THE ETHICS OF USING FINANCIAL INCENTIVES TO ENCOURAGE HEALTHY BEHAVIOUR

REBECCA BROWN

SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS OF THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF LAW
QUEEN MARY, UNIVERSITY OF LONDON

MAY 2013

ABSTRACT

Efforts to encourage healthy behaviour often fail to bring about sustained changes in people's lifestyles. New approaches to tackling chronic disease include the use of financial incentives: rewards paid to individuals conditional upon their achieving some pre-specified target, such as losing weight or quitting smoking. Incentives may provide an extra motivation to adopt healthy lifestyles, and structure behaviour change efforts in ways more conducive to success.

Health incentives have, however, provoked controversy, attracting accusations of 'bribing people to be healthy,' 'rewarding bad behaviour,' and 'wasting taxpayers' money.' It remains unclear how viable health incentives could be as a tool for health promotion; but, even if effective, their contentious nature may still give reason for hesitancy.

Here, I explore whether such ethical concerns present us with convincing reasons not to use health incentives. I begin by looking at the nature of the criticisms of incentives in the media, and grouping these arguments into more general themes for discussion. I then proceed to consider each of these in turn, beginning first with debates about the requirements for the state to act efficiently without overstepping its legitimate sphere of influence. I then move on to concerns relating to the potential for incentives to undermine liberty and autonomy. Next, I discuss whether health incentives are unjust insofar as they are undeserved, and how this relates to agent freedom and responsibility for adopting healthy lifestyles. Finally, I consider the worry that using money as a healthcare intervention could corrupt certain values that we care about. In

concluding, I seek to give an overall idea as to the ethical permissibility of health incentives, and identify some key features that are likely to render incentives more or less acceptable as a means of improving health.

Contents

ACKNOWLEDGEMENTS

My doctoral studentship was funded as part of a Strategic Award in Biomedical Ethics made by the Wellcome Trust to the Centre for the Study of Incentives in Health (PI Marteau: 086031/Z/08/Z). I am grateful to the Wellcome Trust for their faith in sponsoring the Centre: without this generosity, neither my research nor any of that of CSI Health would have been possible.

I have received a great deal of help from the members of CSI Health over the past years, who, without exception, have been unfailingly obliging and insightful. Any accurate representations of health psychology in this thesis are entirely due to the expertise of Theresa Marteau, Eleni Mantzari, Marianne Promberger and Florian Voigt. Similarly, Paul Dolan, Claudia Niza, Matteo Galizzi and Caroline Rudisill have been skilled instructors in behavioural economics. I am indebted to Clive Seale, Stephen John and Martin Wilkinson for their unhesitating willingness to read garbled, developing versions of chapters, and for advice and encouragement at an early stage. Other collaborators and members of the Centre's Advisory Board have also been generous with their time and expertise, and crucial to all the work of CSI Health. Further thanks are due to the staff and students at Queen Mary, University of London, for providing a stimulating and congenial working environment.

My family and friends have been constant in their support, showing inhuman levels of patience and a remarkable readiness to discuss the minutiae of health incentives. Credit goes to Julie, Chris, Matt, Jen, Ellen, Tom, and

many others, without whom I would never have started, let alone completed, this thesis.

Finally and most particularly, my thanks go to Richard Ashcroft, who has proven to be an exceptionally wise and committed supervisor, and invariably a pleasure to know and work with.

INTRODUCTION

BEHAVIOURAL 'CHOICES' AND HEALTH

One of the biggest challenges currently faced by healthcare providers all over the world is how to combat chronic, lifestyle-related disease. According to the World Health Organization (WHO), noncommunicable diseases (NCDs) kill more people each year than all other diseases combined. The majority of these deaths are due to cardiovascular disease, cancer, diabetes, and chronic lung disease. In turn, the likelihood of suffering from NCDs is greatly increased by exposure to four behavioural risk factors: tobacco smoke, poor diet, insufficient physical activity, and excessive alcohol consumption (WHO [2011: vii]).

In developed nations such as the UK, it is something of a puzzle why the disease burden resulting from behavioural risk factors is so high. After all, it would seem that most individuals have it within their power to choose the less risky behaviours of not smoking, eating healthily, taking a bit more exercise and avoiding drinking too much alcohol. Such behaviours, common sense tells us, are surely accessible to the vast majority of UK citizens. So too is basic healthcare provision from the National Health Service (NHS). If this is the case, then why do so many continue to suffer from avoidable disease?

Part of the problem could be a failure of knowledge: people simply do not know that such behaviours are likely to harm their health. In this case, education may be needed to ensure that people understand what behaviours are likely to damage and promote their health. However, although interventions designed to promote this kind of knowledge may succeed in educating

individuals, and even in altering intentions about behaviour, they typically have a much smaller effect on actual behaviour (Marteau et al. [2012]; Webb and Sheeran [2006]; Marteau et al [1996]).

