

Defining Paternalism

Sebastien Bishop-Bennett

UCL

MPhilStud Philosophical Studies

Word Count: 29,249

I confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

---

## Abstract

Much of the philosophical engagement with the issue of paternalism, especially in the last couple of decades, has focused on important normative issues such as: 'what, if anything, is morally problematic about paternalism?', 'when is paternalism justified?', and 'how concerned should we be, morally speaking, with paternalism?'.

My thesis seeks to take a step back and asks a more fundamental, conceptual question, upon which these more practical, normative issues supervene: *What precisely defines an act of paternalism?*

To this end, this thesis is divided into five parts; an introduction followed by four chapters. As well as setting out the aims of the thesis, the introduction outlines some of the basic, uncontroversial features of paternalism. Chapters 1, 2, and 3 then examine the rich philosophical literature on defining paternalism; each chapter examining a different approach to defining paternalism.

Through the outlining of some of the uncontroversial features of paternalism, and subsequent investigation of where philosophers have gone wrong in defining paternalism, I develop seven '*Challenges*' an accurate definition of paternalism must meet. I also argue that no definition of paternalism currently in the philosophical literature, can meet each of these Challenges (or even just meet Challenges 5, 6, and 7; three connected Challenges that are particularly important).

Finally, in Chapter 4, I present an original definition of paternalism consisting of three

3

individually necessary, and together sufficient conditions. I argue that this definition can in fact meet each of the seven 'Challenges'.

Table of Contents

Abstract	2
Table of Contents	4
<i>Thesis Introduction</i>	6
<i>Thesis Structure</i>	7
<i>Thesis Motivation</i>	8
<i>Paternalism: three uncontroversial features</i>	12
 <i>Chapter 1: an examination of those more minimalist, straightforward definitions of paternalism</i>	 21
<i>Part 1: Minimalist definitions that emphasise paternalist interferences as benevolent acts that are contrary to Y's wishes/desires/what Y has consented to</i>	22
<i>Part 2: Minimalist definitions that emphasise paternalist interferences as benevolently motivated coercive, or liberty-limiting actions</i>	31
 <i>Chapter 2: defining paternalism as a benevolent interference with Y's autonomy</i>	 39
<i>Part 1: Dworkin (1983)</i>	39

<i>Critique of Dworkin (1983)</i>	40
<i>Part 2: Dworkin (2014)</i>	48
<i>Critique of Dworkin (2014)</i>	48
<i>Part 3: Clarke</i>	50
<i>Critique of Clarke</i>	52

*Chapter 3: some less conventional definitions of paternalism*      60

<i>Part 1: Gert and Culver</i>	61
<i>Critique of Gert and Culver</i>	67
<i>Part 2: Jonathan Quong</i>	73
<i>Critique of Quong</i>	74

*Chapter 4: a new definition of paternalism*      80

<i>List of Challenges</i>	81
<i>Challenge 1</i>	82
<i>Challenge 2</i>	87
<i>Challenge 3</i>	87
<i>Challenge 4</i>	88
<i>Challenge 5</i>	94
<i>Challenge 6</i>	97
<i>Challenge 7</i>	105
<i>Three Points Clarified</i>	108
<i>Our Final Definition</i>	115
<i>Implications of this Definition</i>	116
<i>Bibliography</i>	119

## Thesis Introduction

Consider the following three cases:

### CAKE

*Dmitri (Y) is looking forward to getting home, opening his fridge, and eating the chocolate cake he bought earlier. But when he gets home, he finds that his brother Ivan (X) has thrown away his chocolate cake, because Ivan thinks Dmitri shouldn't eat so much unhealthy food.*

### SUICIDE

*Anna (Y) has rationally chosen to commit suicide by jumping in front of a train. But her friend (X), despite knowing that this is a decision Anna has come to rationally, forcibly prevents Anna from jumping in front of the train.*

### BLOOD TRANSFUSION

*Rodion (Y) is ill and in hospital. He tells his doctor (X) that because of certain ethical beliefs, no matter what happens he does not want a blood transfusion. Despite this, when Rodion falls unconscious, his doctor administers a blood transfusion. The doctor acts because he thinks that it is for Rodion's own good.*

In each of these cases X (the paternaliser) paternalises another agent, Y (the paternalisee).<sup>1</sup> Much of the philosophical engagement with the issue of paternalism, especially in the last couple of decades, has focused on important normative issues like:

'what, if anything, is morally problematic about paternalism?', 'when is paternalism

<sup>1</sup> Throughout this thesis, I will use 'X' to refer to the paternaliser/would-be paternaliser, and 'Y' to refer to the paternalisee/would-be paternalisee

justified?', and 'how concerned should we be with state paternalism?'

My thesis seeks to take a step back and asks a more fundamental, conceptual question, upon which these more practical, normative issues supervene: What precisely defines an act of paternalism? Of course, a rough and ready answer might be that X paternalises Y when he interferes with Y's life, against Y's will, for Y's own good. But I want to investigate the issue more rigorously; examining precisely what kinds of elements are necessary for paternalism, what kinds of attitudes, interferences and intentions characterise paternalism, and what different forms paternalism can take.

### *Thesis Structure*

To this end, this thesis is divided into five parts; an introduction followed by four chapters. As well as setting out the aims of the thesis, the introduction outlines some of the basic, uncontroversial features of paternalism. In particular, three features are emphasised. Chapters 1, 2, and 3 then examine the rich philosophical literature on defining paternalism; each chapter examining a different approach to defining paternalism. Finally, in Chapter 4, I present an original definition of paternalism.<sup>2</sup>

Now there are at least two reasons why the analysis of the literature and the outlining of the basic features of paternalism, conducted in the introduction and first three chapters of

---

<sup>2</sup>Of course, we have already noted that much of philosophical engagement with the issue of paternalism concerns more practical, normative questions. Yet there nonetheless exists an extensive and interesting philosophical literature on defining paternalism. Some philosophers dedicate entire essays to the question of how to define paternalism, although more commonly this question is examined as a precursor to examining the practical issues mentioned above.

Either way, generally the philosopher will present a battery of cases- some supposedly uncontroversially cases of paternalism, others supposedly uncontroversially not cases of paternalism- before attempting to provide a definition of paternalism that is not only extensionally accurate with regards to these cases, but also furnishes us with a precise, philosophically rigorous description of the would-be paternaliser's engagement with the would-be paternalisee

this thesis, are a necessary precursor to our presentation of an original definition of paternalism in Chapter 4. First, the analysis of the philosophical literature on defining paternalism provides the motivation for the presentation of a new definition of paternalism. For as I will argue, no definition currently featured in the literature, accurately captures what it means for one agent to paternalise another. And thus a new definition is called for.

Second, through outlining some of the basic features of paternalism in the introduction, and examining where various philosophers have gone wrong in defining paternalism in chapters 1, 2, and 3, I deduce a series of '*Challenges*' that an accurate definition of paternalism must meet e.g. '*Challenge 1: an accurate definition of paternalism must reflect how paternalism necessarily involves a benevolent motive on the part of the paternaliser, concerning the the paternalisee*'. These '*Challenges*', seven in total, guide our investigation of paternalism, and ultimately the construction of a new definition of paternalism in Chapter 4.

### *Thesis Motivation*

I now want to say a few words regarding the motivation for writing a thesis on defining paternalism. In particular I highlight three motivating factors.

First, I take 'paternalism' to be an intrinsically interesting concept, worthy of philosophical study and clarification. In this sense, paternalism may be taken as similar to concepts like '*coercion*' and '*freedom*', which have rich philosophical literatures dedicated to simply clarifying the meaning of these concepts. What makes paternalism so interesting?

Perhaps the most obvious answer, is the way paternalism combines conflicting morally

relevant features. As I will argue in this thesis, when X paternalises Y, he interferes with Y's life (in a way that will be clarified) while disregarding Y's relevantly informed judgements/wishes concerning what is in his own best interests. Such aspects are at least prima facie morally problematic. But he does so in order to promote Y's good. Because he has Y's best interests at heart. This paints the picture of a conflicted, morally complicated concept.

Second, I take the rich philosophical literature on defining paternalism to be itself something worthy of study. Within this literature, philosophers like Gerald Dworkin, Joel Feinberg, Richard Arneson, Richard Thaler and Cass Sunstein, offer competing characterisations of paternalism based on an analysis of a battery of cases. I take it to be a philosophically interesting issue whether any of these philosophers are successful, and if none are (which is the conclusion I ultimately argue in favour of), where precisely these philosophers went wrong, and whether we can do one better and offer a satisfying definition of paternalism ourselves.

Third and finally, as noted at the very beginning of this introduction, there are several important normative questions regarding paternalism; questions such as 'what is (morally) problematic about paternalism?', 'when is paternalism justified?'. But I take it that how we answer the conceptual question this thesis investigates (i.e. what precisely defines paternalism?), crucially informs the investigation of these more practical, normative questions.

After all, what kind of moral problem paternalism poses, can plausibly be viewed as the product of on the one hand, what (morally relevant) features necessarily make up an act of

paternalism, and on the other hand, what different forms paternalism can take. And deciding what morally relevant features necessarily make up an act of paternalism and what forms paternalism can take, plausibly requires defining paternalism.

What's more, in this thesis I argue that even the most well respected and widely accepted definitions of paternalism, upon which we often find the normative analysis of paternalism based, are guilty of distorting the concept in various, normatively significant ways. For instance, Dworkin's 1983 and 2014 definitions of paternalism are two of the most widely accepted definitions in the philosophical literature. And it is unsurprising to find many philosophers base their normative analysis of paternalism upon these definitions.<sup>3</sup>

But in several ways we will investigate in more detail later, Dworkin's accounts are problematic. Take Dworkin's suggestion that paternalist acts necessarily interfere with Y's *autonomy or liberty*. In Chapter 2 I dispute this claim. If I am right, then this has consequences for the normative question. For if paternalism really does necessarily involve an interference with Y's autonomy or liberty, this gives us an important moral reason to avoid paternalism. But if paternalism need not involve an interference with Y's autonomy or liberty, perhaps there are forms of paternalism that are not so morally problematic.

I thus find myself sympathetic to Louis Groarke's suggestion that:

*“If we hope to avoid wildly inaccurate conclusions, it behooves us to base our analysis of paternalism on genuine rather than specious examples of*

---

<sup>3</sup>James and Morden (1988), Wilson (2011), Bergelson (2009), Perez (2008), Uhl (2011) to name a couple of examples

*paternalistic intervention”<sup>4</sup>*

Now recently Kalle Grill (2015) has cast doubt on whether answering the conceptual question (i.e. what precisely defines an act of paternalism?) really is so important for investigating the normative questions surrounding paternalism (i.e. what is morally problematic about paternalism?, when can paternalism be morally justified?).

Grill suggests that an effective way to investigate the normative questions surrounding paternalism, is to analyse the claims of the anti-paternalist. And the best way to interpret the anti-paternalist's claims, he convincingly argues, is as a special claim about the function of reasons in different contexts. Specifically, the claim that a benevolent motive regarding Y (the motivation involved in paternalism), does not count as a reason in favour of, or justification for, whatever else paternalism involves (e.g. interfering with Y's life in some way, perhaps acting contrary to Y's relevantly informed will, perhaps interfering with Y's voluntary conduct).

Why should we think that this shows the normative debate about paternalism can be addressed without entering into the conceptual debate? Grill's argument seems to be that, on this analysis of anti-paternalism, the precise details of what paternalism involves are not so important. For whatever features make up an act of paternalism apart from a benevolent motive regarding Y- be it an interference with Y's autonomy, or an interference with Y's actions, or an action that is known to be contrary to Y's wishes, or something else entirely- the anti-paternalist's critique of paternalism is going to be the same: That these things are not justified by the goal of promoting Y's good.

---

<sup>4</sup>Groarke (2002, p205)

But while I take Grill's argument to be a shrewd analysis of the structure of the anti-paternalist's claims, I dispute whether it really, without being supplemented by an answer to the conceptual question, answers the normative questions paternalism raises. For it stands to reason that if we want to answer the normative questions that surround paternalism, we need to know more than the structure of the anti-paternalist's claim (i.e. that a benevolent motive regarding Y does not justify whatever else is involved in paternalism). We need to assess whether this claim is justified, or not. And assessing whether the anti-paternalist's claim is justified, plausibly requires that we isolate precisely what kind of features make up a paternalist act other than a benevolent motive; before investigating whether these features really are not justified by a benevolent motive. In other words, we still need an answer to the conceptual question.

### *Paternalism: three uncontroversial features*

I want to end this introduction by drawing out three general, uncontroversial features of paternalism that virtually all philosophers agree upon. This will help clarify the concept a little, as well as lay the groundwork for the rest of the thesis.

#### *i) a benevolent motive*

First, it is widely accepted within the philosophical literature that paternalism necessarily involves a benevolent motive on X's part regarding Y. That is to say, it is widely accepted that an accurate definition of paternalism must feature what we might call a benevolence condition.

Of course, there are different ways of characterising a benevolent motive. Indeed, later I will argue that many philosophers actually fail to characterise a benevolent motive accurately. But the point remains that virtually all philosophers agree that paternalism necessarily involves a benevolent motive, although they may disagree on the precise details of what that is. Thus, Gerald Dworkin (2014) writes that paternalism involves X acting only because he believes his actions: “will improve the welfare of Y (where this includes preventing his welfare from diminishing), or in some way promote the interests, values, or good of Y”. Danny Scoccia demands that in order to paternalise Y, X must act out of a “beneficent concern for Y’s welfare”.<sup>5</sup> Simon Clarke states that X acts paternalistically towards Y only “to the extent that X does so in order to promote Y’s good”.<sup>6</sup> While Gert and Culver’s definition of paternalism requires that X believes “that his action is for Y’s good”.<sup>7</sup>

Before we move on, it should be noted that while it is very widely accepted within the philosophical literature that paternalism involves a benevolent motive on X’s part regarding Y, this acceptance is not quite unanimous. In particular Seana Shiffrin (2000) has forcefully argued that rather than endorsing a benevolence condition, we should instead endorse what I will refer to as a *superciliousness condition*. More precisely Shiffrin claims that a necessary condition of paternalism, should simply be that X’s actions are motivated by: *X’s belief that his judgement or agency regarding some interest or matter that legitimately lies within Y’s control, is superior in some respect to Y’s judgement or agency regarding said interest or matter.*

The superciliousness condition is importantly broader than the benevolence condition. As

<sup>5</sup>Scoccia (2008, p2), ‘their’ changed to ‘Y’s’

<sup>6</sup>Clarke (2002, p82)

<sup>7</sup>Gert and Culver (1979, p199), changed subjects. ‘S’ changed to ‘Y’

Shiffrin points out, whenever X meets the benevolence condition, he will necessarily also meet the superciliousness condition. After all, if X believes he knows better than Y how to promote Y's good, he surely counts as believing his judgement/agency to be superior to Y's regarding some interest that lies within Y's legitimate control. By contrast, X can meet the superciliousness condition without also meeting the benevolence condition; for instance, when X interferes with Y's life just because he is annoyed at watching Y fail, and thinks he could do better.

Shiffrin argues in favour of the *superciliousness condition* over the traditional benevolence condition, on the following grounds: That cases where X interferes with Y's life and meets the superciliousness condition, are relevantly similar normatively speaking, to cases where X interferes with Y's life and meets the benevolence condition. This is (allegedly) evidenced by the fact we have relevantly similar *normative reactions* to both types of case.<sup>8</sup> Coupled with the assumption that our classification of whether a given case counts as paternalist should depend on the normative significance of those cases<sup>9</sup>, Shiffrin concludes that if we think that X paternalises Y when X interferes with Y's life and meets the benevolence condition, we should also think that X paternalises Y whenever X interferes with Y's life and meets the superciliousness condition.

The key problem with Shiffrin's argument, is that it is far from clear whether these two types of case really are necessarily normatively similar. Consider how, when X interferes with Y while also meeting the benevolence condition, X necessarily interferes with Y for *Y's own good*. This is a necessary feature of such cases. However, when X interferes with Y and meets the superciliousness condition, X need not interfere with Y for Y's own good (although he might). For instance, X might simply interfere with Y because he wants to

<sup>8</sup>Shiffrin (2002, p218)

<sup>9</sup>Shiffrin (2002, p212)

protect Z, whom Y is supposed to be (but is failing in) protecting.

Contra Shiffrin, this points to a significant moral difference between cases where X interferes with Y and meets the benevolence condition, and cases where X interferes with Y and meets the superciliousness condition. When X interferes with Y in order to protect Z, his interference can be (at least *prima facie*) morally justified by appealing to *the harm principle*. Yet no such appeal can ever be made in the case where X interferes with Y for Y's own good (when X interferes with Y and meets the benevolence condition).

Actually, the issue cuts both ways. For X can also meet the superciliousness condition (but not the benevolence condition), by interfering with Y merely because he is personally tired of seeing Y fail in some matter Y has legitimate control of. For instance, X might prevent Y from being able to cook his own food because X is sick of watching Y make such a bad job of it. Or, to take one of Shiffrin's examples, X might talk over Y while Y is trying to express a point, because X thinks he can express the point so much more clearly. Interfering with Y in this way (because one simply wants to, and with no regard for Y's own good), is very different, morally speaking, to interfering with Y for his own sake, as X necessarily does when he meets the benevolence condition.

The issue can be expressed in terms of appreciating the moral significance of the *separateness of persons*.<sup>10</sup> In the former case, X's actions can be (at least *prima facie*) morally justified to Y. Why? Because although X imposes some sacrifice on Y, Y stands to gain from this sacrifice. In the latter case however, no such justification can be provided; for X imposes a sacrifice on Y so that some other agent can gain. Appreciating the difference between these cases amounts to appreciating that there is a (at least *prima*

---

<sup>10</sup>See Otsuka and Voorhoeve (2009, p179)

facie) moral difference between making Y suffer for his own sake, and making Y suffer for someone else's sake. That there is a certain moral difference between the intrapersonal case and the interpersonal case. If Shiffrin's account cannot reflect this moral difference, then so much the worse for Shiffrin.

With Shiffrin's objection replied to, let us say that the first '*Challenge*' an accurate definition of paternalism must meet, is as follows

#### Challenge 1

To reflect that paternalism necessarily involves a benevolent motive on X's part regarding Y.

#### *ii) voluntariness*

The second uncontroversial feature of paternalism is a little more complicated. It relates to a distinction often drawn in the literature that Joel Feinberg describes as of the "first importance"<sup>11</sup>. The distinction between *hard paternalism* and *soft paternalism*. When philosophers talk about paternalism (and I have been no different in this respect), what they are usually actually talking about is 'hard paternalism'. Hard paternalism (or just plain 'paternalism') is standardly taken to involve, in some rough and ready way that this thesis aims to clarify, a benevolent interference with Y's *voluntary conduct*. Soft paternalism, by contrast consists in X benevolently interfering with Y, but not aiming to interfere with Y's voluntary conduct.<sup>12</sup>

Along these lines, it is commonly assumed that an accurate definition of paternalism (that

<sup>11</sup>Feinberg (1986, p12)

<sup>12</sup>See Grill (2011)

is, *hard paternalism*) must discount two different types of cases that qualify as merely *soft paternalism*. First, cases like BRIDGE where X merely aims to benevolently interfere with Y's *non-voluntary* conduct. Second, cases like ONCOMING TRAFFIC where X benevolently interferes with Y, merely in order to *ascertain* if Y is acting voluntarily or not.<sup>13</sup>

### BRIDGE

*Y is about to step onto a bridge. Little does he know, the bridge is broken, and thus will collapse and kill him if and when he steps onto it. By contrast, X knows that the bridge is broken. X sees Y walking towards the bridge, and working on the assumption that Y is unaware that the bridge is about to collapse, X decides to forcibly prevent Y from stepping onto the bridge, say by tackling him to the ground before he gets to the bridge.*

### ONCOMING TRAFFIC

*Y is about to cross the road and step into oncoming traffic. X sees this, but is unsure whether Y is acting involuntarily or not. After all, Y might be attempting to deliberately end his own life, or he may simply be crossing the road absent-mindedly. X decides to forcibly prevent Y from crossing the road so he can quickly ascertain whether Y is acting voluntarily or not. If it turns out that Y is acting voluntarily, X will not stand in his way.*

In light of this, it is tempting to say that the second Challenge an accurate definition of paternalism must meet, is to reflect that (hard) paternalism necessarily involves X interfering with aspects of Y's life that are voluntary. Actually, later on in Chapter 4 I will argue that this principle is incorrect. Still, I agree that X does not paternalise Y in cases like BRIDGE or ONCOMING TRAFFIC. Thus, let us simply install the following as the second Challenge an accurate definition must meet:

---

<sup>13</sup>ibid

### Challenge 2

That X does not paternalise Y in cases like BRIDGE where X interferes with Y merely because he believes Y is acting *involuntarily*. Nor in cases like ONCOMING TRAFFIC where X interferes with Y merely in order to *ascertain* if he is acting voluntarily or not.

