



How Wrong is Paternalism?

David Birks

Department of Politics and International Relations, University of Oxford

david.birks@politics.ox.ac.uk

Abstract

In this paper, I argue against the commonly held view that paternalism is *all things considered* wrong when it interferes with a person's autonomy. I begin by noting that the plausibility of this view rests on the assumption that there is a morally relevant difference in the normative reasons concerning an intervention in a person's self-regarding actions and an intervention in his other-regarding actions. I demonstrate that this assumption cannot be grounded by wellbeing reasons, and that autonomy-based reasons of non-interference also cannot adequately explain the difference. Following this, I propose that the difference in the reasons related to an intervention in a person's self-regarding actions and an intervention in his other-regarding actions can be explained by the value of holding a person responsible for his choices. Nonetheless, this does not result in paternalistic behavior that interferes with autonomy being *all things considered* wrong. Instead, I show that the reason to hold a person responsible for a diminution of his wellbeing does not necessarily defeat the wellbeing reasons that count in favor of paternalistic behavior.

Keywords

paternalism – autonomy – wellbeing – responsibility-catering prioritarianism

Paternalism is often thought to be wrong. According to the dominant account of the wrongness of paternalism, paternalism is wrong when it interferes with a person's autonomy. For instance, suppose that I secretly switch your soft drink for a diet soft drink because sugar is bad for your health. Or imagine that I hypnotize you to pursue a career in philosophy rather than investment banking because you would have a better life as a philosopher. On this view, these

paternalistic actions are wrong because they interfere with your autonomous decisions, even if they would make your life go better.¹

This paper focuses on the question, how wrong is paternalism? More precisely, it examines whether paternalistic behavior that interferes with a person's autonomy is *all things considered* wrong, and so impermissible, or only *pro tanto* wrong, and so possibly permissible, or even sometimes obligatory.² This is an important distinction. After all, it is a commonly held liberal view that if a person is making a self-regarding autonomous choice, it is always *all things considered* wrong to interfere paternalistically with this choice, irrespective of the benefits of the interference to the person's life. This view is often attributed to John Stuart Mill,³ and has been defended by many philosophers, such as Richard Arneson, Michael Cholbi, John Hodson, Joel Feinberg, Jeffrie Murphy, and Donald VanDeVeer.⁴

-
- 1 I am referring to the view of autonomy that is frequently described as personal autonomy, understood as an agent "determining his conduct by his own most highly cherished values," Stephen Darwall, 'The Value of Autonomy and Autonomy of the Will', *Ethics* 116 (2006), pp. 263–84. p. 265. This contrasts with the notion of moral autonomy, the view (broadly speaking) that a person is morally autonomous if his principles of morality originate from him.
 - 2 There are other accounts of the wrongness of paternalism not grounded in autonomy. For example, Jonathan Quong has developed an account that grounds the wrongness of paternalism in its demeaning of moral status. See: Jonathan Quong, *Liberalism without Perfection* (Oxford: Oxford University Press, 2011), p. 98. I am unable to address these other accounts in this paper. However, given its prominence, and the number of its influential proponents, my focus on the autonomy view is justified. For a good overview of the different accounts of the wrongness of paternalism see: Nicolas Cornell, 'A Third Theory of Paternalism', *Michigan Law Review* 113 (2015), pp. 1295–336.
 - 3 John Stuart Mill, *On Liberty* (London: Penguin Classics, 1985). For this interpretation of Mill, see: H.L.A. Hart, 'Natural Rights: Bentham and John Stuart Mill', in H.L.A. Hart, *Essays on Bentham: Jurisprudence and Political Philosophy* (Oxford: Oxford University Press, 1982), pp. 79–105, pp. 101–3.
 - 4 Richard Arneson, 'Mill Vs. Paternalism' *Ethics* 90 (1980), pp. 470–80. Michael Cholbi, 'Kantian Paternalism and Suicide Intervention', in C. Coons and M. Weber (eds.), *Paternalism: Theory and Practice* (Cambridge: Cambridge University Press, 2013), pp. 115–33. John D. Hodson, 'The Principle of Paternalism', *American Philosophical Quarterly* 14 (1977), pp. 61–9. Joel Feinberg, *The Moral Limits of the Criminal Law: Harm to Self* (Oxford: Oxford University Press, 1986). Jeffrie Murphy, 'Incompetence and Paternalism', in J. Murphy, *Retribution, Justice, and Therapy: Essays in the Philosophy of Law* (London: Pallas Paperbacks, 1979), pp. 165–182. Donald VanDeVeer, *Paternalistic Intervention: The Moral Bounds of Benevolence* (Princeton, New Jersey: Princeton University Press, 1986). For some of the practical implications of this view, see: Tim Bayne, and Neil Levy, 'Amputees by Choice: Body Integrity Identity Disorder and the Ethics of Amputation', *Journal of Applied Philosophy* 22 (2005), pp. 75–86, pp. 79–80. T.M. Wilkinson, *Ethics and the Acquisition of Organs* (Oxford: Oxford University Press, 2011),

One fruitful way of analyzing the wrongness of paternalism is to view the competing considerations as normative reasons. When we consider how we should act, we are considering the normative reasons for and against these actions.⁵ Consequently, the autonomy view of the wrongness of paternalism could be explained by what I call an autonomy-based reason of non-interference, or more succinctly, an ABRONI. An ABRONI is a normative reason not to interfere with a person's autonomous decisions and actions.⁶ Let us assume that in the cases mentioned above, there is an ABRONI not to hypnotize you to change your career, and not to swap secretly your soft drink for a diet soft drink. If a person fails to act in accordance with an ABRONI, and interferes with another person's autonomous decisions or actions, he is doing something that is at least *pro tanto* wrong.

Let us also assume that a person acts wrongly *all things considered* if there is a reason (or set of reasons) not to perform the action that defeats any reason(s) that count in favor of performing the action. So, in order for it to be *all things considered* wrong to act contrary to the ABRONI, the ABRONI must defeat all of the reasons that count in favor of acting contrary to the ABRONI, or the ABRONI must be among the set of reasons that collectively defeat all of the reasons that count in favor of acting contrary to it. One likely candidate for a reason to act contrary to the ABRONI is a wellbeing reason. For instance, suppose that your life will go far better if you become a philosopher rather than an investment banker. If you were a philosopher, you would write a number of important papers and you would have a very happy life, whereas given your temperament, if you were an investment banker, you would suffer from stress and anxiety, and your life would be full of misery. Given that your life would

p. 20. L.W. Sumner, *Assisted Death: A Study in Ethics and Law* (Oxford: Oxford University Press, 2011), Chapter 2.

5 See: Roger Crisp, *Reasons and the Good* (Oxford: Oxford University Press, 2006), p. 8, pp. 36–61. For an alternative view, see: John Broome, 'Reasons', in R.J. Wallace, P. Pettit, S. Schefler and M. Smith (eds.), *Reason and Value: Themes from the Moral Philosophy of Joseph Raz* (Oxford: Clarendon Press, 2006), pp. 28–55.

6 To be clear, an ABRONI is a reason to *respect* that person's autonomy, rather than a reason to *promote* that person's autonomy. For example, we do not act in accordance with an ABRONI when we intervene to prevent a person from autonomously committing suicide, even if the intervention promotes the paternalizee's autonomy by extending his life. Instead, the anti-paternalist view considered in this paper is concerned with autonomy generating a reason not to interfere with that person's autonomous decisions and actions, even when it results in an overall reduction in the autonomy in that person's life. For a discussion of this distinction see: Thomas Scanlon, *What We Owe to Each Other* (London: The Belknap Press of Harvard University Press, 2000), pp. 78–107.

go better as a philosopher than as an investment banker, there is a wellbeing reason to hypnotize you to lead you to change your career to philosophy.⁷

In light of this conflict between the wellbeing reason and the ABRONI, one could hold that it is *all things considered* wrong to act paternalistically by claiming that an ABRONI is a *wellbeing-defeating* reason. That is, it is a *type* of reason that always defeats reasons of wellbeing, regardless of their weight.⁸ Once we hold this view, it follows that even if hypnotism resulted in you having a life that goes significantly better, if the hypnotism is contrary to your ABRONI, the wellbeing reason to hypnotize you would be defeated by the ABRONI, and so, *ceteris paribus*, hypnotizing you would be *all things considered* wrong.