If lack of information and understanding is not the main barrier to behaviour change, then we might consider a further two possibilities. First, individuals may be well aware of the risks of unhealthy behaviours, but simply not care: information about lifestyle risk factors simply does not make people want to change their behaviour. Second, individuals may lack the ability to alter their behaviour effectively: whilst it might seem that most people have 'access' to healthy lifestyles (they can afford to eat healthier diets, are free to not smoke, to take more exercise and to drink less), in fact, their capacity to behave in these ways is undermined by psychological, social, and economic factors.

The first case is primarily a 'problem' of motivation. This is the problem as straightforward economic thinking might present it. Roughly, in order to get agents to change their behaviour (swap their unhealthy habits for healthy ones) economic thinking might suggest we need to alter agents' preferences for those behaviours we are interested in. To do this, we must make some behaviour more appealing than it previously was by altering the outcomes, and hence desirability, associated with that behaviour.

The second case is a problem of implementing the will of the agent to change her behaviour and requires that barriers to engaging with healthy behaviours be removed. These may include psychological and physical barriers, such as 'weakness of the will' and cravings which make it difficult for the agent to act consistently in line with her 'healthy' preferences. Other barriers may relate more to socio-economic, structural factors, and the environment the agent inhabits. For instance, stressful working conditions, being surrounded by individuals engaging in unhealthy behaviours, or the prospect of social stigma for seeking to alter one's lifestyle may scupper agents' efforts to change their behaviour.

From these two different problems with effecting behaviour change, we may seek one solution: in both cases, incentives may provide a way of changing the behaviour both of those who have little motivation to do so and of those who are unable to implement their healthy intentions. On the 'economic' model, a sufficiently desirable incentive can tip the balance of preferences held by the agent in favour of the healthy behaviour. On the 'psychological' model, the

incentive may help to structure the agent's approach to changing her behaviour in such a way that she is more likely to succeed.

THE CENTRE FOR THE STUDY OF INCENTIVES IN HEALTH

Health incentives may have wider application than the behavioural risk factors mentioned in the opening paragraphs. There are many instances where decisions made by the agent have a large impact on her health outcomes. Some of these are intricately entwined with one's everyday habits, repeated frequently, and integrated into one's overall lifestyle (including those risk factors relating to diet and exercise). These kinds of 'complex' behaviours are particularly resistant to change (WHO [2008]; Marteau et al. [2009a]; Dolan et al. [2012]). Other health-relevant behaviours may be more occasional, only implemented on a single or few occasions, and less familiar to the agent (such as behaviours relating to getting vaccinated or tested for disease).

My concern here is with providing an ethical account of the use of health incentives quite generally. This means that I am concerned with a whole range of behaviours that could be targeted by health incentives, both those 'lifestyle' behaviours which can have an impact on health, and decisions relevant to health that are enacted less frequently. Inevitably, pursuing this greater breadth of discussion will restrict the extent to which I am able to consider specific cases in more detail. It will, however, be necessary to examine the ethical implications of a given incentive scheme more closely prior to implementation. This thesis is intended to be a 'first look' at the ethics of incentives, and certainly not the last word. Such an account can hope only to provide some indication of the permissibility of incentives quite generally, and to serve as a framework for further consideration of specific health incentive interventions.

This analysis of the ethical issues relating to the use of incentives is situated within a wider project to consider the larger question of if and when it is right to use health incentives. This research operates under the auspices of the 'Centre for the Study of Incentives in Health' (CSI Health), and is a collaborative project with a core team of philosophers, health psychologists and behavioural economists, with wider links to researchers from other disciplines. Through the psychology and economics strands of CSI Health, research is conducted into how effective incentives are at altering behaviour; what incentives are effective; what behaviours are amenable to influence through incentives; the mechanisms

by which incentives encourage behaviour change; the 'spillover' effects of using incentives to alter behaviour; lay assessments of the acceptability of health incentives; and more.

In order to most effectively contribute to the research objectives of CSI Health, as well as to establish some necessary limitations on scope, this thesis will be restricted to considering *positive*, *personal*, *conditional*, *financial*, *health* incentives.

Positive here indicates that incentives should be true 'incentives' and not 'disincentives'; rewards rather than punishments. As will become apparent in Chapter Four, on occasion this distinction is not easy to make. For the most part, however, it is relatively straightforward to focus only on schemes which seek to reward healthy behaviours rather than those punishing unhealthy behaviours. The same focus on positive rather than negative incentives is adopted by the empirical research conducted within the other strands of CSI Health.

Personal identifies those incentives which are paid directly to the individual whose behaviour is subject to change. This rules out incentives paid to healthcare professionals for meeting targets, or to retail organisations for selling healthier products, and so on. Incentives can operate at the level of third parties, and such incentives would carry with them their own range of ethical (as well as empirical) issues. However, this thesis (and the work of CSI Health generally) is not primarily concerned with third party incentives.