Before we move on to the third uncontroversial feature of paternalism, I want to head off an objection. It may follow from what we have just said, that a father (X) does not (hard) paternalise his young son (Y), when he benevolently interferes with the young son's actions against the son's will e.g. by banning the son from playing his Playstation before he has done all his homework. After all, depending on how young the son is, his decision/desire to play on his Playstation may not count as sufficiently voluntary.

Some philosophers object to this conclusion. Gert and Culver (1979) point out that the concept of (hard) paternalism derives from how a parent treats their young child when they benevolently interfere with the young child's life/actions/decisions/etc. From this they infer that it is “paradoxical” to suggest that such a case is not a case of (hard) paternalism.<sup>14</sup>

But Gert and Culver's argument is flawed. The concept of (hard) paternalism may indeed derive from how a parent treats their young children. Yet this is not because the latter is an example of the former. Rather, it is because, as Feinberg (1986) suggests, to label X's treatment of Y as (hard) paternalist is to draw a certain analogy. It is to claim that X is treating Y, as a parent typically treats their young child.<sup>15</sup>

---

<sup>14</sup>Gert and Culver (1976, p54)

<sup>15</sup>Feinberg (1986, p23)

This furnishes us with an explanation of how the concept of hard paternalism relates to the parent/young child case, without assuming that the parent necessarily (hard) paternalises the young child in the parent/child case.

### *iii) rational persuasion*

The third and final uncontroversial feature of paternalism I want to draw attention to, is how X does not paternalise Y when he merely engages Y in *rational persuasion*.<sup>16</sup> It follows from this, that had X merely tried to rationally persuade Y not to eat so much cake in CAKE, or merely tried to rationally persuade Y against jumping on the train tracks in SUICIDE, then X would not have paternalised Y in these cases. Along these lines, let us make the following the third *Challenge* an accurate definition of paternalism must meet.

#### Challenge 3

To reflect that X does not paternalise Y (even when all other conditions are met) when he merely aims to engage Y in rational persuasion.

At this point, we must be careful here not to conflate rational persuasion with the presentation of argument. For the essence of mere rational persuasion, and the reason that it precludes paternalism, is that mere rational persuasion does not interfere with Y in any significant sense. But sometimes X can present an argument to Y, while interfering with Y. Consider, for instance, the case where X deploys highly rhetorical devices, such as powerful emotional appeals. Similarly, consider the case where X persists with the presentation of argument against Y's will, knowing that Y does not want to hear any such arguments. In such cases X may well present an argument to Y, but his actions go beyond

---

<sup>16</sup>See Scoccia (2008)

merely presenting Y with the facts.

Of course, we do not want to equate rational persuasion with nothing but the dry presentation of logical argument either. Very few discussions or arguments consist in nothing but logical, rational appeal, and so demanding that rational persuasion be so austere a concept would be implausible. Instead, the rather more subtle point is that sometimes X can make an argument too forcefully, or depend too much on emotional manipulation; and in such cases X's actions go beyond the limits of mere rational persuasion.

### Conclusion

With these three points clear, and their corresponding Challenges outlined, let us now move to Chapter 1. There we will examine two of the more straightforward, perhaps more commonsensical approaches to defining paternalism.

Chapter 1: an examination of those more minimalist, straightforward definitions of paternalism

Introduction

In the introduction to this thesis, we looked at three uncontroversial, general features of paternalism. We said that an accurate definition of paternalism must reflect each of these features, and so an accurate definition must at least meet the following three Challenges:<sup>17</sup>

Challenge 1

To reflect that paternalism necessarily involves a benevolent motive on X's part regarding Y.

Challenge 2

To reflect that paternalism involves X aiming to interfere with aspects of Y's life that are voluntary. And that as such X does not paternalise Y in cases like BRIDGE or ONCOMING TRAFFIC.

Challenge 3

To reflect that X does not paternalise Y (even when all other conditions are met) when he merely aims to engage Y in rational persuasion.

What's more, we said that throughout this thesis, as we learn more about the nature of paternalism, we will develop new Challenges. Indeed, by the end of this chapter we will have isolated seven *Challenges* an accurate definition of paternalism must meet. These seven Challenges will guide the rest of our investigation of paternalism.

<sup>17</sup>As in the introduction 'X' will stand for the paternaliser/would-be paternaliser. While 'Y' will stand for the paternalisee/would-be paternalisee

But identifying new Challenges is not the central aim of this chapter. Rather, I want to take a closer look at some of the more minimalist, more straightforward, perhaps more commonsensical ways paternalism has been defined in the literature. And whether any of these more minimalist, straightforward definitions can meet the Challenges an accurate definition of paternalism must meet.

Along these lines, I divide those more minimalist, straightforward definitions of paternalism into two groups. On the one hand those definitions that conceptualise paternalist acts as *benevolent interferences that are simply contrary to Y's wishes/desires/what Y has consented to*. And on the other hand, those definitions that conceptualise paternalist acts as *benevolent interference that violate Y's liberty*.

Ultimately I suggest that both types of minimalist, straightforward definition are problematic, and that therefore we should turn to examining some of the more complicated, more nuanced definitions of paternalism in the literature.

Part 1: Minimalist definitions that emphasise paternalist interferences as *benevolent acts that are contrary to Y's wishes/desires/what Y has consented to*

Let's start by taking a look at those more minimalist definitions in the literature that conceptualise acts of paternalism as benevolently motivated acts that are contrary to Y's wishes/desires/what Y has consented to. In particular I want to look at the definitions of paternalism offered by Peter Suber, Dan Brock, Donald VanDeVeer, and finally Paul Hershey.

As we saw in the introduction to this thesis, it is uncontroversial enough to suppose that paternalism necessarily involves a benevolent motive on X's part. But what about the notion that the only additional aspect to a paternalist act, is that X acts in a way that is contrary to Y's wishes/desires/what Y has consented to? Can this general idea be cashed-out so as to produce a plausible definition of paternalism?

I highlight two potential problems for this kind of minimalist definition of paternalism. These two problems will lead me to introduce two new challenges; Challenge 4 and Challenge 5. But while I argue that this type of minimalist definition may be able to meet Challenge 4, I submit that Challenge 5 presents a more enduring problem.

#### *i) Challenge 4*

##### *Peter Suber*

Let's start by looking at Peter Suber's definition of paternalism. Suber argues that to paternalise someone, "*is to act for the good of another person without that person's consent*".<sup>18</sup> This definition adequately deals with at least some of the cases of paternalism we have looked at in this thesis. For instance, it explains why X paternalises Y in cases like SUICIDE, CAKE and BLOOD TRANSFUSION. For in each of these cases, X benevolently acts without Y's consent; in SUICIDE X stops Y from jumping on the train tracks without Y's consent. In CAKE X has not obtained Y's consent when he throws away Y's cake. While in BLOOD TRANSFUSION Y is incapacitated and so cannot consent to X's administering a blood transfusion, nor has Y given his consent in the past.

---

<sup>18</sup>Suber (1999), my emphasis

Yet a closer look at Suber's definition reveals it to be too broad. Consider the following case we might call BEE.

### *BEE*

*X and Y are in the park. A bee flies on Y's back with the seeming intent to sting Y. Y does not seem to notice the bee, but X does. Assuming that Y would not want to be stung if he knew there was a bee around, X decides to push Y to one side and swat away the bee. X acts without Y's consent for he believes there is no time to waste.*

I take it as clear that X does not paternalise Y in BEE. We will explore a precise answer to this question later in this thesis. But to pre-empt our answer a little, the answer relates to how X is acting in a way he believes Y approves of, or at least would approve of if he was relevantly informed. As such, X is not trying to paternalistically ride roughshod over Y's judgement regarding his own interests. Or paternalistically substitute out Y's judgement or wishes regarding what is in Y's best interests, and substitute in his own. Rather, X has taken into consideration what Y's relevantly informed wishes are/would be, and is ensuring this guides his interaction with Y. In this sense, BEE is relevantly similar to CIGARETTES.

### *CIGARETTES*

*Y wants to finally quit smoking. To this end, Y asks his friend X to come to his house, and throw away all his cigarettes. X is happy to oblige as he thinks quitting smoking will do Y the world of good.*

The problem is that Suber's definition fails to reflect this point about approval. The fact that

X has thought about what Y's wishes are or would be if he was relevantly informed, and believes his actions to be in line with these wishes, is not a relevant consideration on his account. Rather, all that matters is that benevolently motivated X acts without Y's consent. As such, Suber's account delivers the (implausible) verdict that X paternalises Y in BEE and CIGARETTES.

This point about how X does not paternalise Y in cases like BEE and CIGARETTES can be expressed in the form of a Challenge. A Challenge that Suber's definition fails to meet.

#### Challenge 4

To reflect that X does not paternalise Y in cases like BEE and CIGARETTES where X has considered what relevantly informed Y (would) approve of, and merely aims to act in a way he believes relevantly informed Y does/would approve of.

#### *Dan Brock*

Perhaps in light of the problems we have seen with Suber's account, it is best to drop talk of consent. Along these lines, Dan Brock defines paternalism as follows:

“paternalism is action by one person [X] for another's [Y's] good, but *contrary to their [Y's] present wishes or desires*”<sup>19</sup>

But it is unclear whether Brock's account, and in particular his focus on Y's *present* wishes and desires can reflect that X does not paternalise Y in BEE. After all, X may well think that Y will approve once he finds out why X is pushing him aside. But at the time X pushes Y

<sup>19</sup>Brock, (1983, p238) my emphasis

aside, Y presumably objects to X's behaviour. That is, X acts in a way that is contrary to Y's present wishes or desires.

*Donald VanDeVeer*

It is with this kind of problem in mind that Donald VanDeVeer offers his definition of paternalism. According to VanDeVeer, X paternalises Y IFF:

“1. *X deliberately does (or omits) his action*

2. *X believes that his doing (or omitting) his action is contrary to Y's operative preference, intention or disposition at the time X does (or omits) his actions (or when X's actions affects Y)*

3. *X acts (or omits to act a certain way) with the primary or sole aim of promoting a benefit for Y [a benefit which, X believes, would not accrue to Y in the absence of X's doing (or omitting) his action, or preventing a harm to Y, a harm which, X believes, would accrue to Y in the absence of X's doing (or omitting) X”* <sup>20</sup>

What is particularly interesting about VanDeVeer's definition, is the way it asks us to look not only at whether X acts contrary to Y's current wishes/preferences, but also whether X acts contrary to Y's *disposition*. Unfortunately, despite this, VanDeVeer's definition still incorrectly labels BEE as a case of paternalism. For VanDeVeer's second condition states that X must act contrary to Y's operative preferences *or* Y's disposition at the time. Such a disjunctive condition is met, even when only one of its disjuncts is satisfied. And as X acts contrary to Y's operative preferences in BEE, X's actions meet this second condition, and so are (implausibly) counted as potentially paternalist according to VanDeVeer's definition.

<sup>20</sup>VanDeVeer (1986, p22), changed subjects, my emphasis

However, what if we take VanDeVeer's second condition, and focus only on Y's disposition? What if we demand simply that benevolently motivated X must act contrary to Y's *disposition* at the time, in order to paternalise Y? This modified version of VanDeVeer's second condition would explain why X does not paternalise Y in BEE. For in this case, there is an important sense in which X acts in line with Y's *disposition* at the time. For presumably Y has a disposition such that he does not want to be stung by a bee/experience needless pain. As such, X actually acts in line with Y's dispositional desires in BEE.

Yet I suggest that this account too is flawed. For consider a case where X interferes with Y's life in a way he believes Y will object to/not approve of when Y finds out, but where Y also has no relevant dispositional desires at the time of X's interference.

### *SAVINGS ACCOUNT*

*X surreptitiously sets up a savings account in Y's name. For the sake of argument we can say that Y has never considered the possibility of having a savings account and so has no relevant wishes or desires on the matter. But X does believe that were Y to be confronted with the possibility of having a savings account, he would unwisely decline.*

X's actions are clearly paternalist in this case. Yet X neither acts contrary to Y's *actual* desires/wishes/preferences at the time, nor Y's *dispositional* desires/wishes/preferences. After all, how can Y be said to have dispositional desires regarding whether to set up a savings account, if Y has never even heard of the concept? As such, our account would incorrectly discount X's actions as paternalist in SAVINGS ACCOUNT.

*Paul Hershey*

All in all, I think the best way to meet Challenge 4 while also explaining why X does not paternalise Y in SAVINGS ACCOUNT, is to emphasise the idea of paternalism involving a disregard for Y's wishes and desires on X's part. I will not offer a full explanation here,<sup>21</sup> but Paul Hershey captures the basic idea. According to Hershey, X paternalises Y just so long as:

*“(1) the paternalistic action is primarily intended to benefit the recipient [Y], and (2) the recipient's [Y's] consent or dissent is not a relevant consideration for the initiator [X]”*<sup>22</sup>

Hershey's definition correctly discounts X's actions from being paternalist in BEE and CIGARETTES. For in this case, Y's dissent *is* a relevant consideration for X; it's just that X thinks relevantly informed Y would not dissent. What is more, it also correctly labels X's actions as paternalist in SAVINGS ACCOUNT; as in this case the fact that relevantly informed Y would not consent and would dissent to X's actions, clearly is of no concern to X.

*ii) Challenge 5*

However, there is a problem with Hershey's definition. A problem that also applies to *all* those minimalist, straightforward definitions of paternalism that emphasise acts of paternalism as merely *benevolent acts that are contrary to Y's wishes/desires/what Y has consented to.*

---

<sup>21</sup>This will come in Chapter 4

<sup>22</sup>Hershey (1985)

It is common for philosophers to suppose that paternalism necessarily involves some kind of 'interference' with Y's life. A corollary of this, is that when X interacts with Y's life in a way that does not 'interfere' (whatever this means) with Y's life in any significant way, then X does not paternalise Y. Thus, consider LAWYER; a case presented by Gerald Dworkin in his highly influential 1983 essay '*Paternalism: Some Second Thoughts*':

*“Consider a father [Y] who wants his daughter [X] to become a lawyer. The daughter [X] believes that she would make a very good lawyer. Indeed, she believes it likely that she would be more successful professionally than her father [Y], who has managed to survive only on a marginal basis. Because she believes that such success would make her father [Y] very unhappy, the daughter [X] decides to become a doctor instead... Yet I think that this is not a case of paternalism.”*<sup>23</sup>

According to Dworkin's influential analysis, which I agree with, X (the daughter) does not paternalise Y (the father) in LAWYER, because X's actions do not interfere with Y in the way necessary for paternalism.<sup>24</sup> After all, when we recall the cases of paternalism we have examined in this chapter (e.g. CAKE, SUICIDE, BLOOD TRANSFUSION), X seems to interfere with Y's life in quite substantial ways; violating Y's liberty, or at least restricting Y's actions. By contrast in LAWYER, X merely seems to act contrary to Y's wishes, without really interfering with his life.

Now note that each of the more minimalist, straightforward definitions of paternalism we have seen in this chapter, license the opposite verdict on LAWYER. For these definitions

<sup>23</sup>Dworkin, (1983, p106). I have inserted Xs and Ys to make clear that it is the daughter, not the father, who is the potential paternalizer

<sup>24</sup>We will examine Dworkin's precise reasoning in the next chapter

impose no requirement that benevolently motivated X must 'interfere' with Y's life in some significant way. They require simply that X acts in a way that relevantly informed Y does not/would not like.

To recall Suber's definition of paternalism, the daughter acts without the father's consent. To recall Brock and VanDeVeer's definitions, the daughter acts contrary to father's wishes and desires (he wishes for her to go into law). To recall Hershey's definition, the daughter acts in a way that she knows the father will not like but acts in this way anyway. As such, each of these definitions (implausibly) implies that X paternalises Y in LAWYER.

This problem would be somewhat minimised if LAWYER was a one-off, anomalous case. Yet there are many others that share the same basic structure; cases where benevolently motivated X acts in a way relevantly informed Y does not like, but where X's actions do not paternalise Y because (in some rough and ready sense) they do not interfere with him. Take for example, POSITIVE THOUGHTS.

### *POSITIVE THOUGHTS*

*Y has an interview for a big promotion later today. In order to give Y a little bit of extra luck, X decides to 'send out good thoughts to the universe on Y's behalf' and maybe say a little prayer for Y, even though X knows Y has a preference that X not do so.<sup>25</sup>*

Or consider the following case presented by Simon Clarke:<sup>26</sup>

### *ROSES*

*X grows roses in his garden, close to where his neighbour's garden starts. X does so in an*

<sup>25</sup>This is a modified version of a case suggested to me by Dr James Wilson in private discussion, 2015

<sup>26</sup>Clarke (2002, p90)

*attempt to benefit Y. After all, it nice to be able to look out of one's window and see beautiful roses. X knows Y has a preference that he not do so.*

In all these cases X's actions seem to be discounted by how, whatever else is happening, X does not aim to interfere with Y's life in any significant way. Along these lines, let us say that Challenge 5 that an accurate definition must meet (a Challenge that those more minimalist definitions of paternalism that characterise paternalist acts as benevolent acts by X that relevantly informed Y does not like, fail to meet), is as follows:

#### Challenge 5

To reflect that, in some rough and ready way in need of clarification, paternalism requires that X aims to 'interfere with Y's life in some significant way'. As such, X does not paternalise Y in cases like ROSES, POSITIVE THOUGHTS, and LAWYER, where X does not aim to interfere with Y's life in a significant way.

#### Part 2: Minimalist definitions that emphasise paternalist interferences as benevolently motivated coercive, or liberty-limiting actions

There is, however, a different kind of intuitive, minimalist definition of paternalism. Specifically, those definitions that characterise paternalist interferences as *benevolent violations of Y's liberty or benevolent rights-violations*.

*Arneson, Pope, Valdés*

Consider, for instance, Richard Arneson's definition of paternalist acts as: "restrictions on a

person's (Y's) liberty which are justified exclusively by consideration for that person's (Y's) own good or welfare, and which are carried out either against his present will (when his present will is not explicitly overridden by his own prior commitment) or against his prior commitment".<sup>27</sup> Or consider the definition of paternalism offered by Thaddeus Pope, that restricts paternalist acts to acts where benevolently motivated X intentionally limits Y's liberty.<sup>28</sup> While Ernesto Garzón Valdés<sup>29</sup> characterises paternalist interferences as benevolently motivated 'coercive interventions'.<sup>30</sup>

These accounts of paternalism seem to deal well enough with the kinds of cases of paternalism we examined in the introduction to this thesis (i.e. CAKE, SUICIDE, and BLOOD TRANSFUSION). For in each of these cases benevolently motivated X violates Y's liberty; in CAKE X violates Y's liberty by destroying his property (Y's cake), in BLOOD TRANSFUSION, X coerces Y's by interfering with unconscious Y's body without his consent, while in SUICIDE X coerces Y by forcibly preventing Y from jumping on the train tracks.