The view that an ABRONI is a wellbeing-defeating reason can be summarized as the following position that I will call *Strong Anti-Paternalism*:

Strong Anti-Paternalism: An ABRONI with respect to some person P always defeats P's wellbeing reasons that count in favor of intervention, regardless of the weight of the wellbeing reasons that count in favor of the intervention.

Although not expressed in these terms, this is the position held by many defenders of strongly anti-paternalistic views, such as Arneson, Cholbi, Feinberg, Mill, and Murphy.⁹ I propose that *Strong Anti-Paternalism* should be rejected in favor of the following position:

7 This assumes that it is possible for a person's wellbeing and his autonomous decisions to diverge. For a defense of this view see: George Sher, *Beyond Neutrality* (Cambridge: Cambridge University Press, 1997), pp. 57–67. Richard Arneson, 'Human Flourishing Verses Desire Satisfaction', *Social Philosophy and Policy* 16 (2009), pp. 113–42, pp. 135–41. Steven Wall, *Liberalism, Perfectionism and Restraint* (Cambridge: Cambridge University Press, 1998), pp. 189–197. T.M. Wilkinson, 'Against Dworkin's Endorsement Constraint', *Utilitas* 15 (2003), pp. 175–93.

8 To clarify, a wellbeing-defeating reason does not necessarily defeat non-wellbeing reasons. It may be the case that reasons generated by values other than wellbeing defeat a wellbeing-defeating reason. Moreover, when a reason is defeated, the defeated reason is not eliminated, but still has weight. Ideally, we would conform with all reasons, including the wellbeing reason, and the wellbeing-defeating reason. But given the conflict between these reasons, we cannot act in accordance with both, and so *ceteris paribus* we should act in accordance with the wellbeing-defeating reason.

9 These authors argue that a necessary condition of morally permissible paternalism is that it is in accordance with the paternalizee's autonomy. Paternalism contrary to the paternalizee's autonomy is impermissible, regardless of the benefit of the paternalism to the paternalizee. For example, Cholbi writes "interference with an individual's liberty is permissible only if, by interfering, we stand a reasonable chance of preventing that agent from performing actions

Moderate Anti-Paternalism: (1) An ABRONI with respect to some person P will defeat P's wellbeing reasons that count in favor of intervention if and only if the wellbeing reasons are insufficiently weighty to outweigh the ABRONI; and (2) P's wellbeing reasons are sometimes sufficiently weighty to outweigh P's ABRONI.

If we accept *Moderate Anti-Paternalism* rather than *Strong Anti-Paternalism*, it follows that although there is an ABRONI not to behave paternalistically, it is not a wellbeing-defeating reason, and so it could be *all things considered* permissible, or even sometimes obligatory to behave paternalistically. I will not discuss specific cases in which the ABRONI is defeated by wellbeing reasons, but simply aim to show that ABRONIS do not always defeat the wellbeing reasons that count in favor of the intervention, and as a result, *Strong Anti-Paternalism* is mistaken.

My argument proceeds as follows. I begin by noting that the plausibility of *Strong Anti-Paternalism* rests on the assumption that there is a morally relevant difference in the reasons concerning an intervention in a person's self-regarding actions and an intervention in his other-regarding actions.¹⁰ I show that this assumption cannot be grounded by a relevant difference in wellbeing reasons, and that the ABRONI itself also cannot adequately explain the distinction without resulting in implausible implications. I argue that that *Moderate Anti-Paternalism* does not suffer from this weakness, and consequently, we should reject *Strong Anti-Paternalism*. Following this, I propose that the

she chose due to distorted reasoning and which would result in that agent's rationally chosen ends not being as fully realized as they would have been had she so acted," Cholbi, 'Kantian Paternalism and Suicide Intervention', p. 118. Joel Feinberg writes "the state has a right to prevent self-regarding harmful conduct *only* when it is substantially nonvoluntary or when temporary intervention is necessary to establish whether it is voluntary or not," Joel Feinberg, 'Legal Paternalism', *Canadian Journal of Philosophy* 1 (1971), pp. 105–24. p. 113. Hodson writes "Paternalistic interventions are justified if and only if (i) there is good evidence that the decisions with respect to which the person is to be coerced are encumbered, and (ii) there is good evidence that this person's decisions would be supportive of the paternalistic intervention if they were not encumbered." Hodson, 'The Principle of Paternalism', p. 65. Murphy writes "it is plausible to regard incompetence as a necessary condition for justified paternalistic coercion because I find it plausible the general Millian position on the importance of human liberty". Murphy, 'Incompetence and Paternalism', p. 178 n2. I should be clear that in a later paper, Arneson changed his view but his earlier paper remains influential. See Richard Arneson, 'Joel Feinberg and the Justification of Hard Paternalism', *Legal Theory*, 11 (2005), pp. 259–84.

¹⁰ From now on, I refer to a morally relevant difference as a relevant difference, for short.

relevant difference between an intervention in a person's self-regarding actions and an intervention in his other-regarding actions can be explained by the value of holding a person responsible for his choices. Nonetheless, this does not result in paternalistic behavior being *all things considered* wrong. Instead, I show that the reason to hold a person responsible for the diminution of his wellbeing does not necessarily defeat any wellbeing reason that counts in favor of paternalistic behavior.¹¹

Before I proceed to the argument of the paper, I will make a few clarificatory points. So far, I have mentioned the wrongness of paternalistic behavior without providing a definition of paternalism. For the purposes of this paper, I hold that an agent behaves paternalistically when he intends to benefit another person while being motivated by a negative judgment about the mental abilities (including the volitional capacities) of the beneficiary in some respect.¹² However, my argument challenges all autonomy-based accounts of the wrongness of paternalism, regardless of how they define paternalism.¹³

My second clarification concerns the autonomy view of the wrongness of paternalism. On this view, paternalism is wrong because it is contrary to the paternalizee's autonomy. However, there is not just one autonomy view but rather, there are two main accounts of when paternalistic behavior is contrary to an ABRONI.¹⁴ According to one account, a necessary and sufficient condition for paternalistic behavior to be contrary to an ABRONI is that the paternalizee would desire for the behavior not to take place when rational and without faulty beliefs.¹⁵ According to another view, a necessary and sufficient condition

11 For different grounds to reject strongly anti-paternalistic views, with varying degrees of success, see: Sarah Conly, *Against Autonomy: Justifying Coercive Paternalism* (Cambridge: Cambridge University Press, 2013), Arneson, 'Joel Feinberg and the Justification of Hard Paternalism', Peter De Marneffe, 'Avoiding Paternalism', *Philosophy & Public Affairs* 34 (2006), pp. 68–94. Jason Hanna, 'Paternalism and Impairment', *Social Theory and Practice* 37 (2011), pp. 434–60. Danny Scoccia, 'In Defense of Hard Paternalism', *Law and Philosophy* 27 (2008), pp. 351–81.

12 This definition is adapted from Quong, *Liberalism without Perfection*, p. 80.

13 For other prominent definitions of paternalism see: Allen E. Buchanan, 'Medical Paternalism', in R. Sartorius, (ed.), *Paternalism* (Minneapolis: University of Minnesota Press, 1983). Seana Valentine Shiffrin, 'Paternalism, Unconscionability Doctrine, and Accommodation', *Philosophy and Public Affairs* 29 (2000), pp. 205–50, Bernard Gert, and Charles M. Culver, 'Paternalistic Behavior', *Philosophy and Public Affairs* 6 (1976), pp. 45–57.

14 For an excellent discussion of the difference between these two views see: Danny Scoccia, 'Paternalism and Respect for Autonomy', *Ethics* 100 (1990), pp. 318–34.

15 This autonomy view of the wrongness of paternalism is held by Gerald Dworkin, 'Paternalism' *The Monist* 56 (1972), pp. 64–84, p. 29. Note that Dworkin includes lack of information under the requirement of rationality. See also: Donald VanDeVeer, *Paternalistic*

for a paternalistic behavior to be contrary to an ABRONI is that the paternal-izee's self-regarding conduct is sufficiently voluntary.¹⁶

I do not need to take a stand on which of these views is correct here because there is a considerable overlap on both these views when behavior is contrary to an ABRONI. For example, it would be contrary to an ABRONI to hypnotize you to change your career either because you would desire for the hypnotism not to take place if you were rational and without faulty beliefs, or that your career choice is sufficiently voluntary. When I refer to actions contrary to an ABRONI, I am referring only to actions that would be contrary to an ABRONI on both of these views.