By *conditional* I mean to refer to incentives that are offered prior to a behaviour change being enacted, and are given to the individual only once she has successfully altered her behaviour. The behaviour change required for the incentive to be paid should be sufficiently specific, and the conditions for determining successful behaviour change should be established when the offer of the incentive is made.[1] There are schemes which offer 'incentives' intended to encourage healthier behaviour, such as providing grocery vouchers to pregnant women in the hope of encouraging healthier eating. I will not consider the ethics of schemes like these, where the incentive is not conditional upon a behaviour change.

---

[1] Often, participants must be weighed, pass a breath test, or similar in order to show they have met the conditions of incentive payment.

The focus on *financial* incentives is rather more flexible: whilst some incentive schemes provide cash payment or other direct monetary transfers, many use gift vouchers, lottery tickets, or other kinds of rewards. However, all of the incentives considered should have some identifiable monetary value. This would rule out 'incentive' schemes where the reward takes the form of praise or winning a competition (with no distinct prize) or similar.

Finally, I am interested in *health* incentives. Once again, this still allows quite a lot of breadth in terms of the behaviours potentially included, as a diverse array of activities can have an impact on health. For instance, level of education and qualifications attained may influence occupation, income, and other factors which are known to play a role in an individual's expected health outcomes. Moreover, there are incentive schemes that target education-related behaviours such as reading and school attendance, and so which can plausibly be understood to have an impact on health. However, I am more concerned with those schemes that are implemented primarily as a means of improving health, and where the behaviour change and the health effect are more directly connected.

It is worth pointing out one further, notable exclusion: that of taxation. It has become quite common for states to impose greater taxes on unhealthy items or activities that they might wish to discourage, and similarly to reduce taxes on healthy things. Tax breaks on healthy foods, for example, could act as an incentive for buying them (saving an identifiable sum of money). Although taxation, as a means of encouraging healthy behaviour, would provide an interesting subject for ethical appraisal, I do not consider it in this thesis.

In terms of the geo-political location of this thesis, I do not set out to restrict my focus too narrowly on either the UK or the NHS case. Incentives can and have been used quite widely, targeting different populations and provided by a range of institutions (see for example Nglazi et al. [2012]; Lagarde et al. [2009]; Reilly et al. [2000]). It is true, however, that this thesis is likely to have most relevance to the use of incentives within the healthcare system of the UK. Most of the examples I use come from the UK, and the analysis of media reporting of incentives was conducted using UK media sources. This means that the issues dealt with in this thesis are directed by British concerns. This need not preclude applicability to other domains, however, where relevant similarities are shared. An exclusion, perhaps, exists where it is directly relevant

that incentives are provided by a socialised healthcare system, such as the NHS (in particular, when considering concerns relating to desert, fairness, and solidarity, for instance). At other times, criticisms and arguments may be applicable in a wider range of contexts.

Not only may the issues addressed in this thesis have a UK bias, but the approach to analysing them will be similarly coloured by my own bias as a philosopher from the UK (in terms of both citizenship and training). I can only accept this, and regret that I am ill-equipped to conduct research that would be able to better represent the diverse concerns of others (in particular, those whose concerns may go systematically under-represented in bioethical research).

The Ethical Concerns Raised by Health Incentives

An example of a health incentive would be a scheme where an individual is given £10 worth of grocery vouchers at the end of each week that she successfully meets some pre-specified weight loss target. This scheme could run for a number of weeks, the weight loss targets might be altered, and the size of the incentive might vary over the course of the scheme. In another case, a health incentive could involve a one-off payment, say, £5 in cash to an individual who returns a completed chlamydia testing kit.

In Appendix A, I provide a number of examples of incentive schemes that have been operational in the past or which are currently active. Throughout this thesis, I also use invented incentive schemes as examples to illustrate my points. Sometimes this is useful, as the 'real life' schemes are more limited in variety and often more complex than examples I would wish to use. Although I am interested in real world effects, and wish for the content of this thesis to be applicable in practical contexts, it will sometimes be helpful to work at a certain level of abstraction. Since I do not intend to draw firm conclusions about any particular schemes here, little is lost by using invented (as well as real world) examples.

Health incentives such as these are typically quite controversial, and often treated with a mixture of suspicion and hostility. Health incentives can be referred to in terms of 'paying people to be healthy,' or sometimes 'bribing people to be healthy.' Emotive language in popular media and academic treatments of incentives suggests a level of discomfort with the idea of using

money as a direct tool of healthcare intervention (see for example Shaw [2007]; The Mirror [2009]). This discomfort may stem from a range of factors, including misunderstanding as well as a lack of familiarity with health incentives, and it may seem rather spurious once considered more closely. However, such negativity may also result from more robust moral concerns with the appropriateness of using health incentives.