But what about Challenge 5? Can this second type of minimalist definition of paternalism explain why in cases like LAWYER, ROSES and POSITIVE THOUGHTS, X's actions do not seem to interfere with Y's life in the kind of way required for paternalism? The answer is yes. For this second type of minimalist definition makes it a necessary condition of paternalism that X interferes with Y's life qua violating Y's liberty. And in each of the three cases mentioned in Challenge 5, X does not violate Y's liberty. After all, in LAWYER Y (the father) may prefer that X (the daughter) go into law, but her not going into law hardly

<sup>27</sup>Arneson (1980, p471). Again I have added Xs and Ys

<sup>28</sup>Pope (2004)

<sup>29</sup>Valdés (1990)

<sup>30</sup>Or as Valdés puts it two years earlier in a 1988 essay, paternalism consists in "la intervención coactiva en el comportamiento de una persona a fin de evitar que se dañe a sí misma". "*The coercive intervention with the behaviour of a person [Y] to the end of preventing harm to that person*" (p155, my translation)

violates his liberty. Similarly, in ROSES X may provide Y with the additional option of being able to look at a bed of roses, but this does not violate Y's liberty. While in POSITIVE THOUGHTS, Y may prefer that X not pray on his behalf, but X's silent prayer does not coerce Y or undermine his liberty.

Have we already found then, the accurate definition of paternalism we are looking for? I suggest not. For I want to propose two more Challenges an accurate definition must meet. Not only will these be the final Challenges I will outline in this thesis, I argue that the second kind of minimalist, straightforward definition of paternalism we are looking at, meets neither Challenge.

Recall that according to Challenge 5, an accurate definition of paternalism must reflect that paternalism involves X aiming to *interfere* with Y's life in some kind of significant way. With this in mind, I want to dedicate Challenges 6 and 7 to reflecting all the various ways X can count as interfering with Y's life. However, I want to keep the form of interference with Y's life known as paternalist 'nudges' and 'shoves', separate from all other forms of interference. This is for the sake of clarity, and because 'nudges'/'shoves' constitute a different kind of paternalist interference with Y. Thus, let us say that Challenge 6 requires reflecting- leaving aside the issue of so called paternalist 'nudges' and 'shoves'- all the ways in which X can count as 'interfering with Y's life'.<sup>31</sup> While Challenge 7 involves reflecting how X can 'interfere with Y's life, simply by 'nudging' or 'shoving' Y.

### *i) Challenge 6*

---

<sup>31</sup>Here I keep the issue of paternalist 'nudges' and 'shoves' separate from all other types of paternalist interference with Y's life. This is because, as we will see later on in this thesis, paternalist 'nudges' and 'shoves' interfere with Y's life in a special, different kind of way

Challenge 6 immediately poses a problem for the second kind of more minimalist, straightforward definition of paternalism. For even leaving aside the issue of 'nudging'/'shoving, there are many ways in which X can interfere with Y's life such that his actions are potentially paternalist, without X violating Y's liberty. Thus, consider PILLS; a case offered by Dworkin in '*Paternalism: Some Second Thoughts*'.

### *PILLS*

*A husband, knowing his wife to be suicidal (let's say rationally suicidal so as to make it clear the husband's behaviour is hard paternalist, not soft), decides to hide his sleeping pills in order to thwart his wife's attempts to kill herself.<sup>32</sup>*

Or consider PROFESSOR, another case of Dworkin's.

### *PROFESSOR*

*A professor (X) refuses to recommend her PhD student (Y) to a certain university, because she (X) is worried the student will be out of his league, and will therefore struggle and be unhappy.<sup>33</sup>*

Finally, consider DEATHBED.

### *DEATHBED*

*Y is on his deathbed. Y asks X how his family are. X lies to Y, telling Y that Y's family are all fine, even though the truth is that they were all just murdered. X lies so as to soothe Y in his dying moments, even though X knows Y wants to know the truth.*

---

<sup>32</sup>p106

<sup>33</sup>ibid

In all three of these cases, X paternalises Y without coercing Y or interfering with Y's liberty. In PILLS X merely hides *his* pills from his wife. As Dworkin points out, this is not an infringement of Y's liberty as the concept is normally construed, as they are X's pills and he can put them wherever he wants. In PROFESSOR, X may well prevent Y from going to a certain university, but this hardly coerces Y/violates his liberty. While in DEATHBED X simply deceives Y, which does not obviously result in Y's liberty being violated.

Thus, while it still seems fair to say that paternalism involves X interfering with Y's life in some significant way (see Challenge 5), we should reject the idea that paternalism necessarily involves X violating Y's liberty. Rather, what we need is an account of paternalism that can reflect all the types of interference with Y's life through which X can paternalise Y e.g. through violating Y's liberty, through deception, through preventing Y from acting in certain ways (this list will grow larger throughout this thesis). Again, we can put this in the form of a Challenge.

### Challenge 6

Following on from Challenge 5, to reflect the wide variety of interferences with Y's life ('nudges' and 'shoves' aside) through which X can paternalise Y. e.g. through violating Y's liberty, through controlling what actions Y can do, through deception, etc.

### ii) *'nudges' and 'shoves' and Challenge 7*

What about Challenge 7? Challenge 7 involves reflecting how X can interfere with Y's life such that his actions are potentially paternalist, simply by 'nudging' or 'shoving' Y. Now the

related concepts of 'nudging' and 'shoving' were thrust into academic and public consciousness by Richard Thaler and Cass Sunstein's groundbreaking 2008 work '*Nudge: Improving Decisions about Health, Wealth, and Happiness*'. But what do we mean when we talk of 'nudging' and 'shoving'?

And as one might expect philosophers have interpreted the related two concepts in subtly different ways. As I will understand the terms, to 'nudge' Y, is to deliberately affect, through means other than rational persuasion, a choice of Y's, without closing any of Y's options or forcing any option upon Y. Or even exerting significant decisional pressure on Y's choice(s).<sup>34</sup> Something exerts non-significant decisional pressure, when whatever influence on Y's choices that thing exerts, can be potentially easily outweighed or neutralised by other factors (e.g. conflicting preferences). Thus, a typical example of a paternalist nudge might be SHOPKEEPER.

'Shoves' meanwhile also affect (through means other than rational persuasion) what choice(s) Y makes without closing Y's options or forcing any option upon Y. But they are less gentle than nudges and involve exerting significant decisional pressure. Thus, X 'shoves' rather than 'nudges' Y in PRICE HIKE.

### *SHOPKEEPER*

*X (a shopkeeper) paternalistically nudges Y (a customer), by placing the more healthy food at the front of the shop where it is easiest to access, and the unhealthy food at the back of the shop where it is a little bit more difficult to access.*

<sup>34</sup> ' To count as a mere nudge, the intervention must be easy and cheap to avoid. Nudges are not mandates. Putting the fruit at eye level counts as a nudge. Banning junk food does not". Thaler and Sunstein (2008, p9)

*PRICE HIKE*

*Shopkeeper X 'shoves' customer Y, by significantly increasing the the price of unhealthy food, in a way that does not prevent or restrict Y from buying the unhealthy food, but does make buying unhealthy food significantly less appealing for X.*

The problem, I submit, is that those minimalist accounts that make X's violating Y's liberty a necessary condition of paternalism, fail to reflect the existence of paternalist nudges and shoves. After all, nudges and shoves, by definition, do not violate Y's liberty. They do not coerce Y or even restrict Y's options. Coercion and liberty- violation completely takes away Y's choice; nudges and shoves do not.

Once again, the point can be expressed in the form of a 'Challenge'. A Challenge that those minimalist accounts of paternalism that make X's violating Y's liberty a necessary condition of paternalism, fail to meet.

Challenge 7

Finally, also following on from Challenge 5, to reflect how X can interfere with Y's life/paternalise Y through merely 'nudging' or 'shoving' Y.

Conclusion

In this opening chapter, we have examined two distinct types of a more minimalist definition of paternalism. We have assessed whether these more minimalist definitions of paternalism can meet the seven *Challenges* an accurate definition of paternalism must meet, and suggested they do not.

I want to place particular emphasis on three connected Challenges: Challenges 5, 6, and 7. Taken together these Challenges reflect that paternalism necessarily involves some kind of interference with Y's life (Challenge 5), and that this kind of interference comes in many forms e.g. through X violating Y's liberty, restricting Y's actions, deceiving Y, (Challenge 6), or simply through X nudging/shoving Y (Challenge 7). As will become clear, these three Challenges are perhaps the most difficult Challenges for an accurate definition of paternalism to meet. And will thus play a crucial role in the rest of this thesis.

Having suggested that the more minimalist, straightforward approaches to defining paternalism fail to meet the seven Challenges an accurate definition must meet, in Chapter 2 I turn to some of the more complicated, more precise, more finely tailored definitions of paternalism offered within the philosophical literature. In particular I look at three definitions; Dworkin (1983), Dworkin (2014), and Clark (2002). Each of three definitions conceptualise paternalist acts as interferences with Y's autonomous sphere, or some closely related concept. The question now is, can any of them meet the seven 'Challenges' an accurate definition must meet?

## Chapter 2: defining paternalism as a benevolent interference with Y's autonomy

### Introduction

In chapter 1 we looked at some of the more minimalist, straightforward ways of defining paternalism. Two kinds of more minimalist, straightforward definition of paternalism were identified; but ultimately it was suggested that neither is able to meet the seven '*Challenges*' than an accurate definition of paternalism must meet.

Out of all the seven Challenges, we said that Challenges 5, 6, and 7 should be taken together, and may in the end be the most important (and difficult) Challenges to meet.

Meeting these Challenges involves reflecting how paternalism necessarily involves some kind of interference with Y's life (Challenge 5), while also reflecting the various ways X can interfere with Y's life e.g. through X violating Y's liberty, restricting Y's actions, deceiving Y, (Challenge 6), or simply through X nudging/shoving Y (Challenge 7). Along these lines, it was disappointing to find that neither of the more minimalist ways of defining paternalism, was even able to meet these three Challenges.

In this second chapter, I want to look at a different, more nuanced approach to defining paternalism. In particular I want to look at three definitions of paternalism that conceptualise paternalist acts as benevolent interferences with Y's autonomy (or some closely related concept): Dworkin (1983), Dworkin (2014), and Clarke (2002). I then assess whether any of these definitions can meet the seven '*Challenges*' identified in this thesis, or at least meet Challenges 5, 6, and 7.

### Part 1: Dworkin (1983)

No author looms quite as large over the philosophical debates about how to define paternalism, as Gerald Dworkin. His 1983 essay, '*Paternalism: Some Second Thoughts*' is something of a landmark in the literature; a careful critique of several famous definitions of paternalism culminating in Dworkin presenting his own definition. According to the definition developed in this essay, paternalism should be characterised as a benevolently motivated interference with Y's 'autonomy'.<sup>35</sup> But talk of 'autonomy' can be vague. What precisely does Dworkin mean when he refers to a violation of an agent's autonomy? According to Dworkin, X can interfere with Y's autonomy in two ways:

*“either by preventing people [Y] from doing what they have decided or by interfering with the way they [Y] arrive at their decisions”*<sup>36</sup>

As I interpret Dworkin, the first kind of paternalist interference he identifies (where X interferes with Y by preventing Y from doing what Y has decided) corresponds to all types of paternalist interference other than cases of paternalist nudges/shoves, and so corresponds to Challenge 6. While the second kind of paternalist interference (where X interferes with Y by interfering with the way Y arrives at his decisions) corresponds to cases of paternalist nudges/shoves, and so corresponds to Challenge 7.

### Critique of Dworkin (1983)

I want to point to four problems with Dworkin's account.

<sup>35</sup>As Dworkin puts it, in order for someone's actions to count as paternalist, “there must be a violation of another person's autonomy (which I conceive as a distinct notion from that of liberty)” (1983, p107)

<sup>36</sup>Ibid

*i) decisions Y has actually made*

Let's start by just focusing on the first kind of paternalist interference identified by Dworkin: where X prevents Y from doing what he has decided. I suggested just above that for Dworkin, this corresponds to all types of paternalist interference, other than cases of paternalist nudges/shoves. But contra Dworkin, I submit that (leaving aside the issue of 'nudges' and 'shoves') X can in fact paternalise Y without preventing Y from doing as he has decided. For X can paternalise Y by interfering with Y's life (say by violating Y's liberty, or coercing Y) in a way he believes Y would object to and would decide against if ever given the chance, although Y may never have actually made such a decision. Thus, consider the case of PAINT.

*PAINT*

*X paints Y's house a light shade of blue while Y is out at work. X predicts that when Y gets home, he will be angry. And that had X asked Y whether he wanted his house painted blue, Y would have said 'no', for X knows Y doesn't like the colour blue. However, at no point in Y's life has Y ever sat down and considered what colour he would want your house to be painted. Y has never 'decided' that he does not want a blue house. X has therefore not prevented Y from doing anything he ever decided to do.*

Or recall SAVINGS ACCOUNT from Chapter 1.

*SAVINGS ACCOUNT*

*Y has never considered the possibility of having a savings account and so has no relevant wishes or desires on the matter. X, a friend of Y's, thinks that Y really should set up a*

*savings account, although X also believes that if confronted with the possibility of setting up a savings account, Y would unwisely decline. For this reason, X surreptitiously sets up a savings account in Y's name.*

In both cases X paternalises Y, even though X does not interfere with any decision Y has actually made. Rather, X interferes with Y's life in a significant way that he believes to be, though technically not contrary to any decision Y has made, contrary to what Y's wishes would be if Y was relevantly informed and knew about X's plans. Dworkin's definition of paternalism then, is revealed as being too narrow.

Now a defender of Dworkin might try and respond to this objection in the following way. He might concede that in PAINT and SAVINGS ACCOUNT X does not interfere with any decision Y has made. But claim that in these cases, to recall the other way X can paternalistically interfere with Y's life according to Dworkin's account, *X interferes with the way Y arrives at a decision*. After all, in both cases X forces a decision upon Y; in PAINT X forces Y to have a blue house, while in SAVINGS ACCOUNT X forces Y to have a savings account. And surely, so the argument goes, by forcing a decision upon Y in these ways, X thereby interferes with the way Y arrives at a decision.

However, this defence of Dworkin fails because it conflates a decision being forced upon Y, with Y's making a decision. For in PAINT and SAVINGS ACCOUNT, Y does not arrive at any decision. It is not Y who decides to paint the house blue, it is X. It is not Y who sets up a savings account in his own name, it is X. Just because a decision is made in Y's name or affects Y or is forced upon Y, does not mean it is Y who arrived at that decision. Otherwise in the case of an act of rape committed against Y, one would have to say that in some

sense Y arrived at the decision to engage in sex. Which of course is nonsense.

*ii) Challenges 5 and 6: what counts as something we can decide upon?*

Still, I submit that even leaving aside this issue (that is, leaving aside cases like PAINT and SAVINGS ACCOUNT where X has made no decision) there is another, more general problem with Dworkin's claim that benevolently motivated X paternalises Y when he *prevents Y from doing what he (Y) has decided*. Specifically, this part of Dworkin's account is too ambiguous.

According to Challenge 6, an accurate definition of paternalism must reflect (leaving aside the issue of 'nudging' and 'shoving') all the ways in which X can interfere with Y's life such that his actions are potentially paternalist. The problem is, it is so deeply ambiguous what kinds of things Y can, and cannot decide to do, that it is virtually impossible to assess whether Dworkin's account meets Challenge 6.

For instance, it is obvious enough that X can paternalise Y by restricting Y's actions. The case of SUICIDE where benevolently motivated X paternalises Y, by refusing to let Y throw herself on the train tracks demonstrates this point, as does a case like PILLS (a case of Dworkin's). And in Dworkin's defence, it is natural to think that X prevents Y from doing as he has decided, when he prevents Y from acting in a certain way.

### *PILLS*

*A husband (X), knowing his wife (Y) to be suicidal (let's say rationality suicidal so as to make it clear the husband's behaviour is hard paternalist, not soft), decides to hide his*

*sleeping pills in order to thwart his wife's attempts to kill herself.*<sup>37</sup>

But what about a more subtle case of paternalism? What about a case where X does not interfere with how Y can *act*, but say interferes with what Y can know? Thus, consider DEATHBED:

#### *DEATHBED*

*Y is on his deathbed. Y asks X how his family are. X lies to Y, telling Y that Y's family are all fine, even though the truth is that they were all just murdered. X lies so as to soothe Y in his dying moments, even though X knows Y wants to know the truth.*

Or what about a case where X doesn't interfere with how Y can *act*, but rather interferes with what Y can possess?

#### *FORCED KNIGHTHOOD*

*Y has been offered a knighthood by his country. X compels Y to accept the offer and so possess the title, even though X knows that Y wants to refuse the knighthood because he considers titles to be vain. X so acts because he thinks that there is much Y will gain from possessing a knighthood (e.g. a feeling of pride), and thinks Y will deeply regret refusing the knighthood in the future.*

In both these cases X paternalises Y, and so Dworkin must show (assuming that X does not interfere with how Y arrives at a decision in these cases, which he clearly does not) that X *prevents Y from doing as he has decided* in these cases. Yet it is unclear whether forcing Y to possess something he does not want to possess, prevents Y from doing what

---

<sup>37</sup>p106

he has decided. Similarly, it is unclear whether deceiving Y prevents Y from doing what he has decided.

Similarly, recall our discussion of Challenge 5, and the cases of LAWYER and POSITIVE THOUGHTS. Meeting Challenge 5 involves reflecting how X does not paternalise Y in these cases because X does not interfere with Y's life. Can Dworkin's account reflect this? Or does X count as 'preventing Y from doing as he has decided' in a case like LAWYER where X decides not to go into law as her father desires. Again, matters are unclear.

The real problem is this: part of defining paternalism is clarifying the 'interference with Y's life' that paternalism involves (nudges and shoves aside). In order to do this, we need an account of what does and does not count as an interference with Y's life. Yet instead of giving us this and clarifying the concept of interference, Dworkin simply substitutes talk of 'interference' for talk of the hardly more precise notion of '*X preventing Y from doing what he has decided*'. At this point we might hope that Dworkin will clarify the concepts of 'preventing' and 'doing'. But Dworkin never really does this either.

### *iii) autonomy and nudges*

I now want to take a look at the second kind of paternalist interference Dworkin's 1983 account identifies. That is, how Dworkin claims benevolently motivated X paternalises Y, when *X interferes with the way Y arrives at his decisions*.

As noted above, this part of Dworkin's account is intended to explain how X can paternalise Y, even when the only way in which X interferes with Y's life is that X merely

'nudges'/'shoves' Y. But while reflecting how X can paternalise Y by merely nudging/shoving Y is important (see Challenge 7), it is difficult to square with one of Dworkin's more general claims: specifically, that all instances of paternalism involve X interfering with Y's autonomy. For Dworkin indicates that he construes the notion of autonomy in terms of the *power of self-determination*. Yet nudges and shoves, especially the former, do not obviously interfere with Y's autonomy qua self-determination.

Recall for instance, the case of SHOPKEEPER in Chapter 1. In this case X paternalistically nudges Y by placing healthy food near the front of the shop; thereby making the healthy food more visible to Y, and thus encouraging Y to buy the healthy food and improve his diet. But it seems implausibly strong to suppose that X has therefore undermined Y's power of self-determination. Indeed, by definition nudges do not close Y's options, or force an option upon Y, or even exert significant decisional pressure on Y. As such, I submit that contra Dworkin, nudges do not so much interfere with Y's self-determination. Rather they interfere with Y's ability to make decisions in a kind of decisional vacuum free of too much outside interference.