My third clarification concerns the formulation of Strong Anti-Paternalism and Moderate Anti-Paternalism. As presented, they do not refer to the motives or intentions of the agent performing the intervention. Yet, by definition, in order for an intervention to be paternalistic, the agent has to intend to benefit the object of the intervention while being motivated by a negative judgment about the beneficiary's mental abilities. To be clear, on the autonomy view of the wrongness of paternalism, the intentions and motives of the intervening agent are irrelevant to the wrongness of the intervention.¹⁷ While certain motives and intentions of the intervening agent are necessary for an intervention to be labeled as paternalistic, they do not determine whether an ABRONI defeats the wellbeing reasons that count in favor of the intervention.

Finally, I should clarify that some Strong Anti-Paternalists state their view in the language of rights, specifically, that each person has a right of autonomy.¹⁸ My framework of ABRONIS fits well with this rights-based approach. We do not lose anything from our discussion if we understand the ABRONI as one way of expressing the rights-based corresponding duty not to interfere with the autonomy of the right-holder.¹⁹

Intervention: The Moral Bounds of Benevolence (Princeton, New Jersey: Princeton University Press, 1986), pp. 87–9. John Rawls, *A Theory of Justice*, Revised ed. (Cambridge, MA: The Belknap Press of Harvard University Press, 1999), pp. 218–20.

16 Feinberg, *The Moral Limits of the Criminal Law*, pp. 98–171. Arneson, 'Mill Vs. Paternalism'.

17 For a discussion of motive-based accounts of the wrongness of paternalism distinct from autonomy-based views, see: David Birks, 'Moral Status and the Wrongness of Paternalism', *Social Theory and Practice* 40 (2014), pp. 483–98. Chris Mills, 'The Problem of Paternal Motives' *Utilitas* 25 (2013) pp. 446–62.

18 See for instance: Feinberg, *The Moral Limits of the Criminal Law*, pp. 52–94. Also, see the discussion in VanDeVeer, *Paternalistic Intervention*, pp. 58–63.

19 In the literature on reasons, a duty is expressed as a mandatory categorical reason. As John Gardner and Timothy Macklem explain: "To have a duty to do something is to have a reason to do it that, (i) does not depend for its existence on one's goals at the time, and

1 Self-regarding and Other-regarding Interventions

The plausibility of *Strong Anti-Paternalism* rests on the assumption that there is a relevant difference in the reasons concerning the following two types of intervention:

- (1) An intervention in a person's self-regarding actions
- (2) An intervention in a person's other-regarding actions

An example of Intervention (1) would be a standard case of paternalism in which an intervention prevents the paternalizee from making a decision to kill himself, whereas an example of Intervention (2) is an intervention that prevents a murderer from killing another person.²⁰

Typically, the literature defending *Strong Anti-Paternalism* focuses on the *all things considered* wrongness of Intervention (1), and ignores the arguments in support of Intervention (2). This is why Strong Anti-Paternalism is formulated in terms of whether P's ABRONI always defeats P's wellbeing reasons that count in favor of an intervention. This focus is understandable, given that the argument takes place in a discussion of paternalism, and as noted above, a necessary condition of paternalistic behavior is that it must aim to benefit the paternalizee, rather than benefit someone else.

However, I will show that the arguments that would prohibit interventions in (1) also would prohibit many interventions in (2). This reveals the implausibility of *Strong Anti-Paternalism*, and as a consequence, we should reject it in favor of *Moderate Anti-Paternalism*.

Before I set out my argument, it is crucial to emphasize the importance of the distinction between Interventions (1) and (2) for the plausibility of *Strong Anti-Paternalism*. If there is no relevant difference between the two interventions then an absolute prohibition of one entails an absolute prohibition of the

(ii) is also a reason not to act for certain conflicting reasons." John Gardner and Timothy Macklem, 'Reasons', in J.L. Coleman, S. Shapiro, and K.E. Himma (eds.), *The Oxford Handbook of Jurisprudence and Philosophy of Law*, (Oxford: Oxford University Press, 2002), pp. 440–475, p. 465.

20 There are a number of cases that are located on the borderline between these two types of intervention. For example, if A commits suicide, this is likely to have a devastating effect on the wellbeing of A's family. Nonetheless, I assume that this distinction can be made, and I will focus on the clear cases. For a critical discussion of this distinction see: James Fitzjames Stephen, *Liberty, Equality, Fraternity* (Indianapolis: Liberty Fund, 1993), pp. 86–96, 230–2. John Kleinig, *Paternalism* (Manchester: University of Manchester Press, 1984), pp. 32–7.

other. It is clear that the plausibility of *Strong Anti-Paternalism* requires this distinction. After all, without it, the Strong Anti-Paternalist would be unable to prohibit Intervention (1) without also being committed to prohibiting (2). Yet prohibiting (2) would have implausible consequences. For example, it would mean that it would be *all things considered* wrong to act contrary to the ABRONI in order to prevent a murderer killing an innocent person for his amusement. I will assume that this implication is sufficiently counterintuitive to warrant rejecting the argument.²¹ Thus, if *Strong Anti-Paternalism* results in prohibiting (2) in addition to prohibiting (1) then it should be rejected. I will refer to the idea that there is a relevant difference between Interventions (1) and (2) as the *Asymmetry Claim*.

Asymmetry Claim – There is a relevant difference between intervening in a person's self-regarding actions and intervening in a person's other-regarding actions.

In the following sections I will consider two possible explanations for the *Asymmetry Claim*. First, I examine the claim that there is a relevant difference between harming oneself and harming another person. Second, I consider whether the relevant difference can be explained in terms of ABRONIS in the two types of interventions. Neither of these explanations is adequate to ground the *Asymmetry Claim*. As a result, I propose that we should reject *Strong Anti-Paternalism* in favor of *Moderate Anti-Paternalism*.

2 Self-regarding and Other-regarding Reasons of Wellbeing

One traditional explanation for the *Asymmetry Claim* is the thought that there is a relevant difference between an intervention to prevent a person harming himself and an intervention to prevent a person harming another person. Assume that harm should be understood as a diminution of actual or prospective wellbeing. Once we draw this distinction, we could claim that it is (at least) permissible to intervene in order to prevent a person harming another person, while maintaining that it is *all things considered* wrong to act contrary to the ABRONI in order to prevent a person harming himself. This view is often attributed to John Stuart Mill, and his Harm Principle.²² Mill wrote:

²¹ Gerald Dworkin notes that it would only be some anarchists or extreme pacifists who would question this view, 'Paternalism', p. 64.

²² For the clearest defense of this interpretation of Mill and the Harm Principle, see: Arneson, 'Mill Vs. Paternalism', p. 471. See also: Dworkin, 'Paternalism', pp. 74–6. Feinberg 'Legal

[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.²³

One possible explanation of the Harm Principle can be seen if we draw a distinction between two types of intervention:

- (1) It is always *all things considered* wrong for A to act contrary to the ABRONI in relation to B in order to increase (or prevent a diminution of) B's wellbeing.
- (2) It is not always *pro tanto* or *all things considered* wrong for A to act contrary to the ABRONI in relation to B in order to prevent B from diminishing C's wellbeing.

This distinction allows the Strong Anti-Paternalist to claim that the ABRONI always defeats the reasons of wellbeing in (1), whereas the ABRONI may not defeat the reasons of wellbeing in (2). Nevertheless, this distinction does not work. If the reasons of wellbeing are always defeated by the ABRONI in Intervention (1), it follows that unless there is a relevant difference between the wellbeing reasons in the two cases, the wellbeing reasons also should always be defeated by the ABRONI in Intervention (2). However, there is no relevant difference between the wellbeing reasons in the two cases. Therefore, if there are grounds for the *Asymmetry Claim*, it cannot be due to a relevant difference in the wellbeing reasons in (1) and (2).

Let me explain. Imagine an example of Intervention (1), where A intervenes with B in order to prevent a diminution of B's wellbeing of 50 wellbos.²⁴ Now contrast this with a case of Intervention (2), where A intervenes with B in order

Paternalism', p. 108 n5. There are difficulties reconciling this interpretation with Mill's Principle of Utility. However, I am unable to address this issue here. See: Stephen, *Liberty, Equality, Fraternity*, pp. 34–6. Roger Crisp, *Mill on Utilitarianism* (London: Routledge, 1997), pp. 173–200.

²³ Mill, *On Liberty*, p. 68.