My approach in this thesis will be to offer a largely 'negative' account of health incentives, considering in turn each of a number of lines of criticism of incentives. My project is to see if incentives can be defended against these criticisms, and to consider the extent to which such critiques provide good reasons for thinking that incentives shouldn't be used as a method of healthcare intervention. This approach contrasts with providing a more 'positive' account which would set out to develop an account of health incentives and provide an argument in favour of their use.

I have chosen to structure the thesis in this way because I wish to treat health incentives just as another form of healthcare intervention, and to assess them as such. Language of 'paying' or 'bribing' people to be healthy, and a focus on money and non-medicalised settings, distances incentives from their role within healthcare. However, health incentives are directed quite specifically at improving health. Just as the efficacy of healthcare interventions should be judged against the efficacy of other healthcare interventions intended to have similar effects, the ethical analysis of health incentives should be done in a way compatible with other forms of healthcare intervention. For the most part, healthcare interventions need little positive justification over and above cost-efficacy (relative to other interventions). This will not always be the case, but it is likely that an ethical assessment of a healthcare intervention will take this 'negative' form.

Such an approach to analysing the ethics of health incentives will necessarily be limited in the conclusions it can draw. At most, a negative account of this form could conclude that there are no convincing ethical objections to the use of health incentives, which is by no means the same as concluding that health incentives *should* or *must* be used. What is left out are debates about the priority of tackling particular health problems; rationing debates; efficacy and cost-effectiveness concerns; wider political significance of interventions; and so on. Such factors would need to be considered, in the

context of current healthcare provision, in order to provide a judgement as to whether, in fact, health incentives should be used. My project is more circumscribed, and I do not attempt to answer the much more demanding question of whether a positive account for incentives can be provided here.

It may strike the reader that this thesis draws upon a rather eclectic range of literature, and, on occasion, seems to venture into the territories of other disciplines. There are a number of reasons for this. First, the range of issues with health incentives (all of them ethically relevant) cover a wide range of topics. Although this thesis is far from a comprehensive consideration of health incentives, I do wish it to be reasonably broad, and to deal with a good number of these issues.

Relatedly, I have used the popular media as a 'way in' to the arguments concerning health incentives (Chapters One and Two). I have combined this with a more traditional philosophical approach, considering the kinds of concerns that typically arise in relation to the ethics of healthcare interventions. By starting with the media, however, I hope to address some concerns that might otherwise be overlooked (or perhaps, more minimally, to lend such concerns slightly more consideration). Those contributing to the media have rather less respect for disciplinary boundaries than the academy might have, and hence, in considering some of these issues, it has been necessary to move beyond the philosophical literature.

This piece of research, as mentioned, is also intended to complement the wider research of CSI Health. It also makes sense, therefore, to integrate some of the strands of research from health psychology and behavioural economics into this thesis. Untrained in these subjects as I am, I cannot do full justice to the complexity of the scientific theories at work and the research conducted. However, I think it would be a shame not to incorporate (and at least acknowledge) some of the important work being done in these disciplines which is also contributing to the wider question of the appropriateness of using incentives to encourage healthy behaviour. It is my conviction that an ethical analysis of this sort will be helpfully informed by a sensitivity to these wider issues.

Finally, I wish for this research to have at least some potential for practical application. This means that my approach to considering the range of

criticisms of health incentives has always sought to be responsive to ordinary notions of morality and acceptability. One possibility for writing this thesis would have been to adopt a very specific philosophical framework and then set about assessing the legitimacy of health incentives from this perspective. For instance, one might lay out a theory of virtue ethics and then look to understand the impact that using health incentives could have on the capacity for people to develop virtuous characters.

I have not done this. Such a method of researching this subject may well turn out some interesting results, and provide new ways of looking at the issues. However, as a 'first look' it seems to me that it will be more profitable to retain a degree of flexibility, and to be rather more relaxed about the kinds of philosophical theories used to assess arguments relating to incentives. Further, it would seem a rather arbitrary choice to select a single theory from which to assess incentives, as I have no settled position on what ethical theory would make the most useful starting point for this research.

That said, I certainly have more sympathy with some philosophical traditions than others. In my quest to provide an account of incentives that is both plausible and as close to what I feel is 'right' as possible, I do seek to selectively apply theory which I feel will be most constructive. Where this is done, I have tried to be clear and to provide sufficient background information on the theory or theories used so as to make the analysis intelligible. The slight fudging I indulge in with regard to the theoretical framework of the thesis as a whole is intended to avoid alienating readers with a range of different philosophical inclinations. I hope I do not lose too much by way of philosophical rigour in this pursuit of applicability and acceptability.

Trade-Type and Aid-Type Incentives

It is worth saying a little more at this stage about incentives as they relate to the different disciplinary motivations for employing them, as mentioned earlier. This is the different thinking that I have connected to the economic and psychological approaches to behaviour change. In the first case, we may think of the incentive as an extra inducement to make one behaviour more appealing than it previously was, in the hope of altering an individual's preferences and getting her to change her behaviour. Where an incentive takes this form, I shall describe it as a *trade-type* incentive, for it acts as an economic trade.