Now this might seem like a relatively minor point: that there is a discrepancy between Dworkin's definition of paternalism, and his more general claim that all cases of paternalism involve an interference with Y's autonomy qua self-determination. But one of the advantages of Dworkin's account was that it potentially furnished us with an explanation of what unites all the disparate kinds of interference with Y's life through which X can paternalise Y. Now however, it seems as though this explanation ultimately falls apart under closer scrutiny.

*iv) Challenge 3 and rational persuasion*

Fourth and finally, in the introduction to this thesis, we said that it was a widely accepted notion in the literature, that X does not paternalise Y when he merely engages Y in rational persuasion.<sup>38</sup> Indeed, we even said that Challenge 3 for an accurate definition of paternalism, was to reflect this idea. Yet, it is not so clear that Dworkin's account can meet this Challenge.

The problem is that, according to Dworkin, benevolently motivated X paternalises Y when he *prevents Y from doing as he has decided*. But benevolently motivated X can plausibly prevent Y from doing as he has decided, by merely engaging Y in rational persuasion. Consider for instance, a case where Y decides to eat a chocolate cake, but before Y is able to do so, is engaged in rational persuasion by his friend X. By informing Y of the various downsides to eating too much cake, X manages to persuade Y not to eat the cake.

In this case X has merely engaged Y in rational persuasion, and as such, should not count as paternalising Y. Yet, X has prevented Y from doing as he has decided. Y decided to eat a chocolate cake, and was prevented from doing so. And as such will implausibly qualify as paternalising Y (providing X is also benevolently motivated) on Dworkin's account.

Now I take this to be a relatively minor problem. For Dworkin could respond to this problem by simply adding an extra necessary condition in his account, that limits cases of paternalism to cases where X does not engage Y merely in rational persuasion.

---

<sup>38</sup>I take it that Dworkin would agree that X cannot paternalise Y through mere rational persuasion, and that as such we are dealing with an oversight in his account, rather than a deliberate divergence from the literature on Dworkin's part

Alternatively he might claim that he is deploying the word '*prevented*' in a special way, such that 'prevention' picks out forceful interferences with Y's decisions. Still, in its current form Dworkin's account would seem to allow that X can paternalise Y through mere rational persuasion, and is all the weaker for it.

## Part 2: Dworkin (2014)

Now more recently Dworkin has offered a modified version of his 1983 account. Does this version succeed where the other one fails? According to Dworkin's updated 2014 definition of paternalism:

*"X acts paternalistically towards Y by doing (omitting) Z:*

1. *Z (or its omission) interferes with the liberty or autonomy of Y.*
2. *X does so without the consent of Y.*
3. *X does so only because X believes Z will improve the welfare of Y (where this includes preventing his welfare from diminishing), or in some way promote the interests, values, or good of Y."*<sup>39</sup>

## Critique of Dworkin (2014)

I want to suggest that despite the modifications to his original account, Dworkin's more recent definition not only solves none of the problems associated with his original account; it even suffers from an important additional problem.

*i) Dworkin (2014) addresses none of the problems associated with Dworkin (1983)*

<sup>39</sup>Dworkin (2014)

Each of the four problems with Dworkin's 1983 account, relate to the idea that benevolently motivated X paternalises Y when he interferes with Y's *autonomy* qua X either *preventing Y from doing what Y has decided* or *interfering with the way Y arrives at his decisions*. But in Dworkin's 2014 definition, X still paternalises Y when he interferes with Y's autonomy in this way, just so long as X also acts without Y's consent.

As Y has not consented to X's actions in any of the counter-examples we evoked in our critique of Dworkin (1983) anyway, I thus suggest that Dworkin's updated 2014 account cannot respond to any of the four objections levelled against his 1983 account.

*ii) benevolent motive*

Second, in his 1983 account, Dworkin referred rather generally to the idea that paternalism necessarily involves a benevolent motive on X's part regarding Y. And indeed, this is surely right. But what we didn't get in the 1983 account, was a precise description of what it means for X to hold a benevolent motive with regards to Y. In his 2014 account, Dworkin corrects this oversight. In his updated account, he argues that in order to paternalise Y, X must act:

*“only because X believes Z [his actions] will improve the welfare of Y (where this includes preventing his welfare from diminishing), or in some way promote the interests, values, or good of Y”*

But I submit that this description of what it means to hold a benevolent motive with regards

to Y, is too narrow. The problem lies in Dworkin's assumption that in order for X to hold a benevolent motive with regards to Y, the *only* reason for X's actions must be that X believes his actions will benefit Y. Yet consider a case I will call ALCOHOL PROBLEM.

### *ALCOHOL PROBLEM*

*Y has a serious alcohol problem. X has a genuine concern for Y and wants X to stop drinking before he hurts himself one day. He also, additionally, is concerned for Y's partner, Z, whom he knows suffers when Y drinks. So motivated, X breaks into Y's house and throws away all the alcohol he can find.*

In this case, I take it that X has a benevolent motive with regards to Y. This seems to be established by how X genuinely wishes to promote Y's good, and how this seems to give him sufficient reason to act. But this does not entail that benefiting Y is actually the only reason that X acts. For X is also additionally motivated by the prospect of promoting Z's good. On Dworkin's account, this additional motivating factor discounts X from counting as benevolently motivated. But this is implausible.

Ultimately then, I suggest that not only does Dworkin's updated 2014 account fail to address the problems with his original 1983 account; in the end it may suffer from an additional problem that the original did not.

### Part 3: Clarke

Still, I want to look at one more definition of paternalism that characterises paternalism along similar lines as Dworkin; that is, as a benevolent act by X that undermines Y's self-

determination, or autonomy, or ability to act as he wishes, or some related notion. Thus, let us take a look at the definition of paternalism offered by Simon Clarke (2002).

Clarke's definition is the result of a careful examination of both a wide range of cases, and a wide range of different accounts of paternalism, including Gerald Dworkin's 1983 definition. According to Clarke, X paternalises Y IFF X:

*“1) aims to close an option that would otherwise be open to Y, or X chooses for Y in the event that Y is unable to choose for himself; and 2) to the extent that X does so in order to promote Y's good”<sup>40</sup>*

*Clarke's account then, distinguishes two ways in which X can interfere with Y's life such that his actions are potentially paternalist. X can aim to close an option that would otherwise be open to Y. Or X can choose for Y in the event that Y is unable to choose for himself. The latter refers to a relatively small class of paternalism cases where Y is in some way incapacitated and so unable to make a choice. Indeed, Clarke claims that this part of his account is motivated by the existence of paternalism cases like BLOOD TRANSFUSION<sup>41</sup>; a case we saw in the introduction to this thesis where Y is unconscious. The former therefore, refers to all other cases of paternalism. That is, cases of paternalism where X is not incapacitated.*

Unlike Dworkin, Clarke does not explicitly conceptualise paternalist interferences as benevolent interferences with Y's autonomy. But by conceptualising paternalist interferences as benevolent behaviour that limits what options are available for Y to choose (either by closing those options or choosing on Y's behalf), Clarke's account

<sup>40</sup>Clarke (2002, p89). I have also capitalised the subjects (X and Y) in Clarke's definition

<sup>41</sup>Clarke (2002, p84)

nonetheless evokes the idea of a paternalist act as a benevolent interference with Y's self-determination.

What's more, as Clarke's account does not conceptualise paternalist acts as interferences with Y's *decisions/decision-making*, instead introducing the idea of a paternalist act as an interference with Y's *options/choosing*, Clarke's account makes an interesting contrast with Dworkin's two accounts of paternalism.

### Critique of Clarke

Still, I want to take a closer, more critical look at Clarke's account. In total I highlight four issues with Clarke's account.

#### *i) Challenges 5 and 6: ambiguity in Clarke's account*

I want to start by examining how Clarke's account deals with Challenges 5 and 6; but leaving the issue of 'nudges' and 'shoves' (Challenge 7) to one side. With this in mind, meeting Challenge 5 involves reflecting how paternalism requires that X aims to interfere with Y's life in some significant way. While meeting Challenge 6 involves reflecting, nudges and shoves aside, the variety of ways through which X can interfere with Y's life in some significant way.

Now in order to assess whether Clarke's account can meet these two Challenges, it is important to get a handle on what precisely Clarke means when he talks about 'options', and how X can count as closing Y's options.

One point that Clarke makes clear enough, is that whatever counts as an *option* of Y's on his account, is closely related to what *choices* Y has. Consider, for instance, Clarke's remarks on cases where Y has no options. Clarke claims that in BLOOD TRANSFUSION, Y's having no options is explained by Y's being unconscious and hence having no ability to choose because:

*"The options of the patient are already closed in virtue of the fact that [Y] is unconscious. Y's unconsciousness prevents her from having the options to choose"* <sup>42</sup>

Furthermore, the idea that to have an option is to be able to *choose*, mirrors the emphasis on choice Clarke places on the latter kind of paternalist act; that is, in Clarke's claim that benevolently motivated X paternalises Y by *choosing for Y in the event Y is unable to choose for himself*. Finally, consider Clarke's comments on David Archard's 1990 definition of paternalism. Clarke says that the definition he offers is virtually identical to the first two conditions (there are three in total) of Archard's definition of paternalism. These two conditions are:

*"(1) X aims to bring it about that with respect to some state(s) of affairs which concerns Y's good Y's choice or opportunity to choose is denied or diminished;*

*(2) X's belief that this behaviour promotes Y's good is the main reason for Y's behaviour"* <sup>43</sup>

Archard's second condition is standard enough and not the issue here. His first condition, however, is more interesting. It characterises paternalist acts as interferences with Y's

<sup>42</sup>Clarke (2002, p84). Changed 'her' to 'Y'

<sup>43</sup>Archard (1990), my emphasis, changed subjects

choices, and taking at face value Clarke's comments about the similarity between his own account and Archard's, lends further support to the idea that to have an option closed is to have one's choice diminished in some way.

Yet, beyond this idea that to close an option of Y's is to interfere with what choices Y has available, matters are left frustratingly ambiguous. What kinds of things can Y make a choice regarding? What kinds of things do, and do not number among Y's options? In particular, the ambiguity in this part of Clarke's account makes assessing whether the account meets Challenge 6 difficult.

For instance, we have seen that, as in a case like CAKE, X can paternalise Y through violating Y's liberty. Similarly, we have seen that as in a case like PROFESSOR, X can paternalise Y through restricting how Y can act. And to Clarke's credit, it seems reasonable enough to assume that X closes an option of Y's when X violates Y's liberty or restricts how Y can act. After all, such interferences seem to interfere with Y's choice to act a certain way. But to mirror a point made earlier when discussing Dworkin (1983), what about cases like DEATHBED and FORCED KNIGHTHOOD? Does preventing Y from knowing the truth or forcing Y to possess a title he does not like, thereby 'close an option of Y's'?

This problem also muddies the waters concerning Challenge 5 (reflecting how X does not paternalise Y in cases like LAWYER, ROSES, and POSITIVE THOUGHTS where X does not interfere with Y's life in any significant way). For instance, Clarke discusses the case of LAWYER in some detail, and claims that X does not aim to close an option of Y's in this case. But doesn't Y (the father) make a certain choice in LAWYER? Specifically that X (his

daughter) should go into law. Doesn't it follow then, that X interferes with Y's choices by deciding not to go into law? Shouldn't X then, count as aiming to close an option in LAWYER, and therefore as paternalising Y on Clarke's account?

Overall then, the problem with this part of Clarke's account is that it is too ambiguous. One of the things that we want an accurate definition of paternalism to do, is shed some light on through what kind of interference with Y's life, X can and cannot paternalise Y (see Challenges 5, 6, and 7). However as we have seen, (leaving aside the issue of Challenge 7 and paternalist nudges/shoves) Clarke's account doesn't really do this. Admittedly Clarke moves past obscurely talking about paternalist acts as involving interferences with Y's life. But only to introduce the marginally less obscure notion of paternalist acts involving closures of Y's options/choices.

#### *ii) Nudges and Challenge 7*

The second issue I want to raise with Clarke's account concerns paternalist nudges. As we have seen throughout this thesis, X can paternalise Y even when the only way in which he interferes with Y's life, is by nudging Y. Indeed, we even dedicated Challenge 7 to reflecting the existence of such paternalist nudges, along with the existence of paternalist shoves. Yet, it is unclear whether Clarke's account, and his conceptualisation of paternalist acts as benevolent acts by X either *aimed at closing an option that would otherwise be open to Y, or choosing for Y in the event that Y is unable to choose for himself*, can actually reflect the existence of such paternalist nudges and meet Challenge 7.

Recall, for instance, the case of SHOPKEEPER.

*SHOPKEEPER*

*X (a shopkeeper) paternalistically nudges Y (a customer), by placing the more healthy food at the front of the shop where it is easiest to access, and the unhealthy food at the back of the shop where it is a little bit more difficult to access.*

In this case X paternalistically nudges Y. But X does not 'choose for Y in the event that Y is unable to choose for himself'. Y is not incapacitated or unable to decide for himself; and besides, Y is still able to make a choice himself.

Nor does X 'aim to close an option that would otherwise be open to Y'. As was suggested in chapter 1, by definition nudges do not forbid or close any of Y's options. Rather, nudges merely constitute an attempt to (gently) influence Y so as to make his choosing a certain option, more likely. As such, Clarke's account would seem to licence the implausible conclusion that X does not paternalise Y in a case like SHOPKEEPER, nor any other case where the only way in which X interferes with Y's life is by X merely nudging Y.

*iii) rational persuasion and Challenge 3*

The third worry I have with Clarke's account concerns the issue of rational persuasion. As we have seen throughout this thesis, an accurate definition of paternalism must reflect how X does not paternalise Y when he merely engages Y in rational persuasion (Challenge 3). However, like Dworkin's accounts, it is unclear whether Clarke can actually meet Challenge 3.

According to Clarke, benevolently motivated X paternalises Y whenever he '*closes an option that would otherwise be open to Y*'. But consider again what kinds of things number among Y's 'options'. We said earlier that this issue is somewhat ambiguous and unclear. Yet it seems plausible to suppose that among Y's options, is whether Y is subject to rational persuasion or not. It follows then, that when X engages Y in rational persuasion, he thereby closes one of Y's options: the option of not being subject to rational persuasion.

As such, Clarke's account will implausibly label X's actions as paternalist, even in those cases where benevolently motivated X merely engages Y in rational persuasion.

#### *iv) Challenge 4*

Fourth and finally, in Chapter 1 we suggested that Challenge 4 for an accurate definition of paternalism, is as follows: to reflect that X does not paternalise Y in cases like BEE and CIGARETTES, where X has thought about what Y's relevantly informed wishes are/would be if Y was relevantly informed, and believes his actions to be in line with said relevantly informed wishes.

#### *BEE*

*X and Y are in the park. A bee flies on Y's back with the seeming intent to sting Y. Y does not seem to notice the bee, but X does. Assuming that Y would not want to be stung if he knew there was a bee around, X decides to push Y to one side and swat away the bee. X acts without Y's consent for he believes there is no time to waste.*

#### *CIGARETTES*

*Y wants to finally quit smoking. To this end, Y asks his friend X to come to his house, and throw away all his cigarettes. X is happy to oblige as he thinks quitting smoking will do Y the world of good.*

We will come to the issue of precisely why X's actions seem to be discounted from being paternalist in these cases, later in this thesis. But in a rough and ready way, the essence lies in how paternalism necessarily involves X disregarding Y's judgement/wishes regarding what is in Y best interests. And how when X has properly considered what Y's relevantly informed wishes are/would be, and takes himself to be acting in line with these relevantly informed wishes, X shows no such disregard.

However, Clarke's account fails to reflect this idea (and hence fails to meet Challenge 4). As we have seen, according to Clarke, it is sufficient for X's paternalising Y, that X is benevolently motivated and aims to close an option that would otherwise be open to Y. It follows that the matter of whether X has carefully considered what Y's relevantly informed wishes are and believes himself to be acting in line with Y's relevantly informed wishes, is rendered irrelevant on Clarke's account. As such, Clarke's account misdiagnoses X's actions in cases like BEE and CIGARETTES as paternalist.

### Conclusion

Overall then, we have seen that it is problematic to define paternalism as a benevolent interference with Y's autonomous sphere (or some closely related concept). Indeed, this way of defining paternalism does not even appear to offer a satisfactory answer to the narrower question of what kind of interference with Y's life paternalist acts involve (see Challenges 5, 6, and 7).

I submit then, that having looked in Chapters 1 and 2 at some of the more conventional ways of defining paternalism and of explaining what kind of interference with Y's life paternalism involves, and having found these more conventional definitions to be inadequate, that in the next chapter we should look at a couple less conventional ways of defining paternalism. To this end, in Chapter 3 we turn our attention to those definitions of paternalism offered by Bernard Gert and Charles Culver (1976), and Jonathan Quong (2011).

### Chapter 3: some less conventional definitions of paternalism

#### Introduction

Throughout this thesis we have identified seven *Challenges* an accurate definition of paternalism must meet. Special emphasis has been placed on Challenges 5, 6, 7, which should be taken together, and may in the end be the most important and most difficult of the Challenges to meet.

Meeting these three Challenges involves reflecting how paternalism necessarily involves, in a rough and ready way in need of cashing-out, some kind of 'interference with Y's life' (Challenge 5); but that there are a wide variety of ways in which X can 'interfere with Y's life' such that his actions are potentially paternalist e.g. violating Y's liberty, restricting Y's actions, deceiving Y, controlling what Y possesses (see Challenge 6), or simply through 'nudging'/'shoving' Y (see Challenge 7).

#### Challenge 5

Following on from Challenge 5, to reflect the wide variety of interferences with Y's life ('nudges' and 'shoves' aside) through which X can paternalise Y. e.g. through violating Y's liberty, through controlling what actions Y can do, through deception, etc.

#### Challenge 7

Finally, also following on from Challenge 5, to reflect how X can interfere with Y's life/paternalise Y through merely 'nudging' or 'shoving' Y.

However, despite having spent Chapters 1 and 2 looking at some of the most popular and well regarded definitions of paternalism offered within the philosophical literature, we have neither found a definition of paternalism that can meet these seven Challenges, or even a definition that can meet Challenges 5, 6, and 7.

Along these lines, in this third chapter I examine a couple less conventional definitions of paternalism; the 1976 definition offered by Bernard Gert and Charles Culver, and the 2011 definition offered by Jonathan Quong. By taking a less conventional approach to defining paternalism, these definitions may just be able to provide us with a new way of meeting the seven Challenges an accurate definition of paternalism must meet; or at least Challenges 5, 6, and 7.

### Part 1: Gert and Culver

According to Bernard Gert and Charles Culver, X paternalises Y IFF:

*“X believes that:*

- 1. his action is for Y's good*
- 2. he is qualified to act on Y's behalf*
- 3. his action involves violating a moral rule (or doing that which will require him to do so) with regard to Y*
- 4. he is justified in acting on Y's behalf independently of Y's past, present or immediately forthcoming (free, informed) consent*
- 5. Y believes (perhaps falsely) that he (Y) generally knows what is for his own*

*good.*"<sup>44</sup>

It is Gert and Culver's third condition that sets their account apart from others within the literature. It reflects their attempt to meet what I have termed Challenges 5, 6, and 7. But instead of ruling, as other philosophers have, that paternalism necessarily involves X interfering with Y's life qua violating Y's liberty, or qua undermining Y's autonomy, or qua closing Y's options, Gert and Culver's third condition instead rules that paternalism necessarily involves X interfering with Y's life is qua X violating a 'moral rule' with regard to Y.

Yet Gert and Culver's account, and in particular their third condition, has received a lot of negative critical attention. Perhaps somewhat unfairly. Along these lines, I want to take a look at *three* objections that have been levelled against Gert and Culver's account in the philosophical literature, before explaining how I think Gert and Culver can respond to each one.

