Ownership and Paternalism. *The Journal of Political Philosophy* 17: 399-417.
- Wall, S. 2013. Moral environmentalism. In *Paternalism: Theory and Practice*, ed. C. Coons and M. Weber, 93-114. Cambridge: Cambridge University Press.
- Wilkinson, T.M. 2013. Nudging and Manipulation. *Political Studies* 61: 341-355.