In the second case, the incentive acts to support an individual who already wishes to change her behaviour, to successfully do so. Often, agents suffer from time-inconsistent preferences, and impulsive desires may override more considered, 'cool-headed' decisions about how one would prefer to act. Incentives can act here to aid individuals to act more consistently with their healthier preferences. Incentives operating in this way shall be described as *aid-type* incentives.

This is a distinction that comes in part from different approaches to bringing about behaviour change (persuading people to change versus supporting them to do so), and in part from different ways of describing incentives (paying people to be healthy versus a healthcare intervention). At this level of specification, it can be difficult to distinguish instances of trade-type and aid-type incentives: the incentive provider's intentions may not be fully developed, and agents receiving incentives may have some but not much motivation to alter their behaviour (meaning it is not clear whether incentives act to induce or support them). It is further unclear if facts about the incentive itself can help determine whether it acts as a trade or an aid.

Yet the difficulty with drawing this distinction need not mean it is specious, nor should it undermine its relevance to an account of the ethics of health incentives. It may be more useful to think in terms of a spectrum, with health incentives being more or less trade- or aid-type. One of the key features distinguishing these different sorts of incentives will be the psychological effect on the recipient (whether they experience the incentive as a trade or an aid), yet this will be difficult to assess. There will, however, be other more accessible indicators that could help determine where an incentive should lie on the spectrum.

As the thesis develops, so too will this distinction between trade- and aid-type incentives become more fleshed out, along with the significance of this distinction to an ethical appraisal of health incentives. It will not be particularly helpful, at this point, to try to say more about this distinction.

Structure

The discussion in this thesis will proceed as follows. In the next two chapters I will consider health incentives and the media. Chapter One provides some background theory on the nature of news media content, and how this might

relate to public attitudes regarding incentives. I suggest that there will be value to be gained from looking at the arguments appearing in media publications relating to the use of health incentives. I also note that there will be limitations to what the media can usefully tell us. In Chapter Two, I discuss the arguments for and against incentives identified in a media analysis conducted by Parke et al. (2011) in collaboration with CSI Health (and included in Appendix B).

Beginning with the arguments extracted from the media, I identify a number of more general themes of criticism relating to incentives. It is these themes which I then take as the basis for the substantive consideration of the ethics of incentives in the remaining chapters. These themes relate to the efficiency and efficacy of health incentives; coercion and other forms of pressure placed on the recipient; worries that incentives do not respect principles of fairness and desert; and the potential for values that we care about to be corrupted.

Thus, Chapter Three focusses on some issues which typically arise in the arena of political philosophy. I begin by considering efficiency and efficacy and the sorts of requirements that should be made of a state-provided healthcare system. This leads me to discuss the legitimacy of perfectionist policies and whether the use of health incentives oversteps the limits of acceptable state action. On this theme of political responsibility, I briefly consider some distributive issues relating to incentives, given that the just distribution of resources in society is one of the main requirements of a state.

Chapter Four refocuses the discussion onto the individual and the potential for health incentives to have a harmful impact on recipients. This relates to the pressure placed on agents to change their behaviour, and the kind of influence incentives exert over recipients. My main concern here is with the kind and degree of pressure placed on the agent's will, and whether this is likely to undermine her capacity for free, autonomous action. I begin this chapter with a consideration of coercion, and the typical features of this form of influence. After developing an account of coercion and applying it to the case of incentives, I move on to look at other means of influence: persuasion, manipulation, and nudging. I provide an analysis of how incentives can be thought to invoke such means of influence over recipients, and the extent to which this should be thought problematic. In doing so, I introduce some theory from social psychology relating to the operation of two separate systems of behavioural

control within humans. Some understanding of such mechanisms of human psychology are helpful for philosophical considerations of autonomy in general, as well as being useful for the particular case of how incentives influence behaviour.

Chapter Five responds to concerns that health incentives are unjust, focussing on one particular aspect of justice: the principle of desert. This discussion relates to criticisms that agents in receipt of incentives are not deserving of these kinds of rewards. In considering this criticism, I discuss the requirements for agents to be thought to 'deserve' praise or blame, and the particular form of treatment that follows from such ascriptions of desert. I present some of the psychological and sociological literature which seeks to describe the underlying mechanisms that control lifestyle choices and health-related behaviour, as well as to explain the trends of health outcomes seen across different populations within the UK. I argue that this research has relevance to the assessment of the responsibility of agents for their lifestyles, and the extent to which it is appropriate to blame them for poor lifestyle-related health outcomes they may suffer.

The penultimate chapter considers criticisms that health incentives can corrupt values we find important in the process of seeking to improve health. These criticisms tend to be rather vaguely characterised, and so I spend some time trying to identify the conditions under which corruption is thought to occur, and how and why it is supposed to occur. Related to these worries that incentives can corrupt are arguments about the potential for external rewards to undermine motivation and to have a deleterious impact on desirable (healthy) behaviours. This chapter also provides a brief discussion of some of the claims from research within behavioural economics (for the most part) on motivation crowding theory, and the extent to which we should worry that incentives could have perverse effects on health behaviour.