²⁴ Throughout this paper, I refer to a person's wellbeing in numerical units, which I name wellbos. However, I do not hold that wellbeing can be measured in such precise terms, and I employ this measurement for illustrative purposes only.

to prevent B from diminishing C's wellbeing by 50 wellbos.²⁵ In both (1) and (2), A's intervention is contrary to the ABRONI. If the ABRONI necessarily defeats the wellbeing reason for A to intervene with B in case (1), why would it not also defeat the wellbeing reason for A to intervene with B in order to prevent the diminution of wellbeing of C in case (2)? In both (1) and (2), the subject of the intervention will cause a diminution of wellbeing of 50 wellbos. If it is wrong to intervene in (1) but not (2), it cannot be determined by the wellbeing reasons. The *weight* of wellbeing reasons in both cases is identical, and so we cannot account for a relevant difference between the two cases by appealing to the differing weight of the wellbeing reasons. Moreover, the *type* of reason in both cases is identical, as they both involve a conflict between ABRONIS and wellbeing reasons, and so we cannot appeal to this to account for a relevant difference. Thus, if there is any relevant difference between the two interventions, it cannot be grounded in the value of wellbeing.

The fact that the diminution of a person's wellbeing occurs as a result of his own actions in (1) is wholly irrelevant to the question whether the wellbeing reasons are defeated by the ABRONI.²⁶ After all, wellbeing is equally at stake in both cases, and this is regardless of whether the person has diminished it himself or not.

Now, the argument above does not rule out the possibility that there is a relevant difference between Interventions (1) and (2). Rather, I am only claiming that if there are grounds for the *Asymmetry Claim*, it cannot be due to a relevant difference in the wellbeing reasons in (1) and (2). Therefore, a different strategy is required in order for *Strong Anti-Paternalism* to be plausible.

At this point, it might be objected that the discussion so far has made an implicit assumption that the *Asymmetry Claim* must be grounded by agent-neutral reasons. However, if I allow for agent-relative reasons, there could be a relevant difference between the interventions in the two cases. Briefly, an agent-neutral reason is a reason that all persons have, whereas an agent-relative reason is a reason that necessarily refers to the person who has the reason.²⁷ Once we allow for agent-relative reasons, it might be objected

25 Let us assume for the sake of simplicity that B's wellbeing is not diminished by A's intervention in order to ensure an equal diminution of wellbeing in both (1) and (2).

26 It may be relevant to the question why there is a relevant difference between self-regarding and other-regarding actions, but this will be discussed below.

27 The idea of agent-relative and agent-neutral reasons was first introduced by Thomas Nagel in *The Possibility of Altruism* (Chichester: Princeton University Press, 1970). Nagel later adopted Parfit's label. See: Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984), p. 143. Thomas Nagel, *The View from Nowhere* (Oxford: Oxford University Press, 1986), pp. 152–6.

there could be the following relevant difference between the interventions. In Intervention (1), where A intervenes with B in order to prevent a diminution of B's wellbeing, B could have an agent-relative wellbeing reason that would be different to his agent-relative wellbeing reason in Intervention (2), where A intervenes with B in order to prevent B from diminishing C's wellbeing.

The question concerning the existence of agent-relative reasons is beyond the scope of this paper, but in any case, even if I concede that B has differing agent-relative reasons in the two cases, allowing for agent-relative reasons cannot ground the *Asymmetry Claim*. This is because the fact that B may have differing agent-relative wellbeing reasons in the two interventions does not entail that A has differing wellbeing reasons in the two interventions. Yet, this is what is needed to show that it is *all things considered* wrong to act contrary to the ABRONI in order to prevent a person harming himself, while also maintaining that it is (at least) permissible to intervene in order to prevent a person harming another person. In order to ground the *Asymmetry Claim*, there must be a relevant difference in the reasons that apply to A, given that A is performing the intervention. As a result we could accept that B has different agent-relative wellbeing reasons in the two interventions, but deny that A has any different reasons.

3 Autonomy-based Reasons of Non-interference

Rather than claiming that there is a relevant difference between the wellbeing reasons in Interventions (1) and (2), the Strong Anti-Paternalist could posit that there are other reasons that ground the *Asymmetry Claim*. One reason that potentially could serve this purpose is the ABRONI itself. Although there is no relevant difference in terms of wellbeing reasons in Interventions (1) and (2), there might be a difference in terms of ABRONIS. In this section, I consider possible differences in terms of the number of ABRONIS in Interventions (1) and (2).

First of all, we might think that there is *only* an ABRONI in (1), and not in (2). A does not act contrary to an ABRONI in relation to B in order to prevent B from diminishing C's wellbeing. It might be thought that this is because an ABRONI is not a reason not to interfere with other-regarding actions. Indeed, some proponents of Strong Anti-Paternalism explicitly stipulate that an ABRONI is limited to reasons not to interfere with only self-regarding actions. For example, Feinberg makes it clear that "respect for a person's autonomy is respect for his unfettered voluntary choice as the sole rightful determinant of his actions

except where the interests of others need protection from him."²⁸ Nevertheless, it is difficult to see a good basis for limiting ABRONIS this way. It would be unsatisfactory if it were just a stipulated feature of an ABRONI without further justification. What possible grounds could there be for this feature?

I think there are two possible grounds, and neither are satisfactory. First, we could limit ABRONIS to self-regarding actions if we could show that autonomy is only valuable with regard to self-regarding actions but not other-regarding actions. In which case, given that the value of autonomy generates the ABRONI, there would be an ABRONI in (1) but no ABRONI in (2).

The problem with this is that there is no necessary connection between the value of autonomy and whether the agent is performing a self-regarding or other-regarding action.²⁹ An agent is autonomous when he conducts himself in accordance with his own values, and the value of doing so generates the ABRONI.³⁰ It could be the case that acting in accordance with these values involves other-regarding actions. Indeed, it is possible that interfering with a person's other-regarding action could have a far greater impact on his autonomy than an interference with his self-regarding action. For example, suppose that the most important project in B's life is to murder C. B has spent a long time planning to murder C, and the fruition of this plan matters far more to B than any other of his autonomous decisions. In terms of how B shapes the direction of his life, it is more important for B that he murders C than if he performed a self-regarding action, such as autonomously reading a mildly amusing book. An interference with B autonomously murdering C would prevent B from living his life in accordance with his values to a far greater extent than an interference with B autonomously reading the book. If it is valuable to conduct one's life in accordance with one's values, then this value extends to other-regarding actions and not only self-regarding actions.

It might be objected, however, that this relies on a contentious view of the value of autonomy, and on a different account, there *would* be a necessary

28 Feinberg, *The Moral Limits of the Criminal Law*, p. 68. (emphasis added). See also VanDeVeer, *Paternalistic Intervention*, p. 59.

29 This point is explicitly accepted by some Strong Anti-Paternalists. Mill's view that "liberty consists in doing what one desires" is not necessarily self-regarding. *On Liberty*, p. 109. Cholbi holds it is valuable to act autonomously in accordance with one's own rationally chosen ends that constitute one's conception of the good. 'Kantian Paternalism and Suicide Intervention', p. 119. These ends are not necessarily self-regarding.

30 Stephen Darwall 'The Value of Autonomy and Autonomy of the Will', p. 265. The account employed here is deliberately broad. Although there are numerous variants of what it means to conduct oneself in accordance with one's own values, on all accounts of personal autonomy, the central idea is self-determination.

connection between the value of autonomy and whether the agent is performing a self-regarding or other-regarding action.³¹ Perhaps the value of autonomy is the value of non-interference regarding a person's own view about what makes his own life go best (his conception of the good), but it does not concern his view on what is permissible to do to others.³² B reading a book concerns B's view of what makes his life go best, and so this is valuable autonomy, whereas B's decision to murder C does not involve valuable autonomy, because this concerns his view on what it is permissible to do to other people.

Nevertheless, this distinction is difficult to sustain. After all, seemingly innocuous conceptions of the good directly involve one's view of what it is permissible to do to other people. For instance, conceptions of the good concerning almost all careers, such as medicine, law, and education, require one to hold views about what it is permissible to do to others. Similar thoughts apply to conceptions of the good concerning relationships such as friendship, or being a parent, and so forth. It would be implausible to think that in these cases autonomy is not valuable, and there would not be an ABRONI regarding these conceptions of the good, but in cases where a conception of the good does not involve others, such as writing novels in isolation, there would be an ABRONI.³³ Indeed, a proponent of Strong Anti-Paternalism would be unlikely to accept such an implication.