Let's start with an objection put forward by Gerald Dworkin. According to Dworkin, the problem with assuming that cases of paternalism necessarily involve X interfering with Y's life qua violating a moral rule with regard to Y, is that such an assumption is unable to reflect how X paternalises Y in a case like PILLS.

## PILLS

"A husband [X] who knows his wife [Y] is [rationally] suicidal hides his sleeping pills"<sup>45</sup>

Dworkin argues that although X clearly paternalises Y in PILLS, X does not violate a moral

<sup>44</sup>Gert and Culver (1976, p49-50), changed subjects

<sup>45</sup>Dworkin (1983, p106)

rule with regard to Y.

*“He [the husband, X] violates no moral rule. They are his pills and he can put them wherever he wishes”* <sup>46</sup>

The last sentence is key. Dworkin seems to read the concept of a 'moral rule violation' as equivalent to a rights violation. Hence, Dworkin assumes that X does not violate a moral rule with regard to Y in a case like PILLS because X does not interfere with Y's rights. (“they are his pills”). From this Dworkin concludes that Gert and Culver's account implausibly discounts X's actions from being paternalist in PILLS, and any other case where X does not violate Y's rights.

But Dworkin's objection fails because its premises are false. 'Moral rule violations' are not equivalent to rights violations. Rather, Gert and Culver make it clear that X violates a moral rule with regard to Y any time X acts in a way that requires some kind of (moral) justification. Or, to put it another way, X violates a moral rule when he acts in a prima facie morally problematic way.

*“In our opinion, violating a moral rule involves doing something that would be morally wrong unless one has an adequate justification for doing it”* <sup>47</sup>

Adding that:

*“thus, killing, causing pain (mental or physical), disabling, and depriving of freedom, opportunity, or pleasure are all violations of moral rules. The same is true of*

<sup>46</sup>Ibid

<sup>47</sup>Gert and Culver (1976, p51-52)

*deceiving, breaking a promise, and cheating”*<sup>48</sup>

Gert and Culver then, can in fact explain why X violates a moral rule with regard to Y in PILLS. For although X acts 'rightfully' (does not violate any of Y's rights) in PILLS, he nonetheless acts in a way that is in need of some kind of moral justification (a prima facie morally problematic way). After all, X acts in a way he knows will thwart Y's ends and quite possibly cause Y some degree of mental pain.

With this in mind we can now respond to a second objection levelled against Gert and Culver's account, put forward by Jonathan Quong. Unlike Dworkin, Quong seems open to the idea that X can violate a moral rule with regard to Y without violating Y's rights. In fact, Quong seems to equate moral rule violations with acts that are all things considered wrong. Along these lines Quong rejects Gert and Culver's account because it cannot explain how X can paternalise Y by acting in a “morally admirable fashion”, or otherwise acting in a way that is not all things considered wrong, as when X refuses to cooperate with Y or help Y out.

*“The moralized definition [Quong's term for Gert and Culver's definition] is also unable to explain cases [of paternalism] where one party refuses to assist another person, or refuses to cooperate with that person for paternalistic reasons. If I refuse to loan you £50 because I think you are going to spend it on heroin, I violate no moral rule, yet my refusal may nevertheless be paternalistic”*<sup>49</sup>

But again the problem with this criticism of Gert and Culver, is that it misconstrues the concept of a moral rule violation. X need not all things considered morally wrong Y, in order

<sup>48</sup>Ibid

<sup>49</sup>Quong (2011, p79)

to violate a moral rule with regard to Y. Rather, X simply needs to interact with Y in a way that stands in need of moral justification (in a way that involves a prima facie moral bad).

This explains why, contra Quong, X may in fact violate a moral rule with regard to Y simply by refusing to cooperate with Y, as refusing to cooperate with somebody who wants your cooperation is plausibly at least prima facie morally wrong. Similarly, one can act in an all things considered morally admirable fashion, while still acting in a prima facie bad way, as when one steals a loaf of bread (a prima facie moral bad) in order to save ten dying children.

Finally, I think we can now respond to a third, more practical objection levelled against Gert and Culver's account. Given that Gert and Culver claim all cases of paternalism involve X violating a moral rule with regards to Y, it might seem as though their account of paternalism is problematic, because it begs the moral question in favour of the anti-paternalist.

What precisely is the objection here? I think there are two main reasons Gert and Culver might be accused of begging the moral question in favour of the anti-paternalist. First, some philosophers have claimed that assuming that paternalism necessarily involves some morally problematic feature (i.e. a moral rule violation), entails that it is "conceptually incoherent to think that paternalism is morally permissible".<sup>50</sup> And that it is in this sense that Gert and Culver's account begs the moral question.

Yet as we have seen, to assume that paternalism involves violating a moral rule, does not entail that paternalism is necessarily all things considered wrong. Rather, it is to make the

---

<sup>50</sup>Coons and Weber (2013)

importantly weaker claim that paternalist acts always stand in need of moral justification because they are prima facie morally wrong.

Still, there is a related, more subtle objection. For perhaps the problem is that even if assuming that paternalism involves some prima facie morally wrong features does not assume that paternalism is all things considered wrong, it nonetheless concedes too much to the anti-paternalist. It is, so to speak, to stack the argument against those that think paternalism can be justified (presumably virtually all people), and stack the argument in favour of those that think paternalism can never be justified or only justified on rare occasions.<sup>51</sup>

Still, I think this objection to Gert and Culver's account is somewhat exaggerated. For I submit that for two reasons, there is no need to think that assuming paternalism to be prima facie morally wrong entails that paternalism is hard to justify.

First, as we have seen, some of the actions that qualify as 'moral rule violations' (as prima facie moral wrongs), are fairly benign and relatively easy to all things considered morally justify. As such, assuming that paternalism involves 'moral rule violations' (prima facie wrongs), should not imply that paternalism is necessarily difficult to morally justify.

Second, not only may there be lots of cases of paternalism where the relevant moral rule violation involved with paternalism may be easy to justify, paternalism qua paternalism may also furnish us with a ready-made prima facie moral justification.

After all, as we have seen throughout this thesis, paternalism necessarily involves a

---

<sup>51</sup>See Bullock (2015)

benevolent motive on X's part (see Challenge 1). And X's aiming to provide some benefit for Y, or protect Y from some harm, or otherwise promote Y's good, plausibly provides an at least prima facie moral justification for X's actions.

Of course, there is a debate to be had about how powerful the prima facie justification is that a benevolent motive provides. And how prima facie morally bad the aspects involved in paternalism are. And whether it is appropriate to think the prima facie bad(s) involved in paternalism can be outweighed by the prima facie good(s), or whether such a balancing act is invalid.<sup>52</sup> But such issues take us beyond the limits of this thesis.

### Critique of Gert and Culver

However, despite having taken some time to defend Gert and Culver's account from some of the objections in the literature that have been levelled against it, I now want to offer three of my own objections that I think are more telling.

#### *i) Challenge 5 and 'moral rule violation'*

First, as we have just seen, the concept of a 'moral rule violation', as deployed by Gert and Culver, may be broader than philosophers have generally appreciated.<sup>53</sup> Now, however, I want to suggest that it may in fact be too broad. More precisely, I submit that Gert and Culver's claim that paternalism involves X interfering with Y's life qua X violating a moral rule with regards to Y, leaves their account unable to meet Challenge 5.

Meeting Challenge 5 involves reflecting how X does *not* paternalise Y in cases like

<sup>52</sup>See Grill (2015)

<sup>53</sup>See the first two criticisms canvassed above

LAWYER and POSITIVE THOUGHTS and ROSES, because in these cases X does not aim to interfere with Y's life in any significant way. In these cases however, X may well count as violating a moral rule with regard to Y, even if he does not aim to interfere with Y's life in any significant way. Recall for instance, the case of LAWYER.

*“a father (a lawyer) who wants his daughter to become a lawyer. The daughter believes that she would make a very good lawyer. Indeed, she believes it likely that she would be more successful professionally than her father, who has managed to survive only on a marginal basis. Because she believes that such success would make her father very unhappy, the daughter decides to become a doctor instead”*<sup>54</sup>

We said earlier that, according to Gert and Culver, X violates a moral rule with regard to Y anytime X's actions impact on Y's life in such a way as requires moral justification. Thus, X can violate a moral rule with regard to Y by violating Y's rights, or by restricting how Y can act, or by deceiving Y. But less dramatically, it would seem to follow that X also violates a moral rule with regard to Y when X merely causes Y emotional pain, or when X thwarts Y's ends. The problem with this, is that X therefore plausibly counts as 'violating a moral rule' with regards to Y in LAWYER. After all, X seems to thwart Y's ends and may well cause Y emotional pain in LAWYER.

Overall then, Gert and Culver's definition of paternalism, and in particular their conceptualising paternalism as involving X interfering with Y's life qua X violating a moral rule regarding Y, would seem to render their account too broad. More precisely, it leaves their account unable to discount X's actions as paternalist in cases like LAWYER, where X does not appear to interfere with Y's life in any significant way, but does violate a moral

---

<sup>54</sup>Dworkin (1983, p106)

rule with regards to Y.

*ii) Challenge 7: moral rule violations and nudges*

Second, I submit that Gert and Culver's contention that paternalist acts necessarily involve X interfering with Y's life qua violating a moral rule with regards to Y, also renders their account too narrow.

Recall that Challenge 7 for an accurate definition of paternalism, is to reflect how one of the ways X can interfere with Y's life such that his actions are potentially paternalist, is by simply 'nudging' Y. But 'nudges' do not obviously violate a moral rule with regards to Y.

After all, nudges neither restrict nor compel how Y can act. They do not involve violating Y's liberty, and need not involve causing Y pain or distress. They do not even exert significant decisional pressure on Y; for by definition, the decisional influence a nudge exerts must be potentially easily outweighed/neutralised by other factors (e.g. conflicting desires). As such, it would seem doubtful, contra Gert and Culver, that X necessarily violates a moral rule with regards to Y when he merely nudges Y.

*iii) Challenge 4: past, present, and immediately forthcoming consent*

My third and final worry concerns Challenge 4. Meeting Challenge 4 involves reflecting how X does not paternalise Y in cases like CIGARETTES and BEE, where X has thought about what Y's relevantly informed wishes are/would be if Y was relevantly informed, and X believes his actions to be in line with Y's relevantly informed wishes.

**CIGARETTES**

*Y wants to finally quit smoking. To this end Y asks his friend X to come to his house, and throw away all his cigarettes. X is happy to oblige as he thinks quitting smoking will do Y the world of good.*

**BEE**

*X and Y are in the park. A bee flies on Y's back with the seeming intent to sting Y. Y does not seem to notice the bee, but X does. Assuming that Y would not want to be stung if he knew there was a bee around, X decides to push Y to one side and swat away the bee. X acts without Y's consent for he believes there is no time to waste.*

Why should we think X's actions are discounted from being paternalist in these cases? I have suggested, in a rough and ready way still in need of clarification, that the answer is connected to the issue of *disregard*. Paternalism, so the argument goes, necessarily involves a special kind of disregard on X's part, for Y's relevantly informed judgement/wishes regarding what is in Y's best interests. But in cases like CIGARETTES and BEE, where X has considered what Y's relevantly informed wishes are/would be, and believes himself to be acting in line with these relevantly informed wishes, X does not demonstrate this disregard. On the contrary, X shows respect for Y's relevantly informed judgements/wishes. As such, X's actions should be discounted from counting as paternalist in these cases.

Now there are a number of ways one might reflect this point in one's definition of paternalism. Gert and Culver try to reflect the point (and so meet Challenge 4) through their fourth (necessary) condition of paternalism:

*“4. [X believes] he is justified in acting on Y's behalf independently of Y's past, present or immediately forthcoming (free, informed) consent”<sup>55</sup>*

Gert and Culver's fourth condition limits cases of paternalism to cases where X believes: Y has not consented to his actions already, nor will Y consent in the immediately forthcoming future. A corollary of this condition, is that X does not paternalise Y when X believes he is acting with Y's consent, or in a way Y will consent to in the immediately forthcoming future. It follows from this, that X does not paternalise Y in CIGARETTES and BEE. After all, Y has consented to X's throwing away his cigarettes in CIGARETTES; while in BEE X seems to believe that Y will consent to his pushing Y aside in the immediately forthcoming future, once Y realises that X was only pushing him aside so as to protect Y from a bee sting.

In light of this, it might be tempting to conclude that Gert and Culver's account meets Challenge 4. Actually I think that, for two reasons, we should resist this conclusion. First, it seems suspect to suppose that an agent (Y) can retrospectively give his consent (to X's actions). Y might be able to retrospectively approve of X's actions, or forgive X. But it is unclear if these attitudes amount to consent. Thus, we have reason to be suspicious of Gert and Culver's evocation of 'immediately forthcoming consent'.

Second, I submit that this condition does not in fact meet Challenge 4 after all. Consider a case I will call COMA.

#### COMA

*Y has fallen into a sudden and permanent coma. X, a close friend of Y's, has thought long and hard about the issue, and wholeheartedly believes that Y would, if he could somehow*

<sup>55</sup>Gert and Culver (1976, p49-50), changed subjects

*communicate with X, tell him to pull the plug. Because of this, X pulls the plug. X knows that Y has never actually given his consent; and because his actions will end Y's life, knows Y will never get the chance to actually give his consent in the future.*

X's actions in COMA are relevantly similar to X's actions in CIGARETTES and BEE, and should be discounted from counting as paternalist for the same reason. X has considered what Y's relevantly informed judgements/wishes are concerning his own good, or at least what they would be if Y was relatively informed and in a position to form judgements/wishes, and believes his actions to be in line with these relevantly informed wishes. Thus, an accurate definition of paternalism must reflect how X does not paternalise Y in cases like CIGARETTES, BEE *and* COMA (Challenge 4).

Yet, Gert and Culver's fourth condition fails to discount X's actions from being paternalist in COMA. After all, Y has not consented to X's actions. Nor ever will he. As such, in COMA X must be acting in a way he believes is justifiable independent of Y's past/present/immediately forthcoming consent. And thus X must (implausibly) count as potentially paternalist according to Gert and Culver's account. Gert and Culver's account is too narrow to meet Challenge 4.

Overall then, despite its originality, and despite being more plausible than critics have generally given it credit for, Gert and Culver's account is revealed as being deeply problematic. In particular, three problems have been identified. First, Gert and Culver's conceptualisation of paternalist interferences as moral rule violations is unable to meet Challenge 5. Second, the same part of their account also renders the definition unable to meet Challenge 7. Third, Gert and Culver's account and their conceptualising paternalist

interferences as acts that X believes Y has not consented to nor will consent to in the immediately forthcoming future, is unable to meet Challenge 4.

## Part 2: Jonathan Quong

I turn now to examining the 2011 account of paternalism offered by Jonathan Quong. Like Gert and Culver, Quong defines paternalism in a somewhat unconventional way.

According to Quong's so called '*judgemental definition*', X paternalises Y IFF:

*"1. Agent X attempts to improve the welfare, good, happiness, needs, interests, or values of agent Y with regard to a particular decision or situation that Y faces.*

*2. X's act is motivated by a negative judgement about Y's ability (assuming Y has the relevant information) to make the right decision or manage the particular situation in a way that will effectively advance Y's welfare, good, happiness, needs, interests, or values."* <sup>56</sup>

What really makes Quong's account stand out is his second condition. Taken along with condition 1, the only limit it imposes on X's benevolently motivated actions counting as paternalistic, is that X is motivated by a certain *negative judgement* regarding Y.

Specifically, X must be motivated by the somewhat supercilious, negative judgement that: *without his intervention Y will fail to make the right decision/manage his particular situation in a way that will advance his own ends.*

## Critique of Quong

---

<sup>56</sup>Quong (2011, p80), changed subjects to X and Y

I want to highlight three problems with Quong's *judgemental definition*.

*i) Challenge 5*

First, I have argued that one of the most important Challenges for a definition to meet, is Challenge 5. Meeting Challenge 5 involves reflecting how X does not paternalise Y in cases like LAWYER, POSITIVE THOUGHTS and ROSES, where in some rough and ready sense in need of clarification, X does not aim to interfere with Y's life in any significant way.

With this Challenge in mind, some philosophers have argued that paternalism necessarily involves X interfering with Y's life qua violating Y's liberty. Others have emphasised the notion of paternalism involving an interference with Y's autonomy. While some have argued that paternalism involves a restriction of Y's options. And as we saw earlier in this chapter, Gert and Culver argue that paternalism involves X interfering with Y's life qua X violating a moral rule with regard to Y.

Yet, Quong's account places no such restriction on what can count as an act of paternalism. He claims that in order to paternalise Y, benevolently motivated X must merely hold a certain *negative judgement* regarding Y. As a result, Quong's account is unable to explain why X does not paternalise Y in cases like LAWYER, POSITIVE THOUGHTS and ROSES. After all, in each of these cases X thinks that Y could do with his help. X thus holds a 'negative judgement' regarding Y and a benevolent motive. And thus X (implausibly) counts as paternalising Y on Quong's account.

*ii) Challenge 1: a benevolent motive regarding Y*

My second objection revolves around Challenge 1. In many ways, Challenge 1 is the most basic of all the seven Challenges. It requires that an accurate definition of paternalism reflects how in order to paternalise Y, X must act because he holds a genuinely benevolent motive with regards to Y.

Quong attempts to meet Challenge 1 through his first (necessary) condition. According to this condition, paternalism requires that:

*“Agent X attempts to improve the welfare, good, happiness, needs, interests, or values of agent Y with regard to a particular decision or situation that Y faces”<sup>57</sup>*

In other words then, on Quong's account, a benevolent motive regarding Y, consists in X aiming to promote Y's good. Yet I think this is an implausibly broad description of a benevolent motive. Consider for instance, the following case:

***BET***

*X has placed a bet with Z, regarding Y's blood pressure. X has wagered that when the doctor records Y's blood pressure next week, Y will have low blood pressure. Z has wagered that Y will have high blood pressure. In a bid to help ensure he wins the bet, but with no real concern for Y's well being, X decided to surreptitiously interfere with Y's life in various ways so as to help lower Y's blood pressure, e.g. X replaces Y's unhealthy food with healthier alternatives.*

---

<sup>57</sup>Quong (2011, p80), changed subjects

In BET, X aims to promote Y's good, and so counts as holding a benevolent motive regarding Y according to Quong's account. But this seems implausible. Why? We will investigate precisely what constitutes a benevolent motive in Chapter 4, but to pre-empt that answer a little: X's holding a benevolent motive regarding Y, seems to require more than that X aims to promote Y's good. It seems to also require that X takes promoting Y's good as an end worthy of pursuit, apart from whatever other ends promoting Y's good might secure.

In BET, this is not the case. X takes promoting Y's good to be a valuable end, solely because promoting Y's good will enable Y to secure another end he values (winning a bet). Not because he values the end of promoting Y's good for its own sake. As such, contra Quong, X should not count as holding a benevolent motive in this case.

*iii) Challenge 3: rational persuasion and negative judgement*

My third and final objection concerns Challenge 4. According to Challenge 4, an accurate definition of paternalism must *reflect how X does not paternalise Y when he merely engages Y through rational persuasion*. Quong argues that his second condition meets this challenge. But I want to suggest that Quong's claim relies on an implausible account of the connection between, on the one hand, X's deciding to engage Y in rational persuasion, and on the other hand, X's holding a 'negative judgement' regarding Y.

According to Quong's second condition, paternalism requires that X is motivated by a 'negative judgement' regarding Y's ability to *effectively make the right decision or manage his situation*. Holding such a negative judgement, so Quong argues, is incompatible with

X's choosing to merely engage Y in rational persuasion. And hence, so the argument goes, his account explains why X does not paternalise Y when he merely engages Y in rational persuasion (Challenge 4).