In the final chapter I summarise the key points from the earlier discussions, and draw some overall conclusions about the power of the criticisms of incentives considered here. I indicate how these general conclusions could be developed and applied in more specific cases, but emphasise how the particular context of a given incentive scheme, provider, or recipient must be taken into account in order to make an assessment of the ethical justifiability of a particular use of health incentives.

ONE
THE MEDIA PART I: THEORY AND PUBLIC OPINION

INTRODUCTION

I wish to begin this discussion of health incentives by first considering what the 'ethical concerns' with this form of intervention include. Hence, in these early chapters, I intend to survey what criticisms are to be made of incentives. One source of these criticisms is the media: a number of articles have been written about the use of incentives to alter health behaviour, and the potential for healthcare policy to incorporate incentives into treatment programmes. An analysis of print media articles about incentives was conducted by Parke et al. (2011), identifying articles reporting on the use of incentives, the 'for' and 'against' arguments featuring in those articles, and how these relate to the incentive schemes being discussed. Further data were gathered on the target readership (i.e. whether the media was directed towards an expert or lay audience) and what sources were used in the articles (for example, government spokespeople, incentive scheme participants, healthcare professionals, and so forth).

In the next chapter, I will make considerable use of this media analysis, giving a little more detail on each of the arguments extracted from the media. For ease of reference, I have included a copy in Appendix B.[2] The arguments

[2] As one of the named authors on the article, I should also briefly explain my contribution to the paper, which mainly consisted in assisting with the introductory section, as well as some general input to the rest of the article at the writing up stage.

identified in the media have a significance of their own (as I will discuss in this chapter) and they also serve as a starting point for the subsequent discussion of the ethics of health incentives. The rest of this thesis is not structured around the media arguments narrowly, but rather considers more general 'themes' that relate to (and are largely derived from) these arguments.

As a prelude to this, in the present chapter I wish to look more generally at the significance of the media and what philosophical interest it holds. First of all, I describe how the media may act as a source of arguments from which to draw upon for the present ethical analysis. Next, I go on to give some description of how the particular forces acting upon media copy shape its content. Specifically, I discuss some of the theory that proposes how media content comes to resemble public opinion. The potential to use the media as a proxy for public opinion means that the arguments relating to health incentives which feature in the press may be of special interest.

A Source of Arguments

The media may hold an interest simply as a *source of arguments*. In the case of incentives, the media features a good deal of discussion and debate about whether or not the use of health incentives is a wise idea. This means that a variety of criticisms and defences of incentives are made, employing differing degrees of clarity, reason, reliable evidence, prejudice, politics, and common sense.

There are other ways of coming up with arguments about incentives: one could attempt to read all of the relevant academic literature and see what arguments have previously been made that might be pertinent to the case of incentives. Or one could survey commuters outside Bethnal Green underground station and ask them what they think of the use of health incentives. One might instead begin with some sort of 'fundamental moral principles' (justice and liberty might be suitable candidates), and then evaluate incentives as they relate to such principles.

Different methods will extract different arguments, by accessing information from a variety of lay and expert individuals with different interests, prior assumptions, reasoning processes, and so forth. Some of these methods will seem more appropriate than others. For example, one might argue that picking random 'ethical issues' out of a hat is more likely to provide toothless

inapplicable arguments than would arise from consulting professional philosophers. It is often interesting to consider the forces acting to introduce certain arguments into the debate: political, personal, and business interests may play a role, for instance.

A consideration of health incentives that was tightly focused on the media would surely need to provide an account of such contextual forces. However, my concern seeks to consider the applicability of criticisms to health incentives and focuses more on the philosophical issues than the sociological ones. My use of the media as a source of arguments is thus rather more shallow than it might be, and the arguments and themes of argument extracted from the media are quite abstracted from this original setting in the subsequent discussion I provide. It will, however, still be worth indicating a few of the reasons for thinking that starting with the media is a potentially useful 'way in' to discussing the ethics of health incentives.

Alexander Brown suggests that using the public as a source of arguments helps to avoid the pitfalls of relying upon the philosopher's limited imagination:

> Although philosophers are adept at identifying and clarifying principles of morality, they can't be expected to dream up every idea. Opinion surveys may provide invaluable insights into principles not yet on the philosopher's radar.
>
> Brown (2009: 113)

As Brown points out here, extracting the 'data' from the public realm means some arguments that are often overlooked by academic philosophers, but which are nevertheless of importance, come to light.

One might approach arguments in the media in two ways:

1. Media Arguments should be 'privileged' above Non-Media Arguments;
2. Media Arguments should be treated in exactly the same way as Non-Media Arguments.