31 I thank an anonymous reviewer for this objection.

32 Danny Scoccia has a discussion of this distinction in 'Paternalism and Respect for Autonomy', p. 320.

33 One account I am unable to contend with here is Joseph Raz's view that autonomy is only conditionally valuable, that is, valuable only when employed to pursue valuable actions. Joseph Raz, *The Morality of Freedom* (Oxford: Clarendon Press, 1986), p. 417. On this account if B acts to murder C, this is not a valuable autonomous action, and so A does not act contrary to an ABRONI in relation to B in order prevent B murdering C, whereas A acts contrary to an ABRONI in relation to B in order prevent B from reading an amusing novel, because this is a valuable autonomous action. Regardless of whether it is plausible to hold that autonomy is conditionally valuable, it does not establish there is only an ABRONI in (1) and not in (2). This is because not only other-regarding actions can be disvaluable, but many self-regarding actions can also be disvaluable. Moreover, it is unlikely to be an attractive account of autonomy for a Strong Anti-Paternalist. After all, this account entails that there would not be ABRONIS in a large number of paternalistic cases typically thought to be objectionable. For example, we might think that smoking is a disvaluable activity, and if a person autonomously chose to smoke, an interference to prevent the person from smoking would not be contrary to an ABRONI. For a discussion of this account see: Robert P. George, *Making Men Moral: Civil Liberties and Public Morality* (Oxford: Oxford University Press, 1995), pp. 173–82.

Rather than trying to establish that there is only an ABRONI in (1) and not in (2) due to a difference in the value of the autonomy, there is a second ground to limit ABRONIS to reasons not to interfere with only self-regarding actions. We could view an ABRONI as the expression of the corresponding duty of a right of autonomy, and also hold that rights have a compossibility requirement, that is, the requirement that rights do not conflict.³⁴ On this account, A does not act contrary to an ABRONI with regard to B in order to prevent B from violating C's rights. This is because the corresponding duty of a right is specified such that there is no conflict with other rights. For instance, in the case of the right of autonomy, the corresponding ABRONI is specified so that there is no reason not to interfere with a person violating other rights.

Now, I am unable to engage here with the debate on whether rights compossibility is plausible.³⁵ There are good grounds to think that it is not.³⁶ But it is worth briefly noting that even if it were plausible, by itself this would not be sufficient to ground the Asymmetry Claim, and further arguments would be required, for at least two reasons. One is that we would also have to establish the greater importance of wellbeing over autonomy. This is to ensure that the scope of B's ABRONI is constrained by C's wellbeing, rather than C's wellbeing reasons being constrained by B's ABRONI. If ABRONIS had greater importance than wellbeing reasons instead, it could mean that there would be no reason to promote C's wellbeing when it is in conflict with B's ABRONI. It would be curious for a proponent of Strong Anti-Paternalism based on the significance of autonomy to hold this view. The second reason is that we would also need to establish that wellbeing rights are not inalienable.³⁷ For there to be a difference in ABRONIS in (1) and (2), we would also have to establish that B cannot violate his own rights by causing a diminution of his own wellbeing.

We should instead consider another, more promising possible difference in the number of ABRONIS in (1) and (2). In Intervention (2), when A acts contrary

For criticism of Raz's view, see: Jeremy Waldron, 'Autonomy and Perfectionism in Raz's Morality of Freedom', *Southern Californian Law Review* 62 (1989), pp. 1097–152.

34 Hillel Steiner, *An Essay on Rights* (Oxford: Blackwell, 1994), Chapter 3.

35 I do not discuss these points in detail here because compossibility has not been defended by Strong Anti-Paternalists, and indeed, one of the influential proponents of rights-based Strong Anti-Paternalism explicitly rejects the compossibility requirement. See: Joel Feinberg, *Rights, Justice, and the Bounds of Liberty: Essays in Social Philosophy* (New Jersey: Princeton University Press, 1980), pp. 221–51.

36 For a number of compelling arguments against compossibility see: Judith Jarvis Thomson, *The Realm of Rights* (London: Harvard University Press, 1990), Chapter 3.

37 For arguments for and against alienable rights see, respectively: Hillel Steiner, 'Directed Duties and Inalienable Rights', *Ethics* 123 (2013). Terrance McConnell, *Inalienable Rights: The Limits of Consent in Medicine and Law* (Oxford: Oxford University Press, 2000).

to the ABRONI with B in order to prevent B diminishing C's wellbeing, this could prevent B from acting contrary to an ABRONI with regard to C, whereas in Intervention (1) when A acts contrary to the ABRONI with B in order to prevent a diminution of B's wellbeing, this does not prevent a further action contrary to an ABRONI. Accordingly, the Strong Anti-Paternalist could explain the *Asymmetry Claim* by holding that it is permissible to act contrary to one ABRONI in order to prevent someone else acting contrary to another ABRONI. For example, A can stop B from murdering C, because by murdering C, B would be acting contrary to the ABRONI generated by the value of the autonomy of C.

In summary then, this can be set out as:

- (1) It is always *all things considered* wrong for A to act contrary to the ABRONI with B in order to increase (or prevent a diminution of) B's wellbeing.
- (2) It is not always *pro tanto* or *all things considered* wrong for A to act contrary to the ABRONI with B in order to prevent B from acting contrary to the ABRONI with C.

Although the details of this position need to be fleshed out, let us assume for the sake of argument that this move is sound.³⁸ It enables the Strong Anti-Paternalist to make a distinction between intervening in many actions that are self-regarding and intervening in the actions that are other-regarding, because other-regarding actions are also likely to be contrary to the ABRONI. As a consequence, it could ground the *Asymmetry Claim*. Furthermore, this claim can be made while maintaining that the ABRONI always defeats the reason of wellbeing, and so, it is *all things considered* wrong to act contrary to the ABRONI in order to enhance any person's wellbeing.³⁹

Although this move enables the Strong Anti-Paternalist to circumvent a number of counterintuitive implications, it is insufficient to eliminate many serious unpalatable consequences. For example, suppose that C is a non-autonomous creature, but still has wellbeing, and so it can have a life that goes better or worse. Due to the fact that C is non-autonomous, there is no ABRONI with respect to C. Now suppose that B acts to diminish C's wellbeing, and let

38 This could be understood by holding that it can be sometimes (at least) permissible to do as a right forbids and thus *infringe* that right in order to prevent a right from being violated. See: Thomson, *The Realm of Rights*, p. 122.

39 This explanation for the *Asymmetry Claim* can also make sense of the idea that we ought to respect a person's autonomy insofar it concerns a person's own conception of the good, but it does not involve respecting his view on what is permissible to do to others. We should respect a person's ABRONI, but we are permitted to act contrary to his ABRONI in order to prevent him from acting contrary to another person's ABRONI.

us assume that A's intervention with B would be contrary to the ABRONI.⁴⁰ Moreover, if we hold the Strong Anti-Paternalist view that the ABRONI always defeats reasons of wellbeing, then A's intervention to prevent B from diminishing C's wellbeing would be impermissible. This is regardless of the weight of the reason to prevent a diminution of C's wellbeing.

We can illuminate the controversial implications of this view by fleshing out C. For example, C could be a permanently mentally disabled adult human. B's actions that diminish the wellbeing of C could involve inflicting extreme torture on C for an indefinite period. Perhaps B could ensure that C remains conscious in order to experience the maximum suffering possible. B's actions could result in C having the worst possible life. Nevertheless, on the Strong Anti-Paternalist view, it would be *all things considered* wrong for A to intervene, as the intervention would be contrary to the ABRONI, and according to *Strong Anti-Paternalism*, this reason necessarily defeats all reasons of wellbeing. This is regardless of the extent that C's wellbeing is diminished by B's actions. I consider this implication of *Strong Anti-Paternalism* to be counterintuitive to the extent that it should be rejected.

Nevertheless, *Moderate Anti-Paternalism* does not suffer from this implausibility. This is because the Moderate Anti-Paternalist claims that it is not necessarily *all things considered* wrong to act contrary to the ABRONI if the reasons of wellbeing are sufficiently great. Given that the cases where B tortures or murders C are also cases in which there are likely significant reasons generated by C's wellbeing to intervene, then A's intervention with B could be permissible, and in such cases, the ABRONI would be defeated by the reasons generated by C's wellbeing.⁴¹ Thus, *Moderate Anti-Paternalism* is a more plausible position than *Strong Anti-Paternalism*.⁴²

40 I assume that B is making an autonomous choice to diminish C's wellbeing.

41 I say *likely* in order to allow for the possibility that killing or torture could make C's life go better, for example, if this event prevented C from experiencing a worse fate.