But why should we think X's engaging Y in rational persuasion necessarily precludes X from holding a negative judgement regarding Y's ability to effectively make the right decision/manage his situation? Because, so Quong argues, if X has chosen to merely engage Y in rational persuasion, then X must think Y will listen to rational, reasonable argument. And if X believes Y will listen to rational, reasonable argument, then X must hold a non-negative judgement regarding Y's ability to effectively make the right decision/manage his situation. As Quong puts it, when evaluating a case of rational persuasion and why it could not be considered paternalist:

*“she [X] would be showing respect for his [Y's] capacity to make the right decision in the light of reasons and evidence, and thus no negative judgement would be implied”*<sup>58</sup>

The problem with Quong's account, is that in this instance, action and motivation are not connected in this necessary way. Just because X chooses to engage Y in mere rational persuasion, does not entail that X necessarily believes Y will listen to rational persuasion. For instance, imagine that X believes Y is highly unlikely to listen to reason and rational persuasion. Indeed, let's say that X believes the chances of rational persuasion being successful are something like 1 in 1000. Still, X might decide to engage Y in rational persuasion anyway, because he (X) is severely disabled and so has no other means of influencing Y available to him. X figures that engaging Y in rational persuasion is still better

---

<sup>58</sup>Quong (2011, p81)

than doing nothing and simply letting Y make a disastrous decision, even if he is sceptical whether it will actually work.

In fact the problem cuts both ways. Quong seems to assume that if X chooses to engage Y through means other than rational persuasion (e.g. coercion, liberty-limiting behavior, deception), then unlike the case of rational persuasion, X must hold a negative judgement regarding Y's ability to make the right decision/manage the situation. But this isn't true either.

For what if X thinks Y would likely respond to rational persuasion, but also thinks that coercing Y will work equally well, and is in fact more expedient on the grounds that it is easier and less time consuming than rational persuasion? Clearly in this case X's coercive actions would be paternalist. Yet X seems to think that rational persuasion would work. And as such, on Quong's reasoning, does not hold a negative judgement regarding Y, and thus does not paternalise Y. But this seems an implausible conclusion.

### Conclusion

What unites the accounts of paternalism developed by Charles Gert and Bernard Culver, and Jonathan Quong, is that both define paternalism in a somewhat unconventional way. Yet I have argued that these unconventional definitions fair no better at meeting the seven Challenges an accurate definition of paternalism must meet (or even just meeting Challenges 5, 6, and 7), than the conventional definitions canvassed in Chapters 1 and 2.

With the end of this chapter, comes the end of our careful analysis of the philosophical literature on defining paternalism. We have isolated seven 'Challenges' that an accurate

definition of paternalism must meet; but suggested that no account in the philosophical literature is able to meet these seven Challenges. In the next chapter- the final chapter of this thesis- I construct a new definition of paternalism that I argue meets these seven Challenges.

## Chapter 4: a new definition of paternalism

### Introduction

At least two important lessons can be drawn from the in-depth analysis of the philosophical literature on defining paternalism, conducted in the introduction and subsequent three chapters of this thesis.

First, I have suggested that an accurate definition of paternalism must meet seven '*Challenges*'. Special emphasis has been placed upon Challenges 5, 6, 7, which should be taken together, and may in the end be the most important and most difficult of the Challenges to meet. Meeting these three Challenges involves reflecting how paternalism necessarily involves some kind of interference with Y's life (Challenge 5), while also reflecting the variety of different types of interference with Y's life through which X can paternalise Y e.g. through violating Y's liberty, deceiving Y, restricting Y's actions (Challenge 6), or simply through 'nudging'/'shoving' Y (Challenge 7).

Second, I have suggested, having analysed the most popular and highly regarded definitions of paternalism in the literature, that none of these definitions is able to meet the aforementioned seven Challenges. Or even just Challenges 5, 6, and 7.

In this chapter, I take one final look at each of the seven Challenges. I outline where I believe philosophers have gone wrong in handling these Challenges. This leads me to introduce a new definition of paternalism consisting of three necessary, and together sufficient conditions, that I argue can in fact meet the seven Challenges. I end this chapter, and indeed the thesis as a whole, by briefly taking a look at some of the implications of this

new definition of paternalism.

### List of Challenges

Let's begin by reminding ourselves of the seven 'Challenges' an accurate definition of paternalism must meet.<sup>59</sup>

#### Challenge 1

To reflect that paternalism necessarily involves a benevolent motive on X's part regarding Y.

#### Challenge 2

That X does not paternalise Y in cases like BRIDGE where X interferes with Y merely because he believes Y is acting *involuntarily*, or in cases like ONCOMING TRAFFIC where X interferes with Y merely in order to *ascertain* if he is acting voluntarily or not.

#### Challenge 3

To reflect that X does not paternalise Y when he merely aims to engage Y in rational persuasion.

#### Challenge 4

To reflect that X does not paternalise Y in cases like BEE, CIGARETTES and COMA, where X has thought about what Y's relevantly informed wishes are/would be if Y was relevantly informed, and believes his actions to be in line with said relevantly informed wishes.

<sup>59</sup>As throughout the rest of this thesis, 'X' stands for the paternaliser/would-be paternaliser, while 'Y' stands for the paternalisee/would-be paternalisee

*The last three Challenges can be taken together.*

#### Challenge 5

To reflect that, in some rough and ready way in need of clarification, paternalism requires that X aim to 'interfere with Y's life in some significant way'. As such, X does not paternalise Y in cases like ROSES, POSITIVE THOUGHTS, and LAWYER, where X does not aim to interfere with Y's life in a significant way.

#### Challenge 6

Following on from Challenge 5, to reflect the wide variety of interferences with Y's life ('nudges' and 'shoves' aside) through which X can paternalise Y. e.g. through violating Y's liberty, through controlling what actions Y can do, through deception, etc.

#### Challenge 7

Finally, also following on from Challenge 5, to reflect how X can interfere with Y's life/paternalise Y through merely 'nudging' or 'shoving' Y.

Having clarified the seven Challenges an accurate definition of paternalism must meet, let us, beginning with Challenge 1, take a look at where various philosophers have gone wrong in handling these Challenges, and how best to design one's definition in order to meet these Challenges.

#### Challenge 1

Challenge 1 looks initially like one of the simpler Challenges to meet. In order to meet this Challenge, one's definition must ensure that a necessary condition of paternalism is that X is motivated by a genuinely benevolent concern for Y. But looks can be deceiving. In particular, philosophers seem to make one of two errors when trying to capture what it means for X to hold a benevolent motive with regards to Y.

First, some philosophers define what it means for X to hold a benevolent motive regarding Y, too broadly. Jonathan Quong (2011), for instance, has argued that X has a benevolent motive with regards to Y whenever X aims to “*improve the welfare, good, happiness, needs, interests, or values of agent Y*”.<sup>60</sup> While Simon Clark (2002) conceptualises a benevolent motive as equivalent to *X's acting so as to “promote Y's good”*.<sup>61</sup> The problem with each of these accounts, is that they essentially equate X's having a benevolent motive regarding Y, with X's aiming to improve/promote Y's good. But X can aim to improve/promote Y's good, I argue, without holding a benevolent motive with regards to Y. Thus, consider BET.

### *BET*

*X has placed a bet with Z, regarding Y's blood pressure. X has wagered that when the doctor records Y's blood pressure next week, Y will have low blood pressure. Z has wagered that Y will have high blood pressure. In a bid to help ensure he wins the bet, but with no real concern for Y's well-being, X decides to surreptitiously interfere with Y's life in various ways so as to help lower Y's blood pressure. e.g. X replaces Y's unhealthy food with healthier alternatives,*

<sup>60</sup>Quong (2011, p80), changed subjects

<sup>61</sup>Clarke (2002, p81)

Or consider HELPING Z:

### *HELPING Z*

*Z is a wanted outlaw. Y is a bounty collector and plans on finding and killing Z so he can claim the bounty. X is a close friend of Z's and does not want to see him killed.*

*Accordingly, X decides to pay Y a sizable amount of money to leave Z alone.*

In both BET and HELPING Z, X acts so as to promote Y's good, but without holding a benevolent motive with regard to Y. In the former case, X simply has a selfish motive and couldn't care less for Y, while in the latter case X holds a benevolent motive with regards to another agent (Z), and only tries to benefit Y in order to protect Z.

Second, other philosophers interpret what it means for X to be benevolently motivated regarding Y, too narrowly. Gerald Dworkin (2014) for instance, writes that X holds a benevolent motive with regards to Y, just so long as X acts:

*“**only** because X believes Z [his actions] will improve the welfare of Y (where this includes preventing his welfare from diminishing), or in some way promote the interests, values, or good of Y”<sup>62</sup>*

The problem with this analysis, is that it implausibly assumes that in order for X to hold a benevolent motive with regards to Y, promoting/improving Y's good must be X's **only** reason for acting. Such an account fails to reflect a case like ALCOHOL PROBLEM.

### *ALCOHOL PROBLEM*

---

<sup>62</sup>Dworkin (2014), my emphasis

*Y has a serious alcohol problem. X has a genuine concern for Y and wants X to stop drinking before he hurts himself one day. He also, additionally, is concerned for Y's partner, Z, whom he knows suffers when Y drinks. So motivated, X breaks into Y's house and throws away all the alcohol he can find.*

In ALCOHOL PROBLEM, X clearly has a benevolent motive with regards to Y. But contra Dworkin, promoting Y's good is not the *only* reason X acts. For X is also motivated by a desire to promote Z's good too.

In light of these two problems, how should we design our definition of paternalism so as meet Challenge 1? One thing that seems clear enough, is that X's holding a benevolent motive with regards to Y, involves X attempting to secure the end of promoting Y's good. With this in mind, I want to draw a distinction between X's taking an end of his (e.g. promoting Y's good) to be valuable for its own sake, and X's taking an end of his to be valuable for the sake of something else.

As I interpret these terms, X values an end of his *for its own sake*, when X takes securing that end to be valuable, irrespective of what other ends might be promoted/secured by his securing that end. By contrast, X takes an end of his to be valuable *for the sake of something else*, when the value of that end for X, depends on how securing that ends promotes/secures some other end that X values.

Thus, imagine that X in some way values the end of winning the New York marathon. X values this end *for its own sake*, just so long as X's taking this end to be valuable does not depend on how his winning the New York marathon will secure other ends he values e.g.

prize money, defeating his rival Z. By contrast, X values the end of winning the New York marathon *for the sake of something else*, when the only reason X values this end, is because securing it will secure other ends X values e.g. fame, fortune, defeating Z

Of course, this is not to say that X never values an end *for its own sake* when he believes securing that end promotes some other end that X values. Rather, X still counts as valuing the first end *for its own sake*, so long as X would value the first end even if securing that end promoted no other ends.

Why is this distinction important? Because I propose that X holds a benevolent motive with regards to Y, when he not only values the end of promoting Y's good, but also values this end for its own sake. Thus, I suggest that we can design our definition of paternalism so as to meet Challenge 1, by installing the following as a necessary condition of paternalism:

*'X is motivated by a desire to promote Y's interests, values, goods, etc, where this includes preventing these things from diminishing, and X takes promoting these things to be valuable for their own sake'*

This description of a benevolent motive explains why X does not possess a benevolent motive with regards to Y in BET and HELPING Z. For in these cases, X may aim to promote Y's good, but the value of securing this end for X is dependent on how doing so promotes another end X values: in the former case winning a bet, in the latter case promoting Z's good. As such, in both BET and HELPING Z, X values the end of promoting Y's good, but only *for some other reason* rather than *for its own sake*.

Further, this description correctly rules that X holds a benevolent motive with regards to Y in ALCOHOL PROBLEM. Admittedly, in this case X thinks that securing the end of promoting Y's good will help promote another end he values (protecting Z). But the value of promoting Y's good for X, does not depend on how securing this end promotes Z's good. Rather, X would value promoting Y's good, even if promoting Y's good did not promote Z's good.

### Challenge 2

This takes us to Challenge 2. However, rather than analysing how best to meet Challenge, I instead want to leave it for now to one side; returning to it later once we have dealt with the rest of the seven Challenges. For as I will argue, one of the conditions we develop in order to meet a different Challenge, turns out to provide an effective answer to Challenge 2.

### Challenge 3

*Challenge 3 is probably the most straightforward Challenge. It requires reflecting that X does not paternalise Y when he merely engages Y in rational persuasion. Meeting this Challenge is likewise relatively straightforward. For we can simply make the following a necessary condition of X's paternalising Y:*

*'X does not aim to engage Y merely in rational persuasion'*

Later on, I will argue that this provisional condition is actually redundant, as another

condition we develop in order to meet a different Challenge already precludes the possibility of X's paternalising Y through mere rational persuasion. Still, this condition will suffice for now as a temporary placeholder.

#### Challenge 4

If Challenge 3 was relatively simple to meet, Challenge 4 is more complicated. To recall, Challenge 4 involves reflecting that X's actions seem to be discounted from being paternalist in cases like BEE and CIGARETTES (and other relevantly similar cases like COMA) where X has thought about the issue, and believes relevantly informed Y will/would approve of his actions.

#### *BEE*

*X and Y are in the park, A bee flies on Y's back with the seeming intention to sting Y. Y does not seem to notice the bee, but X does. Assuming that Y would not want to be stung if he knew there was a bee around, X decides to push Y to one side and swat away the bee. X acts without Y's consent for he believes there is no time to waste.*

#### *CIGARETTES*

*Y wants to finally quit smoking. To this end Y asks his friend X to come to his house, and throw away all his cigarettes. X is happy to oblige as he thinks quitting smoking will do Y the world of good.*

I want to highlight just how difficult philosophers have found it too meet Challenge 4. Dan Brock (1983) argues that paternalism necessarily involves X acting in a way he takes to be

'*contrary to Y's present wishes and desires*'. But this focus on Y's *present* wishes and desires is too broad; again failing to reflect that X does not paternalise Y in BEE. After all, at the time X pushes Y in BEE, Y presumably objects to X's behaviour.

Alternatively, Peter Suber (1999) focuses not on Y's wishes and desires, but on what Y has consented to. Suber argues that to paternalise Y, '*is to act without Y's consent, for the benefit of Y*'. Yet this therefore implausibly counts BEE as a case of paternalism. After all, X does not have Y's consent, even if he does believe that Y would approve of his actions if he was relevantly informed.

Gert and Culver (1976) also focus on consent. They argue that in order to paternalise Y, X *must not only act in a way he believes Y has not consented to, but also in a way he believes Y will not consent to in the future*.<sup>63</sup> This condition explains why X does not paternalise Y in CIGARETTES, because Y has already consented to X's actions, and explains why X does not paternalise Y in BEE, because X believes that Y will consent once he realises that X was protecting him from a bee sting.

Yet Gert and Culver's condition is still too narrow, as it fails to discount X's actions from counting as paternalist in COMA.

#### COMA

*Y has fallen into a sudden and permanent coma. X, a close friend of Y's, has thought long and hard about the issue, and wholeheartedly believes that Y would, if he could somehow communicate with X, tell him to pull the plug. Because of this, X pulls the plug. X knows that Y has never actually given his consent; and because his actions will end Y's life, Y will*

<sup>63</sup>Gert and Culver actually refer to the 'immediately forthcoming future', but for our purposes, talking about just the 'future' actually makes their account more plausible

*never get the chance to actually give his consent in the future.*

COMA is relevantly similar to BEE and CIGARETTES, because X's actions are discounted from being paternalist in COMA for the same reason they are discounted from being paternalist in BEE and CIGARETTES; because X believes he is acting in a way Y approves of, or at least would approve of if he was relevantly informed and able to give his approval. But Gert and Culver's account (implausibly) rules that X paternalises Y in COMA. After all, X has acted in a way he believes Y has not consented to, nor ever will get the chance to consent to. The coma was sudden so Y never had the chance to give his consent before; and as X's actions end Y's life, he will never get this chance in the future.

So how do we go one better than these accounts and ensure our definition of paternalism meets Challenge 4? I want to suggest that paternalism necessarily involves a special kind of *disregard* on X's part: *X disregards Y's relevantly informed judgement/wishes concerning what is in his (Y's) best interests.* And that as X does not indulge in this kind of disregard in BEE, CIGARETTES, or COMA, this explains why X does not paternalise Y in these cases.

The most obvious way that X can count as engaging in this kind of *disregard of Y's relevantly informed judgement/wishes concerning what is in his (Y's) best interests*, is as follows: by considering what Y's relevantly informed judgements/wishes are (or would be if Y was relevantly informed), concerning his own best interests, and nonetheless act in a way he knows to be contrary to these relevantly informed judgements/wishes.

When we look back at the cases of paternalism we have examined in this thesis- cases

like SUICIDE, CAKE, SAVING ACCOUNT, PILLS, PROFESSOR, etc - a common thread that runs throughout all of these cases, is that X acts like this. He acts in a way he knows to be contrary to (what would be) Y's relevantly informed judgements/wishes concerning his own good. Yet by contrast, X does behave like this in BEE, CIGARETTES, or COMA. For in each of these cases X has thought about the issue, and believes he is acting in a way Y would want if he was relevantly informed.

Still, we cannot conclude just yet that X does not disregard Y's relevantly informed judgements/wishes in BEE, CIGARETTES, or COMA. For there is another way X can engage in this kind of *disregarding*, and it is as follows: by failing to properly consider what Y's relevantly informed judgements/wishes actually are or would be.<sup>64</sup> For instance, consider the case of CREDIT CARD:

#### *CREDIT CARD*

*X and Y are housemates. One day, as part of a promotional offer, a free credit card arrives in the post for Y. X sees the credit card before Y, and decides to throw it away. X has a clean conscience so to speak, because he sincerely believes that Y would not be interested in the credit card anyway, and wouldn't mind X's throwing it away. However, X is quite mistaken. And had X thought about the issue just a little bit more, he would have realised that Y would have been interested in a free credit card.*

In this case X may well believe his actions to be in line with Y's relevantly informed judgements/wishes concerning Y's own good, but he still *disregards* Y's relevantly informed judgements/wishes anyway. This is because he has failed to give the issue his proper attention. As such, he is still allowing himself to disregard and ride roughshod over

<sup>64</sup>I was made aware of this point by Dr James Wilson in private conversation (2015)

Y's relevantly informed judgements/wishes concerning his own good. Of course, this should not imply that X disregards Y's relevantly informed judgements/wishes concerning his own good, anytime X makes a mistake about what Y's relevantly informed judgements/wishes concerning Y's own good are. But at the very least, X must give the issue his proper attention and try to, so to speak, put himself in Y's shoes.

So there is a second way in which X can paternalistically disregard Y's relevantly informed judgements/wishes concerning Y's own good. But X does not engage in this kind of *disregarding* in the cases of BEE, CIGARETTES, or COMA either. After all, in each of these cases X has really tried to consider what Y's relevantly informed judgements/wishes are (or would be).

Along these lines, I submit that we have our answer to Challenge 4. Paternalism requires that X show a special kind of *disregard* regarding Y. This *disregard* can come in two forms. And as X engages in neither of these forms in BEE, CIGARETTES, or COMA, we have our explanation of why X does not paternalise Y in these cases. All that is left is for us to ensure our definition reflects this point, by installing the following as a necessary condition of paternalism

*'X disregards Y's relevantly informed judgements/wishes, in at least one of two ways:*

***Either*** X believes his actions are contrary to Y's relevantly informed judgements/wishes, or at least what Y's relevantly informed judgements/wishes would be if Y was relevantly informed and in a position to form judgements/wishes

**Or else, X has failed to engage in a minimally sufficient way with the issue of what are (or would be) Y's relevantly informed judgements/wishes'** <sup>65</sup>

Before we move on to the next Challenge, it is worth very quickly preempting a worry one might have concerning the discussion above. It might be asked: why precisely should we associate the special kind of *disregard* involved in paternalism, with the disregarding of Y's *relevantly informed* judgements/wishes, rather than just Y's judgements/wishes (as in both relevantly informed and non-relevantly informed judgements/wishes

The answer is simply that when Y's judgements/wishes concerning what is in his best interests are not relevantly informed, then in an important way those judgements/wishes are not truly *his own*. As such, X cannot be truly said to be disregarding Y's judgements/wishes, in a case where X merely disregards Y's non-relevantly informed judgements/wishes e.g. when Y's judgement is based on an ignorance of the salient facts that he is not aware of.<sup>66</sup>

### Challenges 5, 6, and 7

This leaves us with Challenges 5, 6, and 7. These last three challenges are connected and, as I have suggested, the most difficult to meet.