I intend 'privilege' to mean simply 'give special weight to' or 'pay particular attention to'. Claim 2. is based upon an assumption that arguments are of significance only insofar as they are are well-reasoned, well-founded, applicable

to the case in hand, valid, and so on. That is, features of arguments that exist independent of their provenance. In this case, there is no sense in treating arguments that arise frequently in the media any differently from how one would treat arguments that never appear in the media.

Claim 1. holds a different assumption: the provenance of an argument does (at least sometimes) have a relevance to how we should assess that argument. Specifically, here the suggestion is that there will be some added value to be gained by paying special attention to those arguments that arise in the media relative to those arguments absent from the media.

I will argue that there is something distinct about Media Arguments which relates both to how they come into being (why they appear in the media in the first place), and to their subsequent role in the public domain. For political reasons, Media Arguments often *will* be privileged over Non-Media Arguments, and this is to some extent justifiable on both pragmatic and principled grounds. Given that we are concerned with evaluating the merits of prospective action which will have a number of practical effects on the world, it is worthwhile lending some special consideration to what the media might be able to offer by way of a contribution to the discussion of health incentives.

The news media looks like a promising (and plentiful) source of arguments, with a wide range of outlets producing copy on a range of subjects of public interest. It is, however, worth remembering the process by which a body of arguments present in the media develops, as this will affect what is present and what is absent. It is mostly the actions of journalists and editors which are responsible for determining what stories will get reported and the way in which this is done. Depending on the publication, this may tend to exclude the most nonsensical, wacky, trivial and irrelevant arguments, whilst keeping the relatively sensible and apposite. Alternatively, journalistic and editorial influence may have quite the opposite effect. Certainly, some outlets typically favour more frivolous stories; tales of celebrity scandal which lend themselves to attention-grabbing headlines and salacious accompanying images. It would certainly be too optimistic to suppose that journalism either attempts or succeeds in capturing *only* and *all* the most pertinent arguments.

In an ideal world, perhaps, published arguments would be subjected to additional scrutiny, challenged and countered, supported and bolstered by

engaged readers, interested parties, experts, and other journalists. The most resilient arguments would withstand the challenges directed at them, and be repeated and expanded upon, whilst the weaker arguments would tend to disappear from the coverage.

However, the interests of one providing an ethical account of health incentives and the interests of those involved in the production of media will often diverge. Attributes of a 'good' story for a journalist may include being emotive, simple to grasp, and polemical. After all, a large concern for the journalistic industry is selling papers and communicating news items in a concise, engaging format. It is understandable for news providers to thus hold a preference for stories at the wackier end of the spectrum, over duller, more measured arguments. Without being too cynical about the motives of those involved in the production of news media, or creating a caricatured image of the readership as demanding ever-more simplistic, sensationalist media, it does seem that inevitably those stories (and styles of journalism) able to grab attention and sell more newspapers will be preferentially published. This tendency will, however, be tempered by the professional duties of those in the industry to present a reasonably accurate, balanced picture, and to report high quality stories that are of public interest.

Public Prominence

In addition to acting as a 'selective pool' of arguments, the media plays a significant role in public life. Those arguments frequently appearing in press coverage gain more exposure than those that are absent. The very act of reporting some arguments gives them a status not bestowed upon other arguments. The public nature of the popular press means that the state will be expected to be aware of, and responsive to, high profile arguments in the media. Government response, or lack thereof, to media arguments may be used to make general assumptions about leaders' flexibility and responsiveness to criticism. The extent to which the government acknowledges what arises in the media reporting of an issue may be seen as indicative of the sensitivity of the government to public opinion (or just external opinion) on the matter.

Consider: it would probably be unfair for me to expect Jeremy Hunt to respond to the specific problems I have come up with about his reforms to the NHS, as I have never voiced these concerns to him. I might, however, fairly

criticise his lack of response to concerns that have been widely reported, given that he should surely be aware of those concerns. One way for me to evaluate Hunt's receptiveness to critical opinion is to look at the alacrity with which he responds to the arguments raised in the media.

It is thus important for policy makers to be *seen* to be responsive to arguments present in the media. This relates to the perceived legitimacy of those in positions of power. It need not relate to the most urgent actual concerns people (the public) have about a particular issue, nor the most philosophically interesting. However, political pragmatism means that arguments arising more frequently in media coverage will likely come to receive more attention from those in power.

A Mirror and a Mould

Another reason for privileging Media Arguments over Non-Media Arguments is if media content is representative of public opinion. In this case, by lending special attention to those arguments arising in the media, one is able to elevate those concerns more important to the general public. This will have value insofar as one wishes one's analysis to be responsive to the opinions of the general public. However, the extent to which the media tracks public opinion is debatable. I now intend to introduce some of the theory which describes how media content can act as a proxy for public opinion. Later in this chapter I will question the value such 'public opinion' holds for ethical analyses such as this.