42 An anonymous reviewer has noted that Moderate Anti-Paternalism possibly has curious implications in cases such as the following. Suppose that B's most strongly held lifelong ambition is to break one of C's toes. While C does not want her toe to be broken, C is indifferent if it were broken. In this case, the ABRONI with respect to B is very weighty, given that it is B's strongest ambition. However, the ABRONI with respect to C is not that weighty, and the broken toe might not significantly diminish C's wellbeing. In such a case Moderate Anti-Paternalism seemingly entails that it would be wrong for A to intervene in order to prevent B's behavior. Yet it seems likely that Moderate Anti-Paternalists would want to reject this implication. While I accept that this would be troubling for Moderate Anti-Paternalism, I am unable to address this possible implication here.

This argument is likely to face the following objection, namely the claim that the wellbeing of non-autonomous beings is different to the wellbeing of autonomous beings, whereas my argument treats them alike. The objector could argue that while reasons to avoid a diminution of an autonomous agent's wellbeing are always defeated by the ABRONI, the reasons not to diminish the wellbeing of a non-autonomous agent are not necessarily defeated by the ABRONI. Therefore, if B wishes to torture C, the fact that C is non-autonomous has the result that his wellbeing constitutes a reason that counts in favor of intervention not being defeated by the ABRONI. Thus, it would be permissible for A to intervene in B's act towards C.

This move suffers from two problems. First, it seems to be unacceptably *ad hoc*. In order to make this move, there would have to be an explanation why the wellbeing of a non-autonomous being generates reasons that defeat the ABRONI while the wellbeing of autonomous beings generates reasons that do not defeat the ABRONI. This task is made more difficult by the fact that autonomous beings often have periods of their lives in which they are non-autonomous, such as cases of temporary mental illness, or dementia. Moreover, this position seems to suggest that the wellbeing of non-autonomous beings is more important than the wellbeing of autonomous beings. Although this curious implication may not warrant rejection of the view, it is controversial and it would need to be supported by a powerful argument, and I am unaware of such an argument.

In summary then, it seems that once we have examined the above arguments that could ground the *Asymmetry Claim*, it is clear that *Strong Anti-Paternalism* has implausible implications, namely it has the result that it would be *all things considered* wrong to interfere with a person making an autonomous decision to torture non-autonomous beings. We can avoid these implications by rejecting the idea that the ABRONI is a wellbeing-defeating reason. Instead, we should employ *Moderate Anti-Paternalism* in order to determine the wrongness of acting contrary to the ABRONI. This is a more plausible position, as it accepts that the ABRONI can sometimes be defeated by the reason of wellbeing.

It is important to emphasize that I am not claiming that there is no basis for the *Asymmetry Claim*. In the following section I propose a plausible way to account for the intuitively appealing view that there can be a relevant difference between interventions with a person's self-regarding actions and interventions with a person's other-regarding actions, grounded by the value of holding people responsible for their actions. However, I argue that even if responsibility can ground the *Asymmetry Claim*, it does not result in

paternalistic behavior that interferes with a person's autonomy being *all things considered* wrong.

4 Responsibility as Grounds for the Asymmetry Claim

In this section I will provide a plausible ground for the *Asymmetry Claim*. I will show that there is likely to be a relevant difference between intervening in a person's self-regarding actions that diminish his own wellbeing and intervening in a person's other-regarding actions that diminish the wellbeing of another person, and that this difference relates to the value of holding a person responsible for his choices. While a person who is making an autonomous choice can be held responsible for his own diminution of wellbeing, it is possible that in cases where a person acts to diminish another's wellbeing, the object of the act is not responsible for this diminution. Therefore, the importance of responsibility can ground the *Asymmetry Claim*, without holding that wellbeing reasons are different in other-regarding and self-regarding cases, or that the ABRON provides a reason to intervene. Once this is established, I consider whether the value of holding a person responsible for his choices supports the view that paternalistic behavior that interferes with a person's autonomy is *all things considered* wrong.

Before I proceed to set out my argument, I should note that there is a voluminous literature on the issue of responsibility, and it is beyond the scope of this paper to provide even a brief overview of it.⁴³ However, I am referring to moral responsibility and not causal responsibility, and I hope that the notion of moral responsibility is sufficiently clear that this does not impede the discussion. I will assume that a person is responsible for a diminution of his own wellbeing if he has autonomously acted in a way that brought about or allowed the diminution to take place.

Let me now illuminate how responsibility can ground the *Asymmetry Claim* with the following dilemma. Suppose that we are faced with a choice between two interventions. We are capable of making only one intervention, and the intervention is equally likely to be successful in both cases. In both cases, there will be a diminution of wellbeing of 50 wellbos if the intervention does not

43 For an illuminating discussion of the issue of responsibility see Scanlon, *What We Owe to Each Other*, Chapter 6. For an overview, see: Christopher Kutz, 'Responsibility', in J.L. Coleman, S. Shapiro, and K.E. Himma (eds.), *The Oxford Handbook of Jurisprudence and Philosophy of Law* (Oxford: Oxford University Press, 2002), pp. 548–87.

take place. Consequently, the reason to intervene generated by wellbeing is the same weight in each case.

Dobbie: Dobbie's actions diminish his own wellbeing by 50 wellbos.

Debbie: Debbie's actions diminish the wellbeing of an innocent stranger by 50 wellbos.

To clarify, both Dobbie and Debbie are making an autonomous choice to diminish wellbeing. Let us assume that Dobbie is responsible for the diminution of his own wellbeing. In contrast, the innocent stranger who is the object of Debbie's actions is not responsible for the diminution of his wellbeing (I will refer to him as *Innocent Stranger*). Therefore, the fact that Dobbie is responsible for the diminution of wellbeing results in an additional reason not to intervene, grounded in the value of holding Dobbie responsible. Conversely, Innocent Stranger is not responsible for the diminution of his wellbeing. As a result, when faced with a choice between intervening in the two cases, we should intervene with Debbie in order to prevent a diminution of wellbeing to the Innocent Stranger. This is despite there being equal reason to intervene generated by the value of wellbeing in the two cases. There is an additional reason that counts in favor of not intervening in the case of Dobbie due to the fact that he is responsible for the diminution of his wellbeing. This can be stated as follows:

Reason of Responsibility – If a person is responsible for a prospective or actual diminution of his wellbeing, there is a reason of responsibility not to preserve or restore his wellbeing.

This is a reason to treat people as responsible agents, and we fail to treat them as responsible agents if we preserve or restore their wellbeing if they are responsible for the prospective or actual diminution of their wellbeing. I should also be clear that the purpose of the dilemma is not to show that we would wrong Innocent Stranger if we intervene in regard to Dobbie rather than Debbie. Rather its purpose is to illuminate that there is a reason not to intervene with regard to a person responsible for a diminution of his own wellbeing that is not present when a person is not responsible for a diminution of his own wellbeing. This would be the case even if we were not faced with a choice between preventing a diminution of Dobbie's and Innocent Stranger's wellbeing. The point is that we wrong Dobbie when we intervene with regard to him to prevent a diminution of his wellbeing. Of course, it might be the case that we wrong Innocent Stranger if we do not to intervene

with regard to Debbie, but this would be due to reasons other than the reason of responsibility.⁴⁴

The upshot of the reason of responsibility is that it can account for the view that there is a relevant difference between intervening in a person's self-regarding actions that diminish his own wellbeing and intervening in a person's other-regarding actions that diminish the wellbeing of another person, and that difference grounds the *Asymmetry Claim*. It does not however entail the claim that there is *always* a relevant difference. Rather, there is only a relevant difference between the two types of intervention if there is a difference in responsibility between the individuals who are harmed. If Dobbie were not responsible for diminishing his own wellbeing, or if the Innocent Stranger is not innocent (say he requested that Debbie diminish his wellbeing), then there is no relevant difference between diminishing the wellbeing in the two cases.

This position is different from the common view that the value of autonomy generates a reason of non-interference. On the face of it, the positions seem similar, because we tend to hold people responsible for things that they have autonomously chosen. On both views, if a person makes an autonomous decision that results in him having diminished wellbeing, there is a reason not to promote his wellbeing. However, the reason of responsibility is not grounded in the value of autonomy, but in the value of holding people responsible for the things they can control. These values are distinct. We might think it is good to determine one's conduct in accordance with one's own values, but it is valuable in a quite different sense to be held responsible for one's choices.