Challenge 5 involves reflecting how, in some sense in need of further clarification,

<sup>65</sup>Often philosophers talk about paternalism necessarily involving a special kind of *substitution*, where X substitutes out Y's relevantly informed judgements/wishes concerning what is in Y's best interests, and substitutes in his own. Thus, Brock (1988, p559) writes that paternalism "involves the substitution by the paternalistic interferer of his or her conception of what is good for another for that other's own conception of his or her good". Dworkin (1983, p107) talks of paternalism as involving the substitution of Y's "judgement". While Grill (2012) and Shiffrin (2000, p218)) both refer to paternalism as involving a substitution of Y's "judgement or agency". In this context, I take the notions of *disregard* outlined above, and this notion of substitution to be very similar, if not equivalent

<sup>66</sup>We will come shortly to the issue of what precisely it means, to be 'relevantly informed'

paternalism necessarily involves X aiming to interfere with Y's life in some significant way. Hence, X does not paternalise Y in cases like LAWYER and ROSES where no such significant interference is aimed at.

Challenges 6 and 7 then involve explaining the wide variety of ways in which X can aim to interfere with Y's life such that he may paternalise Y (when all other conditions are met). Challenge 6 involves specifying all the ways in which X can paternalistically interfere with Y's life, other than through 'nudging' and 'shoving' Y. For Challenge 7 is dedicated to reflecting how X can paternalistically interfere with Y's life through merely 'nudging' or 'shoving' Y.

### Challenge 5

In attempting to meet these last three Challenges, I want to start by providing a provisional answer to Challenge 5. That is, I suggest a provisional answer to what kind of interference with Y's life, is necessary for X to paternalise Y. However, I want to leave the issue of paternalist nudges and shoves aside for now.

With this caveat in mind, I submit that there are two ways in which X can count (nudges and shoves aside) as interfering with Y's life such that his actions are potentially paternalist. First, by restricting Y. Second, by compelling Y.

I will come shortly to specifying precisely the various ways in which X can restrict or compel Y, but a quick word on what it generally means to 'restrict' or 'compel'. I take the notion of 'restriction', to pick out cases where X acts so as to limit what Y can do in some

way. For instance, X might restrict Y by preventing Y from being able to act a certain way, say by throwing away all of Y's chocolate bars so X cannot eat chocolate tonight.

By contrast, I take the notion of 'compulsion' to pick out cases where X coerces Y or forces Y to do something. For instance, X might paternalistically compel Y's actions, by force-feeding fruits and vegetables to Y, or telling Y there is a fire that he needs to run away from.

Even in this provisional form, this condition explains why X does not paternalise Y in ROSES, POSITIVE THOUGHTS and LAWYER. For in none of these cases does X aim to *restrict or compel* Y in any way.<sup>67</sup> For instance, recall ROSES.

### ROSES

*X grows roses in his garden, close to where his neighbour's garden starts. X does so in an attempt to benefit Y. After all, it is nice to be able to look out of one's window and see beautiful roses. X knows Y has a preference that he does not do so.*

Clearly in this case X neither aims to restrict or compel Y. X does not aim to prevent Y from being able to look at the roses, nor does he force Y to look at any roses. In fact, far from restricting/compelling Y, X actually adds to what Y can do. For Y now has the additional option of being able to look out of his window and see a bed of roses if he so wishes.

Along these lines, we can contrast ROSES with, on the one hand, a paternalism case like COVERED GARDEN where X aims to restrict what Y can see, and on the other hand, a paternalism case like FORCED TO SEE where X compels Y to see something. I take it as

---

<sup>67</sup>And they are 'nudge' or 'shoves' cases

intuitive that X paternalises Y in these cases, not in ROSES. And I suggest the reason for our intuitions, is that in COVERED GARDEN and FORCED TO SEE, but not in ROSES, X tries to restrict or compel Y.

### *COVERED GARDEN*

*X and Y are neighbours. X thinks Y is a bit of a slacker, and often notices Y spending his time idly looking out his window at X's beautiful garden, instead of working hard.*

*Accordingly, X decides to cover up his garden with a large tarpaulin sheet, so now Y can no longer look at all the beautiful flowers in X's garden, and so is more inclined to do work.*

### *FORCED TO SEE*

*X is once again growing roses in his garden. X believes it would greatly benefit Y to look at his garden of roses each day, even though he knows that Y disagrees. Along these lines, X forces Y to look at his garden of roses each day, physically compelling him.*

The provisional condition also explains why X does not paternalise Y in POSITIVE THOUGHTS and LAWYER. To take the former case, Y might prefer that X not send out positive thoughts to the universe on his behalf, but sending out positive thoughts on Y's behalf does not interfere with Y's life in any significant way. It neither threatens to restrict or compel Y. At most X merely agitates Y.

Similarly, in LAWYER, X's (the daughter's) deciding not to go into law may be contrary to her father's (Y's) wishes, but it hardly restricts or compels. As such, X's deciding not to go into law is not aimed at interfering with Y's life in any way.

## Challenge 6

So as established by Challenge 5, paternalism necessarily involves some kind of interference with Y's life. Challenge 6 involves ensuring our definition reflects (nudges and shoves aside) all the ways in which X can count as interfering with Y's life. That is, it involves clarifying (nudges and shoves aside), through what kind of interferences with Y's life, X can possibly paternalise Y. e.g. how X can paternalise Y through restricting Y's actions, by violating Y's liberty, by deceiving Y.

Challenge 6 is probably the single most difficult Challenge to meet. Within the literature, attempts to meet this Challenge invariably run into at least one of the following three problems. First, some definitions offered within the literature are too narrow to meet Challenge 6; focusing too much on the more obvious cases of paternalism and overlooking those more subtle types of paternalist interference.

For instance, several philosophers- e.g. Arneson (1980) and Pope (2004) and Valdés (1990)- have suggested that (nudges and shoves aside) paternalism necessarily involves X aiming to coerce Y/violate Y's liberty. This might explain cases like SUICIDE and CAKE, but what about a case like PILLS, where X paternalises Y by refusing to tell Y where his sleeping pills are so that Y cannot use them to attempt to commit (rational) suicide? Or PROFESSOR, where a professor (X) refuses to recommend one of her PhD students (Y) to a certain university, because she (X) is worried the student will be out of his league, and will therefore be unhappy? In these cases X may restrict Y from being able to act in a certain way, but X does not coerce Y or violate his liberty.

Dworkin (1983) tries to meet Challenge 6 in a different kind of way. He argues that paternalist acts (nudges and shoves aside) involve X *preventing Y from doing as he has decided*. But this is also too broad. For as evidenced by SAVINGS ACCOUNT from Chapter 2, X can paternalise Y without interfering with a decision Y has actually made.

The second kind of problem definitions of paternalism run into when encountering Challenge 6, is being too broad. Gert and Culver (1976) for instance, argue that in order to paternalise Y, X must act in a way that requires moral justification. But as we saw in Chapter 3, this account fails to discount X's actions from being paternalist in ROSES, POSITIVE THOUGHTS and LAWYER. After all, in each of these cases X may well cause Y significant emotional pain; and such causing of pain plausibly stands in need of moral justification.

Similarly, Quong argues that all cases of paternalism involve X holding a negative judgement regarding Y's ability to make the right decision/manage his situation. Again, this description is too broad as X plausibly holds this kind of negative judgement in non-paternalism cases like ROSES, POSITIVE THOUGHTS and LAWYER.

Finally, the third reason definitions have failed to meet Challenge 6, is because they are too vague and ambiguous. Andre and Velasquez (2011) for instance, argue that paternalist acts involve X interfering with Y's 'freedom'. Yet what precisely these authors mean by 'freedom' is unclear. Similarly, Clarke (2002) suggests that all acts of paternalism involve X aiming to 'close an option' of Y's. But again, the question is, what precisely does and does not count as an 'option' of Y's?

What about the provisional condition we developed in response to Challenge 5? What about the idea that (nudges and shoves aside) paternalist acts involve X interfering with Y's life, qua X aiming to *restrict or compel* Y? The problem with this provisional condition is that, like Andre and Velasquez's, and Clarke's definitions, it is too vague. What we need to do, is to specify precisely what kinds of restrictions and compulsions are involved in paternalist interferences (nudges and shoves aside). So let us do just this.

It should be clear enough already that one of the ways X can interfere with Y's life such that his actions are potentially paternalist, is by restricting or compelling Y's actions.<sup>68</sup> For instance recall the cases of CAKE, SUICIDE and FORCED TO SEE. In each of these paternalism cases X paternalises Y by coercing him. In other words, X paternalises Y by compelling Y to act a certain way.

Similarly, recall the paternalism cases of PROFESSOR and PILLS. In these cases X paternalises Y through aiming to restrict how Y can act. In the former case, X aims to stop Y from being able to study at a certain university; while in the latter case X aims to prevent Y from taking sleeping pills and committing suicide.

Yet an exclusive focus on *actions* is too narrow, and overlooks other, more subtle interferences with Y's life (nudges and shoves aside) through which X can paternalise Y. For instance, I submit that X can also paternalise Y through restricting/compelling what Y knows (or doesn't know). Consider for instance, the following case.

### **DEATHBED**

<sup>68</sup>De Marneffe's definition of paternalism is quite similar to this *action-based* proposal. According to de Marneffe, paternalism involves X interfering with Y qua X "detering Y from choosing to perform an action or by making it more difficult for Y to perform it" (2006, p73) changed subjects ('A' changed to 'Y'). As I will argue, this exclusive focus on *actions* is too narrow

*Y is on his deathbed. Y asks X how his family are. X lies to Y, telling Y that Y's family are all fine, even though the truth is that they were all just murdered. X lies so as to soothe Y, even though X knows Y wants to know the truth.*

X clearly paternalises Y in DEATHBED. Yet X does not aim to restrict Y from acting in a certain way. After all, Y's is on his deathbed and so can't act anyway. Rather, X paternalises Y by restricting what Y can know (via deceiving him).

Furthermore, X can paternalise Y by compelling Y to know something he does not want to know. Thus, consider a case presented by Jason Hanna (2012) I will call FORCED TO KNOW.

#### *FORCED TO KNOW*

*A doctor (X) paternalises a patient (Y) by forcing Y know the relevant risks involved in his different treatment options, even though Y has made it clear that he does not want to know about these things. X interferes with Y in this way despite knowing that Y objects to such interferences, because X thinks that it is for Y's own good that he be forced to know the relevant risks.*

So (nudges and shoves aside), *X can paternalise Y through aiming to restrict or compel Y's actions, or by aiming to restrict or compel what Y knows (or doesn't know).* Do we have then, a description of all the types of interference with Y's life (nudges and shoves aside) through which X can paternalise Y?

No. For I submit that X can also paternalise Y through restricting/compelling what Y can

possess (i.e. preventing Y from being able to possess something/forcing Y to possess something). As well as through restricting/ compelling the possessions themselves (i.e. taking control of Y's possessions, altering them, modifying them).

In this thesis we have already come across a case of paternalism where X paternalises Y through merely compelling Y to possess something; SAVINGS ACCOUNT. In this case X secretly sets up a savings account in Y's name, even though X knows Y would strongly object if he ever found out. Similarly, consider a case we might call FORCED KNIGHTHOOD.

#### *FORCED KNIGHTHOOD*

*Y has been offered a knighthood by his country. X compels Y to accept the offer and so possess the title, even though X knows that Y wants to refuse the knighthood because he considers titles to be vain. X acts against Y's wishes because he thinks that there is much Y will gain from possessing a knighthood (e.g. a feeling of pride), and thinks Y will deeply regret refusing the knighthood in the future.*

X can also paternalise Y through restricting what Y can possess. Thus, consider REMOVED KNIGHTHOOD where X paternalises Y by removing Y's knighthood from him because he believes that Y has become lazy in light of the receiving of his knighthood.

In addition to being able to paternalise Y through compelling/restricting what Y possesses, I submit that X can also paternalise Y through compelling the possessions themselves.

What does it mean for X to *compel a possession of Y's*?

X can compel one of Y's possessions by damaging it, altering its form, or simply taking control of the possession itself. Thus, consider a case where X paternalises Y by remoulding a statue owned by Y, against Y's will, with the intention of making it a more beautiful sight for Y, all for Y's own good. In this case X does not seem to compel or restrict Y's actions, or what Y knows. Nor does he restrict or compel Y what Y can possess. But he does compel one of Y's possessions.

Of course, precisely what counts as something Y possesses can be a complicated issue. But one important possession of Y's, is plausibly Y's own body.<sup>69</sup> Thus, I suggest that X can paternalise Y by merely restricting or compelling Y's body. Along these lines consider AIRBORNE VACCINATION; a case where X paternalises Y, by merely compelling Y's body

#### *AIRBORNE VACCINATION*

*X knows that Y objects to receiving a certain vaccination. Still, X believes Y really should get vaccinated, as doing so will provide great health benefits. Along these lines, X surreptitiously releases an airborne vaccination, that vaccinates. This results in Y being vaccinated against his will.*

Similarly, X can paternalise Y through restricting Y's body. For instance, X might release an airborne substance that removes a vaccination that Y does not want removed.

This point about how X can paternalise Y by compelling Y's possessions, throws up certain interesting distinctions. For instance, it means that X may paternalise Y by cutting Y's hair while he is asleep (because this plausibly counts as a case where X compels a possession

<sup>69</sup>One might also suppose that Y owns his own identity. It would seem to follow then, that X would paternalise Y if he was to take control (compel) Y's identity, against Y's will, for Y's own good. For instance, in a case where X pretends to be Y, and does good deeds

of Y's; Y's hair). But that X does not paternalise Y if he cuts Z's hair. Similarly, it implies that X can paternalise Y by re-arranging Y's garden so as to beautify it without Y's permission, but that X does not paternalise Y if he beautifies a public garden Y occasionally walks through. I take these to be intuitively plausible conclusions. For while compelling (e.g. changing, altering) some possession of Z's, or some public good seems to fall short of any kind of interference with Y's life, and so cannot count as a mode of paternalist interference. By contrast compelling Y's possessions does seem to interfere with Y's life in a significant way.

So X can interfere with Y's life such that his actions are potentially paternalist, through restricting/compelling Y's *actions, what Y knows, and what Y possesses/the possessions themselves*. Are there any other types of interference with Y through which X can paternalise Y (nudges and shoves aside)? Yes, I think there are. For I submit that X can paternalise Y through merely aiming to restrict/compel what *thoughts* Y has. This might seem an unusual claim, but consider the following case.

#### *THOUGHT REMOVAL*

*X paternalises his depressed friend Y, by pressing a special button that somehow stops Y from having certain unhappy thoughts. X interferes with Y because he is concerned for his friend, although he knows that Y objects strongly to his interfering in this way.*

Again I think we have before us a mode of interference with Y's life through which X can paternalise Y, that has been entirely overlooked within the philosophical literature: thought restriction. Along similar lines, I submit that X can also paternalise Y by compelling Y to have certain thoughts. Thus consider a case we might call THOUGHT INSERTION, where

X presses a button that forces depressed Y to have certain happy thoughts.

Actually, I think this focus on *thoughts* is itself a little too narrow. For X can also paternalise Y by restricting or compelling Y's sensations. For instance, X surely paternalises Y in the case where X prevents (restricts) Y from experiencing a certain pleasurable sensation X believes to be sinful. Call this PLEASURE REMOVAL. Similarly, X can paternalise Y by compelling Y to experience certain sensations against Y's will. For instance, X might paternalise Y by somehow compelling Y to experience a certain sensation X believes to be morally praiseworthy, say a feeling of closeness to God. Call this PLEASURE INSERTION.

So X can paternalise Y by aiming to restrict/compel what thoughts Y has. And by aiming to restrict/compel what sensations Y experiences. To capture both these types of paternalist interference, let us say that X can paternalise Y by aiming to restrict or compel Y's 'mental episodes', where the term 'mental episodes' refers to both the various thoughts one might have, and to the sensations one might experience.

I suggest that we are now finally in a position to put forward a condition that can meet Challenge 6. It may initially sound like a grab-bag of cases; but on closer inspection it is a list of all the ways in which X can (nudges and shoves aside) interfere with Y's life such that his actions are potentially paternalist. Along these lines, I submit that a necessary condition of paternalism (nudges and shoves aside), is that X:

*'aims to restrict or compel at least one of the following: Y's actions, what Y's knows (or doesn't know), what Y can possess/the possessions themselves, Y's mental*

*episodes (i.e. thoughts and feelings)'*

### Challenge 7

This leaves us with just Challenge 7 to meet. Challenge 7 involves reflecting how one of the ways in which X can interfere with Y's life such that his actions are potentially paternalist, is by simply 'nudging' or 'shoving' Y.

Nudges and shoves represent a special form of paternalist interference. For rather than involving X restricting or compelling Y (i.e. X taking away Y's choices), nudges and shoves involve X influencing Y's choices. As I have interpreted the terms throughout this thesis, to 'nudge' Y, is to deliberately affect, through means other than rational persuasion, what choice(s) Y makes, without closing (or we might say, restricting) any of Y's options or forcing any option upon Y (or we might say, compelling Y). Or even exerting significant decisional pressure on Y's choices. A nudges exerts non-significant decisional pressure, when whatever influence on Y's choices a that imparts, may be potentially easily outweighed or neutralised by other factors (e.g. conflicting desires). Thus, a typical example of a paternalist nudge might be SHOPKEEPER.

'Shoves' meanwhile, also affect, through means other than rational persuasion, what choice(s) Y makes without restricting Y's options or forcing any option upon Y (compelling Y). But they are less gentle than nudges; exerting significant decisional pressure on Y's choices. Thus, X 'shoves' rather than 'nudges' Y in PRICE HIKE.

**SHOPKEEPER**

*X (a shopkeeper) paternalistically nudges Y (a customer), by placing the more healthy food at the front of the shop where it is easiest to access, and the unhealthy food at the back of the shop where it is a little bit more difficult to access.*

### **PRICE HIKE**

*Shopkeeper X 'shoves' customer Y, by significantly increasing the the price of unhealthy food, in a way that does not prevent or restrict Y from buying the unhealthy food, but does make buying unhealthy food significantly less appealing for X.*

Despite their increasing prominence in debates about paternalism and behaviour modification in general, paternalist nudges are often overlooked by those defining paternalism. Perhaps this is because they are a relatively new phenomenon, having only really received substantial critical attention since Thaler and Sunstein's 2008 work '*Nudge: Improving Decisions about Health, Wealth, and Happiness*'. Alternatively, perhaps nudges and shoves have been often overlooked because they (especially the former) constitute a more subtle, arguably less morally problematic form of paternalism. Either way, it is noteworthy just how many accounts within the philosophical literature fail to define paternalism in a way that meets Challenge 7.

For instance, in Chapter 1 we saw that Arneson (1980), Pope (2004) and Valdés (1990) each define paternalism in terms of some kind of benevolent interference with Y's liberty. These accounts fail to reflect the existence of paternalist nudges and shoves, because paternalist nudges and shoves do not undermine Y's liberty. Similarly, Clarke (2002) argues that all paternalist acts involve X 'closing Y's options'. But nudges and shoves (especially the former) do not so much 'close' Y's options, as they do merely influence how

attractive Y finds his options.