It could be objected that the intuition elicited by this dilemma has a different explanation. It could be explained by the thought that when a person is responsible for an actual or prospective diminution of his wellbeing, the reasons to preserve or restore his wellbeing become less weighty, rather than generating a further reason of responsibility not to preserve or restore his wellbeing. This would have the implication that paternalistic intervention to preserve or restore his wellbeing would not wrong Dobbie, but would still mean we should intervene with Debbie in order to prevent a diminution of wellbeing to Innocent Stranger.

44 I thank an anonymous reviewer for noting that responsibility would not be a satisfying objection to paternalism if one argues that helping people who are (or would be) responsible for their diminution of wellbeing is less urgent than helping people who are not responsible for their diminution of their own wellbeing. After all, it seems that this view is compatible with the claim that there is nothing wrong with paternalism. I agree that this would not be a satisfactory account of the wrongness of paternalism, but this is not what I hope to show with this dilemma.

Now, I concede that there is little to say to defend why it is preferable to explain the intuition by appealing to the reason of responsibility rather than this alternative view. But I will note that my explanation has the advantage of cohering with some plausible views concerning the nature of value and its relationship with reasons. For instance, if we accept the alternative view, we would have to hold one of the following. Either we should hold that the value of a person's wellbeing diminishes in relation to responsibility, and so for example, hold that suffering is less bad if it occurs in a person who is responsible for the suffering. Or alternatively, we should hold that reasons do not track values, which means that the weight of a reason does not increase or diminish in relation to the weight of its corresponding value.⁴⁵ So that even if a person's wellbeing retains the same value irrespective of responsibility, the reasons relating to his wellbeing diminish in weight. Neither of these options are attractive, and I think it is better to think that there is a further value, namely the value of holding someone responsible. However, I accept that this is a difficult issue to resolve, and it requires a discussion of the nature of value and reasons that is beyond the scope of this paper.

5 The Reason of Responsibility as a Wellbeing-defeating Reason

The issue of responsibility could serve as a powerful argument in support of the view that paternalistic behavior that interferes with a person's autonomy is *all things considered* wrong.⁴⁶ If a person is responsible for his own diminution of wellbeing, it could be argued that the reason of responsibility defeats the wellbeing reason that counts in favor of acting paternalistically. This could have the result that if a person is responsible for having diminished wellbeing, we should not intervene to make him better off. Thus, the reason of responsibility could be a wellbeing-defeating reason not to behave paternalistically, and so, in these cases, paternalistic behavior would be *all things considered* wrong.⁴⁷

45 For a discussion on these points see: Gardner and Macklem, 'Reasons', pp. 450–5.

46 Sher has also noted that responsibility might provide an argument against paternalism. See: George Sher, *Equality for Inegalitarians* (Cambridge: Cambridge University Press, 2014), pp. 39–40.

47 The reason of responsibility also has implications that extend beyond cases of paternalism. Suppose Dobbie is responsible for the diminution of his wellbeing, but autonomously requests that we restore his wellbeing. The reason of responsibility is a reason not to restore Dobbie's wellbeing, but given that Dobbie autonomously requests it, it is not an

For example, imagine a person who makes the autonomous choice to refuse a lifesaving blood transfusion. Assume that he is responsible for this decision. Even if we accepted that he suffers a diminution of wellbeing by not receiving the transfusion, it follows that because he is responsible for his diminution of wellbeing, there is a reason of responsibility not to increase his wellbeing. Therefore, this reason of responsibility constitutes a reason not to behave paternalistically by administering the blood transfusion even though the blood transfusion would increase his wellbeing. Furthermore, this reason of responsibility could defeat the reason to enhance his wellbeing. Therefore, it would not be the case that we should intervene and paternalistically administer the blood transfusion.

This argument might appear similar to a position known as Luck Egalitarianism.⁴⁸ Broadly speaking, according to this view, inequalities are unjust unless those who are worse off are responsible for the inequality. Now there are crucial differences between the argument of Luck Egalitarians and those who could claim that the reason of responsibility is a wellbeing-defeating reason. First, as generally understood, Luck Egalitarianism restricts its scope to claims of distributive justice. Thus, it could hold that a person who is responsible for a diminution of his wellbeing is not owed an increase in wellbeing as a matter of justice, while leaving it open as to whether the person is owed an increment of wellbeing due to other non-justice reasons.⁴⁹ Second, Luck Egalitarians could restrict their arguments to claims concerning fairness, where fairness is understood as non-identical to justice.⁵⁰ They could claim that there is a reason of fairness not to restore the wellbeing of the worse off if they are responsible for being worse off. Nonetheless, they could accept that

anti-paternalistic reason not to do so. I thank an anonymous reviewer for this helpful clarification.

- 48 Richard Arneson, 'Luck Egalitarianism: A Primer', in Z. Stemplowska and C. Knight (eds.), *Responsibility and Distributive Justice* (Oxford: Oxford University Press, 2011), pp. 24–50. G.A. Cohen, 'On the Currency of Egalitarian Justice', in M. Otsuka (ed.), *On the Currency of Egalitarian Justice, and Other Essays in Political Philosophy* (Oxford: Princeton University Press, 2011), pp. 3–43. Erik Rakowski, *Equal Justice* (Oxford: Oxford University Press, 1991). John E. Roemer, *Theories of Distributive Justice* (Cambridge: Harvard University Press, 1996).
- 49 This raises many questions concerning the relationship between justice and morality that are beyond the scope of this paper.
- 50 For a discussion on a related matter see: G.A. Cohen, 'Fairness and Legitimacy in Justice, And: Does Option Luck Ever Preserve Justice?', in S. De Wijze, M. Kramer and I. Carter (eds.), *Hillel Steiner and the Anatomy of Justice: Themes and Challenges* (London: Routledge, 2009), pp. 3–21.

this reason of fairness can be defeated by the wellbeing reasons to enhance the wellbeing of those who are responsible for being worse off. Thus, it could be *all things considered* just to enhance the wellbeing of the worse off, even if they are responsible for being worse off.

In any case, I think that it is a mistake to hold that the reason of responsibility is a wellbeing-defeating reason. Furthermore, I think that the plausibility of the view that it is a wellbeing-defeating reason is due to two factors. First, there are significant instrumental reasons to hold people responsible for their actions. For instance, if we create a disincentive for people not to take large risks, it is likely to result in them having lives with greater wellbeing. Thus, if we are concerned with promoting wellbeing, holding people responsible for their actions could encourage people to act in such a way to have greater wellbeing.

Second, and most importantly, in order to enhance, or prevent a diminution of, a person's wellbeing in many cases, it is necessary to either diminish, or refrain from enhancing someone else's wellbeing, due to (a) a reallocation of resources, which could have been used to make other people better off or (b) by requiring a person to act in such a way to restore the wellbeing of a person, when instead they could be pursuing their own projects. Therefore, if we do not hold that responsibility is a wellbeing-defeating reason, then there could be significant costs to the wellbeing of others.

These concerns are important, but given that they are contingent on the consequences of failing to hold people responsible, they do not entail that the reason of responsibility is a wellbeing-defeating reason. If it were a wellbeing-defeating reason, it would have these defeating properties regardless of the above concerns about costs to others. I will illustrate this with the following example, which does not involve costs to others.

Imagine a person, Wally, who is responsible for a permanent diminution of his own wellbeing. Wally engages in a high-risk activity, namely tightrope fire breathing. He understands that if the trick goes wrong, he is likely to be severely burned, and understands that this will result in a permanent diminution of his wellbeing, as it would cause significant suffering for Wally throughout his life. He has not been coerced into engaging in this activity, and he freely chooses to do it.

One day, Wally's tightrope fire breathing trick goes wrong, and results in a permanent diminution of his wellbeing. Without help, Wally will have a life of severe suffering. Let us assume that Wally can be held responsible for the trick going wrong, and so, can be held responsible for this permanent diminution of his wellbeing.⁵¹ Now, imagine that with no cost to others we could restore

51 Also, Wally decided against taking out insurance for accidents. For a discussion on insurance and responsibility, see Ronald Dworkin's discussion on insurance in Ronald

Wally's wellbeing. If we pressed a button, a healing laser would eliminate Wally's suffering, and as a result, this would increase Wally's wellbeing.⁵²

If the reason of responsibility necessarily defeats the reason of wellbeing then the fact that Wally is responsible for his diminution of wellbeing would entail that we should not press the button and costlessly restore Wally's wellbeing. However, it seems more plausible to think that if someone could make Wally's life better, at no cost, then he should press the button and increase Wally's wellbeing. If someone failed to press the button under these costless conditions it seems clear that he is doing something wrong.

Of course, the case of Wally is different from more realistic cases where there is a cost to restoring a person's wellbeing. Although I accept that this complicates matters, even if there is a cost attached to increasing the wellbeing of a person like Wally who is responsible for his diminished wellbeing, it does not follow that the reason of responsibility suddenly transforms into a wellbeing-defeating reason. If there is a value to holding someone responsible for the diminution of his own wellbeing, then this value remains even if there are no costs to others. If there are costs to others, then the disvalue of the costs to others provides the reason to hold someone responsible, and not the reasons generated by the value of responsibility.

It is possible that the reason to restore Wally's wellbeing would eventually be defeated if the costs to others are sufficiently large. In part, this would be determined by the weight of the wellbeing reasons in each case.⁵³ Suppose that Wally is responsible for a permanent diminution of 100 wellbos to his wellbeing, and this results in Wally having the worst possible life. We can restore Wally's wellbeing at a cost of 1 wellbo to Innocent Stranger. Given the significant harm to Wally and the minor harm to Innocent Stranger, it seems plausible to think that we should intervene to prevent the diminution of Wally's wellbeing, even though it has costs to Innocent Stranger. We can imagine numerous scenarios where restoring Wally's wellbeing requires a similarly small diminution of wellbeing for two or more Innocent Strangers. This also may not defeat the reason to increase Wally's wellbeing. However, things may change if we

Dworkin, 'What Is Equality? Part 2: Equality of Resources', *Philosophy and Public Affairs* 20 (1981), pp. 101–31.

52 I assume for the sake of argument that the technology is sufficiently advanced that pressing the button is costless. Furthermore, using the beam on Wally does not deprive anyone else of an increase in his or her wellbeing.

53 I say in part, because there could be other reasons, such as deontic side-constraints. It could be claimed that there are deontic reasons not to impose a burden on any person, even if imposing this burden results in a greater amount of total wellbeing. Nonetheless, these deontic reasons would be independent of the reason of responsibility.

increase the costs, or the number of Innocent Strangers. For instance, things become less clear if there is a cost of 20 wellbos to five Innocent Strangers. I accept that the question of where to draw the line is problematic. All I am willing to conclude here is that the reason of responsibility is not a wellbeing-defeating reason. But it can be a relevant difference between cases where a person harms himself, and cases where a person harms another person, and as such, it grounds the *Asymmetry Claim*. Further, a person can be held responsible for the diminution of his own wellbeing, but this does not entail that the reason of responsibility necessarily defeats the reason of wellbeing. Instead, the reason of wellbeing will be defeated if the competing reasons are sufficiently weighty, such as there being significant costs to others.

In summary then, even if an individual can be held responsible for the actions that diminish his own wellbeing, the reason of responsibility does not necessarily defeat the reason to intervene generated by the responsible individual's wellbeing. At most, it entails that there is an additional reason not to intervene in cases where an individual is responsible for a diminution of his own wellbeing. Therefore, in these cases, paternalistic behavior would not be *all things considered* wrong.⁵⁴

54 An anonymous reviewer has noted that the reason of responsibility cannot adequately account for the wrongness of some of the most common forms of paternalism, namely paternalistic behavior that occurs before the paternalizee is responsible for being worse off. The most striking examples of this are paternalistic laws that prohibit certain behavior. Imagine a law that made it illegal to consume recreational drugs for paternalistic reasons. In order for the reason of responsibility to ground the wrongness of this paternalistic law, it would need to hold the paternalizee responsible for behavior that he has not yet performed. Indeed, even if the paternalizee would never consider consuming recreational drugs because of the existence of the law, the paternalizee would need to be held responsible in order for the reason of responsibility to account for the wrongness of the paternalistic law. Yet it would be implausible to hold a person responsible for this. However, it is important to note that the reason of responsibility can still account for the wrongness of many paternalistic interventions, namely interventions to prevent the effects of actions that have already taken place, and interventions to prevent actions that are about to take place. I am unable to explore here at what point a person precisely can be held responsible for diminutions of his wellbeing. But it is plausible to think that a person can sometimes be held responsible for a diminution of wellbeing, even before he performs the action that results in the diminution. For example, suppose a person was about to autonomously ingest a recreational drug, but before he does so, you knock it from his hand to stop him ingesting it. In this case, it is not implausible to hold the paternalizee responsible for the potential diminution of wellbeing caused by the drug, even though he had not yet ingested it.

The resultant position could be described as ‘Luck Prioritarianism’ because, *ceteris paribus*, there is greater overall reason to enhance the wellbeing of those who are not responsible for diminished wellbeing than those who are responsible for diminished wellbeing, even if the weight of the wellbeing reasons is the same in both cases. As a result, we could describe those who are not responsible for their own diminished wellbeing as having priority over those who are responsible for their own diminished wellbeing. Nonetheless, if the wellbeing reasons of the worse off are sufficiently weighty, it follows that even when a person is responsible for being worse off, there may be greater reason to enhance his wellbeing than others who are not responsible for being worse off.⁵⁵

6 Conclusion

It will be helpful to conclude the paper by summarizing my argument. I began by considering the common thought that it is *all things considered* wrong to interfere with a person’s self-regarding autonomous actions, irrespective of the benefit of the intervention to that person. I argued that this view is only plausible if there is a relevant difference in the reasons concerning an intervention in a person’s self-regarding actions and an intervention in his other-regarding actions. Otherwise, an absolute prohibition of one type of intervention would mean an absolute prohibition of the other, and it would be implausible to think that it is *all things considered wrong* to interfere with a person’s autonomous other-regarding actions. I then considered possible relevant differences in the reasons concerning a person’s self-regarding and other regarding actions. I argued that it cannot be a relevant difference in wellbeing reasons, and that the ABRONI itself also cannot satisfactorily explain the difference without implausible implications. I then considered a plausible alternative basis for the relevant difference, namely the value of holding a person responsible for his choices. I argued that responsibility does not result in paternalistic behavior

55 For a similar view, see Richard Arneson’s responsibility-catering prioritarianism, and Kasper Lippert-Rasmussen’s desert-accommodating prioritarianism: Richard Arneson, ‘Luck Egalitarianism and Prioritarianism’, *Ethics* 110 (2000), pp. 339–49, pp. 344–5. Kasper Lippert-Rasmussen, ‘Private Discrimination: A Prioritarian Desert-Accommodating Account’, *San Diego Law Review* 43 (2007), pp. 817–56, pp. 833–6. Shlomi Segall uses the term ‘luck prioritarianism’ in a different sense, where prioritarianism refers to giving priority to the worse off, rather than to those held responsible for being worse off. See: Shlomi Segall, *Health, Luck, and Justice* (Oxford: Princeton University Press, 2009), pp. 118–9.

being *all things considered* wrong. At most, it entails that there is an additional reason not to intervene in cases where an individual is responsible for a diminution of his own wellbeing. This means those who are not responsible for their own diminished wellbeing have priority over those who are responsible for their own diminished wellbeing.

Acknowledgements

I would like to thank the following for their helpful comments and discussion on the paper: Mattias Brinkman, Kimberley Brownlee, Alena Buyx, Ian Carroll, Matthew Clayton, Roger Crisp, Lorenzo Del Savio, Cécile Fabre, Joanna Firth, Lorenzo Greco, Clare Heyward, Joe Horton, Sam Kiss, Holly Lawford-Smith, Hugh Lazenby, Kasper Lippert-Rasmussen, Andrew Mason, David Miller, Chris Mills, Jonathan Quong, Rebecca Reilly-Cooper, Simon Rippon, Tom Sinclair, Hillel Steiner, Adam Swift, Victor Tadros, Tony Taylor, Frej Klem Thomsen, and Pete West-Oram. Earlier versions of the paper were presented to the Manchester Centre for Political Theory, University of Manchester, the Centre for the Study of Social Justice, University of Oxford, and the Centre for Ethics, Law, and Public Affairs at Warwick University. I am grateful to the audiences at these events for their questions and criticisms. Finally, I would like to thank the anonymous reviewers for their many excellent suggestions and objections. The paper was in part written while funded by the Wellcome Trust (100705/Z/12/Z), and the German Ministry of Education and Research (01GP1311).

Biographical Note

David Birks is a Departmental Lecturer in Political Theory at the Department of Politics and International Relations, University of Oxford, and an Early Career Research Fellow at the Oxford Research Centre for the Humanities (TORCH), University of Oxford. He is also a Research Fellow at the University of Kiel. His research focuses on paternalism, punishment, perfectionism, and public reason.