Dworkin struggles to meet Challenge 7 for a different reason. For while Dworkin's two definitions (1983 and 2014) each adequately reflect the existence of paternalist nudges and shoves, it is hard to see how Dworkin can reconcile this with his more general claim (which underpins both accounts) that all acts of paternalism undermine Y's 'autonomy', where 'autonomy' refers to Y's ability to 'self-determine'. Nudges in particular, do not undermine Y's ability to determine his own life. For by definition, nudges are so weak that X is not forced to act in any way and can easily resist whatever decisional pressure they exert. It would therefore be more accurate to say that nudges interfere with the (perhaps related concept of) Y's power to make decisions in a kind of decisional vacuum free of benevolent influence, rather than Y's self-determination.

How then should we design our definition of paternalism so as to go one step better than these accounts and meet Challenge 7? As we have seen, nudges involve X affecting what choice Y makes, without closing any of Y's options or forcing any option upon Y. In light of this, it might be tempting to say that X nudges/ shoves Y, whenever X *'aims to affect what choice Y makes, without aiming to restrict or compel Y'*.

Actually, I don't think this is quite right. Specifically, it is too broad, for it appears to allow that X can nudge or shove Y through engaging Y in rational persuasion. After all, when X engages Y in rational persuasion, he presumably aims to affect what choice Y makes without aiming to restrict or compel Y. But rational persuasion is normally understood to fall short of nudging or shoving,<sup>70</sup> and besides our definition should not allow that X can paternalise Y through mere rational persuasion (see Challenge 3). Along these lines, I

---

<sup>70</sup>See Hausman and Welch (2010)

submit that the best way to meet Challenge 7, is by stipulating that X can paternalistically interfere with Y's life, by:

*'aiming to affect (through means other than rational persuasion) what choice Y makes, without aiming to compel or restrict Y'*

### Three Points Clarified

We have very nearly arrived at our definition of paternalism. But first we need to clarify a few issues – three to be precise – that are lurking in the background.

First of all, note that our answer to Challenges 5 and 6 on the one hand, and our answer to Challenge 7 on the other hand, need to be combined to form a single necessary condition. For taken together, our answer to these Challenges tells us precisely in what ways X can interfere with Y's life, such that his actions are potentially paternalist. Our answers to these Challenges can be combined to form the following necessary condition of X's paternalising Y:

*'X aims to interfere with Y's life in at least one of two ways.*

***Either*** X aims to restrict or compel one of the following: Y's actions, what Y's knows (or doesn't know), what Y can possess/the possessions themselves, or Y's mental episodes (i.e. Y's thoughts and feelings)

***Or else,*** X aims to affect (through means other than rational persuasion) what

*choice Y makes, without aiming to compel or restrict Y'*

The second point in need of clarification concerns Challenge 2. Earlier when we reviewed Challenge 2, I did not present a solution or condition designed to meet the Challenge. This is okay, as I now want to argue that the condition we developed in order to meet Challenge 4, also ensures our condition meets Challenge 2.

To recall, Challenge 2 involves ensuring our definition discounts two kinds of cases that are normally thought of as cases of mere soft paternalism.<sup>71</sup> First, an accurate definition should discount cases like BRIDGE from being paternalist, where X merely aims to interfere with Y's *non-sufficiently voluntary* conduct (e.g. actions, decisions, desires, etc). Second, an accurate definition of paternalism should discount from being paternalist, cases like ONCOMING TRAFFIC where benevolently motivated X tries to interfere with Y merely in order to *ascertain* whether Y's conduct (e.g. actions, decisions, desires, etc) is voluntary.

### **BRIDGE**

*Y is about to step onto a bridge. Little does he know that the bridge is broken and will collapse and kill him if he steps onto it. X knows that the bridge is broken. X sees that Y is walking towards the bridge, and working on the assumption that Y is unaware that the bridge is about to collapse and does not wish to kill himself, X decides to forcibly prevent Y from stepping onto the bridge, say by tackling him to the ground before he gets to the bridge.*

### **ONCOMING TRAFFIC**

<sup>71</sup>Rather than '*hard paternalism*', otherwise known as just 'paternalism', which we are trying to define in this thesis

*Y is about to cross the road and step into oncoming traffic. X sees this, but is unsure whether Y is acting involuntarily or not. After all, Y might be attempting to deliberately end his own life, or he may simply be crossing the road absent-mindedly. X decides to forcibly prevent Y from crossing the road so he can quickly ascertain whether Y is acting voluntarily or not. If it turns out that Y is acting voluntarily, X will not stand in his way.*

Usually when philosophers attempt to meet Challenge 2, they focus on the issue of *voluntariness*. They reason that X does not paternalise Y in BRIDGE, ONCOMING TRAFFIC and other similar cases, because X does not attempt to interfere with Y's voluntary actions/judgements/conduct/etc in these cases. Thus, it is common for philosophers to claim that whatever kind of interference with Y's life it is that paternalism involves, this interference must be with Y's voluntary conduct.

However, this way of meeting Challenge 2 is more problematic than it first appears.<sup>72</sup> Consider how, as we saw above, X can paternalise Y through benevolently restricting/compelling Y's mental episodes. In cases like this, it is unclear whether X interferes with a voluntary action/judgement/etc of Y's. At least some mental episodes seem to fall short of being voluntary. For instance, the feeling of sadness Y gets when his mind wanders to a recently deceased friend. If X was to benevolently restrict Y from being able to have this mental episode, despite knowing that Y does not welcome X's interference, then X would clearly paternalise Y. Yet such an act of paternalism seems to interfere with Y's merely *non-voluntary* mental episodes.

With this problem in mind, I suggest that the (necessary) condition we developed in order to meet Challenge 4, actually presents a more plausible solution to Challenge 2. That

---

<sup>72</sup>This is yet another point made clear to me by Dr James Wilson in private conversation (2015)

condition was, as follows:

*'X disregards Y's relevantly informed judgements/wishes/desires, in at least one of two ways:*

*'Either X believes his actions are contrary to Y's relevantly informed judgements/wishes, or at least what Y's relevantly informed judgements/wishes would be if Y was relevantly informed and in a position to form judgements/wishes.*

*Or else, X has failed to engage in a minimally sufficient way with the issue of what are (or would be) Y's relevantly informed judgements/wishes'*

This condition explains why X does not paternalise Y in cases like BRIDGE where X believes Y is acting involuntarily. For when X believes his actions are merely contrary to Y's involuntary judgements/wishes, it follows that X does not believe he is acting contrary to Y's relevantly informed judgements/wishes. After all, voluntariness is surely a pre-condition of counting as relevantly informed. Thus, assuming X has engaged with the issue sufficiently, his actions will not meet the condition outlined above.

Similarly, the condition above explains why X does not paternalise Y in a case like ONCOMING TRAFFIC. For in such a case, X does not aim to interfere with Y's relevantly informed judgements/wishes. On the contrary, X is interfering precisely because he does not know what Y's relevantly informed and voluntary judgements and wishes are. Thus, assuming X has engaged minimally with the issue, X's actions will not meet the the condition outlined above, and his actions will be (correctly) discounted from being

paternalist.

Furthermore, this condition correctly allows that X can paternalise Y, through restricting/compelling Y's mental episodes. For although in such a case X may only interfere with Y's non-voluntary mental episodes, X may nonetheless believe he is acting in a way that is contrary to Y's relevantly informed judgements/wishes. After all, X might be aware that Y has certain non-voluntary mental episodes, but prefers that no one interferes with them anyway.

Just before we move on to the third point of clarification, it is worth heading off a potential objection. The objection concerns what precisely we mean when we talk of Y's being 'relevantly informed'. Along these lines, consider the following case.

### *BIG GAMBLE*

*Y has (fully voluntarily) decided to place a big bet on who will win the next world cup. X thinks the bet is a bad idea. Along these lines X goes behind Y's back and cancels the bet.*

X clearly paternalises Y in BIG GAMBLE. Yet it might be claimed that the condition we have developed in order to meet Challenges 2 and 4, implausibly discounts X's actions in this case from counting as paternalist. For in BIG GAMBLE, so the argument goes, X has engaged with what Y's relevantly informed judgements/wishes are, and has merely interfered with a judgement of Y's (to place a bet) that is made *without knowledge of certain crucial facts*. After all, Y does not know who will win the world cup. At best he is making an educated guess.

As such, so the argument concludes, Y's judgement is not relevantly informed in BIG GAMBLE. And so X's actions do not meet the demands of the condition outlined above, and are thus (implausibly) discounted from being paternalist according to my definition.

This criticism can be intensified by pointing out the following. Not only do we need to explain how Y is 'relevantly informed' in BIG GAMBLE, despite being ignorant of certain key facts. We simultaneously need to show, seemingly paradoxically, how Y does not count as relevantly informed in cases like BRIDGE where Y is similarly ignorant of certain key facts.

However, I think this objection can be replied to relatively straightforwardly. The key is to note that X need not to know all the facts, in order to count as 'relevantly informed'. Space does not permit a full explanation of what it means to be relevantly informed, but I submit that the following is a good rule of thumb: *Y's judgements/wishes count as 'relevantly informed' regarding something, just so long as Y has accurate second order beliefs concerning how informed these judgements/wishes are.*

Applied to BIG GAMBLE, this means that Y's judgement to make a bet can still count as relevantly informed even if Y is ignorant of certain facts (e.g. who will win the world cup). For Y recognises that he is ignorant of these facts, and thus qualifies as having accurate second order beliefs regarding how informed he is. This case then, can be contrasted with a case where Y makes a similar bet, but this time labours under the false second order belief that he is fully informed. Such a bet/judgement, would not count as relevantly informed.

At the same time, this condition explains why Y is not relevantly informed in BRIDGE. As in this case, Y is not only ignorant of certain facts (i.e. that the bridge is broken). Y also does not realise he is ignorant of these facts. Thus, Y has inaccurate second order beliefs concerning how informed his judgements/wishes are, and so does not count as relevantly informed.

The third and final point of clarification relates to Challenge 3. Recall that as a response to Challenge 3, we established the following as a necessary condition of X's paternalising Y:

*'X does not aim to engage Y merely in rational persuasion'*

However, note that the condition we developed in order to meet Challenges 5, 6, and 7, actually renders this condition redundant. To recall, this condition established as a necessary condition of X's paternalising Y, that:

*'Either X aims to restrict or compel one of the following: Y's actions, what Y's knows (or doesn't know), what Y can possess/the possessions themselves, or Y's mental episodes (thoughts and feelings)*

*Or else X aims to affect (through means other than rational persuasion) what choice Y makes, without aiming to compel or restrict Y'*

Each of these types of interference with Y's life, is incompatible with rational persuasion anyway. After all, X cannot restrict Y, or compel Y, or affect (through means other than rational persuasion) what choice Y makes, while still merely engaging in rational

persuasion. To engage in any of these three interferences is to go beyond mere rational persuasion.

### Our Final Definition

With these three points clarified, we are finally in a position to present our final, formalised definition of paternalism. I submit that *X paternalises Y IFF:*

1) *X is motivated by a desire to promote Y's interests, values, goods, etc, where this includes preventing these things from diminishing, and X takes promoting these things to be valuable for their own sake*

2) *X aims to interfere with Y's life in at least one of two ways:*

***Either** X aims to restrict or compel any of the following: Y's actions, what Y's knows (or doesn't know), what Y can possess/the possessions themselves, Y's mental episodes (i.e. Y's thoughts and feelings)'*

***Or else,** X aims to affect (through means other than rational persuasion) what choice Y makes, without aiming to compel or restrict Y.*

3) *X disregards Y's relevantly informed judgements/wishes in at least one of two ways:*

***Either** X believes his actions are contrary to Y's relevantly informed*

*judgements/wishes, or at least what Y's relevantly informed judgements/wishes would be if Y was relevantly informed and in a position to form judgements/wishes.*

**Or else,** *X has failed to engage in a minimally sufficient way with the issue of what are (or would be) Y's relevantly informed judgements/wishes*

### Implications of this definition

Paternalism then, on our new definition, consists in *benevolently motivated X*, aiming to *significantly interfere* with Y's life, while *disregarding Y's relevantly informed judgements/wishes*. I want to close this chapter, and the thesis as a whole, by making a few brief remarks regarding to the implications of this definition. I emphasise two points.

First of all, in several significant ways, our definition of paternalism departs from how paternalism has been traditionally defined or conceptualised.

To give an obvious example, we have seen that philosophers have often conceptualised paternalism as necessarily involving an interference with Y's liberty, or at least with how Y can act. But as our definition makes clear, X can in fact paternalise Y by merely nudging/shoving Y. Thus, in some ways, paternalism is a broader concept than has often been appreciated.

To give a more subtle example, even leaving aside the issue of nudges and shoves, there are lots of ways X can paternalise Y without restricting Y's actions or interfering with Y's liberty. For instance, our definitions makes clear that X can paternalise Y by preventing Y from learning a painful truth, or by forcing Y to know something he would rather not, or by

controlling what Y can possess. Thus, again, in many ways paternalism may be a broader, more subtle concept than has often been appreciated.

Furthermore, our definition of paternalism also makes explicit an aspect of paternalism that most definitions either leave as implicit, or else overlook entirely: how paternalism involves *X disregarding Y's relevantly informed judgements/wishes*. As such, paternalism may also be a narrower, perhaps more complicated concept than philosophers have often taken it to be.

The second implication I want to draw attention to concerns our normative evaluation of paternalism. The question of how we ought to normatively evaluate paternalism is not the subject of this particular thesis. But as we have seen, in several significant ways the definition of paternalism offered within this thesis departs from how paternalism has traditionally been understood. And this clearly has implications for what kind of moral problems (if any) paternalism poses.

For instance, we saw above how paternalism has often been conceptualised rather too narrowly; overlooking the possibility of paternalist 'nudges'/'shoves', as well as more subtle forms of interference like forcing Y to know a difficult truth, or by forcing Y to have a certain possession he would rather he didn't possess. This conceptual oversight has normative implications, because these overlooked forms of paternalist interference are among the less morally problematic forms of paternalist interference (they are at least less morally problematic than act of paternalism that involve liberty-violation/act-restriction). As such, paternalism (at least some forms of paternalism) may well be a rather more subtle, and rather less morally problematic concept, than philosophers have often supposed.

In fact, the issue cuts both ways. For another oversight common within the philosophical literature, is to overlook how paternalism involves *X disregards Y's relevantly informed judgements/wishes concerning his own good*. For several reasons such a disregard of Y's relevantly informed judgements/wishes may be interpreted as morally problematic. For instance, we might find such a disregard to be deeply disrespectful and infantilising. Or This implies that in some ways, paternalism may be even more morally problematic than philosophers have supposed.

In conclusion then, despite a wide-ranging and rich philosophical literature existing on the issue, I have suggested that no definition within the literature seems to capture precisely what it means to paternalise another agent. In response to this problem I have argued that, to express the point as simply as is consistent with being accurate, paternalism consists in genuinely benevolently motivated X, aiming to significantly interfere with Y's life, in a way that disregards Y's relevantly informed judgements/wishes.

Although this is intended to settle one debate— *what precisely defines paternalism?*— it touches upon several other important debates in philosophy. Most notably, debates concerning to what degree paternalism should be considered morally permissible. Unfortunately entering into such debates would take us beyond this particular thesis.

### Bibliography

Andre, C & Velasquez, M (1991) 'For Your Own Good', *Issues in Ethics* - V. 4, N. 2 Fall:

<http://www.scu.edu/ethics/publications/iie/v4n2/owngood.html>

Archard, D (1990) 'Paternalism Defined', *Analysis*, Vol. 50, No. 1, pp. 36-42

Arneson, R (1980) 'Mill Versus Paternalism'. *Ethics* 90 (4): pp470-489

Brock, D (1983) 'Paternalism and Promoting the Good', in Rolf Sartorius (ed.), in

*Paternalism*, ed Rolf Sartorius Minneapolis: University of Minnesota Press

Bullock, E (2015) 'A Normatively Neutral Definition of Paternalism', *Philosophical Quarterly*

65 (258):1-21

Clarke, S (2002) 'A definition of paternalism', *Critical Review of International Social and*

*Political Philosophy* 5 (1):81-91

Coons, C and M. Weber (eds.) (2013) *Paternalism: Theory and Practice*, Cambridge:

Cambridge University Press

Dworkin, G (1983) 'Paternalism: Some Second Thoughts', in *Paternalism*, ed Rolf

Sartorius Minneapolis: University of Minnesota Press

Dworkin, G (2013) 'Defining Paternalism', in '*Paternalism*' Coons and Weber (eds.) 2013,

120

pp. 25–39

Dworkin, G 'Paternalism' (2014) *The Stanford Encyclopedia of Philosophy*, (Summer 2014 Edition), Edward N. Zalta (ed.), URL = [<http://plato.stanford.edu/archives/sum2014/entries/paternalism/>](http://plato.stanford.edu/archives/sum2014/entries/paternalism/)

Dworkin, Gerald (2015) 'Defining Paternalism', in *New Perspectives on Paternalism and Health Care*. Springer International Publishing

Feinberg, J (1971) 'Legal Paternalism', *Canadian Journal of Philosophy*, Vol. 1, No. 1. 105-124

Feinberg, J (1986) *Harm to Self*, Oxford: Oxford University Press

Gert, B and Culver, C (1976) 'Paternalistic Behavior' *Philosophy and Public Affairs* 6 (1):45-57

Gert, B and Culver, C (1979) 'The Justification of Paternalism', *Ethics*, Vol. 89, No. 2, 199-210

Grill, K (2011) 'Paternalism', in Chadwick, R (ed.) *Encyclopedia of Applied Ethics*, 2d ed. Academic Press. 359-369.

Grill, K (2015) 'Antipaternalism as a Filter on Reasons'. In *New Perspectives on Paternalism and Health Care*. Springer International Publishing

Groarke, L (2002) 'Paternalism and Egregious Harm: Prader-Willi Syndrome and the Importance of Care', *Public Affairs Quarterly*, Vol. 16, No. 3 (Jul., 2002), pp. 203-230

Hanna, J (2012) 'Paternalism and the Ill-Informed Agent', *Journal of Ethics*, 16 (4):421-439.

Hausman, D and Welch, B (2010) 'Debate: To nudge or not to nudge', *Journal of Political Philosophy*, 18 (1):123-136.

Kleinig, J (1983) *'Paternalism'* Totowa, NJ: Rowman and Allanheld

de Marneffe, P (2006) 'Avoiding Paternalism', *Philosophy and Public Affairs*, 34 (1):68 - 94.

Pope, T (2004) 'Counting the Dragon's Teeth and Claws: The Definition of Hard Paternalism'. *Georgia State University Law Review* 659-722 (2004)

Quong, Jonathan (2011), *Liberalism without Perfection*. Oxford: Oxford University Press.

Scoccia, D (2008) 'In Defense of Hard Paternalism', *Law and Philosophy* 27 (4):351 - 381

Shiffrin, S (2000). 'Paternalism, unconscionability doctrine, and accommodation', *Philosophy and Public Affairs*, 29 (3):205–250.

Suber, P (1999) 'Paternalism' originally appeared in Christopher B. Gray (ed.), *Philosophy*

*of Law:An Encyclopedia*, Garland Pub. Co, II.632-635.

Thaler, R and Sunstein, C (2008) '*Nudge: Improving Decisions about Health, Wealth, and Happiness*', New Haven and London: Yale University Press

Valdés, E G (1988) '¿Es Éticamente Justificable El Paternalismo Jurídico?', *Doxa*. N. 05, ISSN 0214-8876, 155-173

Valdés, E G (1990) 'On Justifying Legal Paternalism', *Ratio Juris* 3 (s1):173-184.

VanDeVeer, D (1986) '*Paternalistic Intervention: The Moral Bounds on Benevolence*', Princeton, NJ: Princeton University Press