Two lines of explanation have been posited for describing how media content and public opinion can come to resemble one another. These are the 'mirror' and 'mould' theories which I will briefly describe below. There is also a sense in which we might take public opinion to *be* that which is measured by media analyses, opinion polls, and the like. This social constructivist approach will also be discussed following the Mirror and Mould sections. This approach assumes that 'Media Opinion' (public opinion as derived from an analysis of media content) and 'Public Opinion' (a phenomenon defined by the mental states of individuals) are one and the same. For now, however, I shall treat Media Opinion and Public Opinion as though they are two separate, non-synonymous entities.

Mirror?

> [T]he often overlooked impact of audiences on media content suggests the ability of media to reflect public sentiment… Media analysis in many ways provides a useful alternative to traditional public opinion measures
>
> Danjoux (2007: 8)

Some theorists have explored how public opinion is reproduced in the media. It is proposed that audiences actively seek out media that presents messages that are in accordance with their own established views. This produces an incentive for media providers to create content that resembles public opinion. Providers seek to establish themselves as "endorsing a particular set of values that reinforces [the] audience's self-perception." (Danjoux [2007: 14]) Additionally, advertisers paying for space in media publications will not wish to be associated with an outlet that is at odds with the views of its own customers. Hence, it is proposed, media providers create material that reflects public opinion: such theorists describe the media as a 'mirror' for public opinion.

A number of authors argue that media providers have become very adept at reproducing public opinion. The success of this mirroring enterprise, it is argued, is apparent in the tendency of politicians to use media analyses, rather than opinion polls, to gauge public attitudes (Growing [1994]; Powlick [1995]).

Mould?

Others argue that the media not only *reflects* public opinion, but also plays a significant role in *moulding* it. Gerbner (1998), for example, describes how 'cultivation analysis' can help identify the influence of television viewing on audience attitudes. This is done by looking at differences in the opinions held by 'light' and 'heavy' viewers.[3] Gerbner proposes his research shows that people who watch a lot of television tend to score higher on the 'mean world index,' which rates to what extent individuals consider people in general to be not altruistic, to be untrustworthy, and to be 'out for themselves.' Heavy viewers

[3] What is counted as 'light' and 'heavy' television viewing is determined on a sample-by-sample basis, to achieve a roughly even three-way split between those who watch the least ('light') television, the most ('heavy') and those who watch an intermediate amount ('medium') (Gerbner [1998: 181]).

also score more highly on a 'sexism scale' (are more likely to hold prejudicial and negative attitudes towards women, supposedly due to exposure to more stereotyped presentations of women) (Gerbner [1998: 185-186]).

If media can affect opinion in the way described by Gerbner, then those creating news media have the capacity to significantly influence the opinions of consumers. McCombs and Shaw (1972) describe how this may have an impact on the reporting of political campaigns:

> In choosing and displaying news, editors, newsroom staff, and broadcasters play an important part in shaping political reality. Readers learn not only about a given issue, but also how much importance to attach to that issue from the amount of information in a news story and its position.
>
> McCombs and Shaw (1972: 176)

In this domain, it is argued that media reporting is likely to be the only source of information to which most people have access.[4] McCombs and Shaw present evidence to support the 'agenda-setting' role played by the media. One example is the correlation between "the emphasis placed on different campaign issues by the media (reflecting to a considerable degree the emphasis by candidates) and the judgments of voters as to the salience and importance of various campaign topics." (1972: 181)

Alternative explanations for such correlations include the possibility that both the media and the voters are responding to the same events, or that the media is mimicking their audience (recall the *mirror* theory). However, the authors argue that the lack of availability of other information outlets makes it unlikely that both groups are responding to the same events, and that "numerous studies indicate a sharp divergence between the news values of professional journalists and their audiences, [thus] it would be remarkable to find a near perfect fit in this one case." (1972: 185)[5]

The development and introduction of new health interventions could well be another domain where the media is responsible for a significant proportion

---

[4] I suspect that this is increasingly untrue (the McCombs and Shaw study was published in 1972), given the growing importance of online information sources created by non-professionals (Twitter and other blogging sites, for instance).

[5] Note that this last criticism, and the response given by McCombs and Shaw, could equally apply to the 'mirror' theory of media content.

of the information available to the public. This 'agenda-setting' function of the media could therefore be influential in determining the public response to such initiatives.

Mixture?

The *mirror* and *mould* theories outlined above are not mutually exclusive: it is possible for both these effects to operate simultaneously to influence media content and public opinion.

Van Gorp (2007) discusses the significance of 'framing' in the production and reception of media. In this context, 'framing' is described as a "bridging concept between cognition and culture" (Van Gorp [2007: 2]):

> Framing refers, on the one hand, to the typical manner in which journalists shape news content within a familiar frame of reference and according to some latent structure of meaning and, on the other hand, to the audience who adopts these frames and sees the world in a similar way as the journalists do.
>
> Van Gorp (2007: 61)

On some definitions, frames can be found in a variety of locations (such as in people's minds, or in the media content itself). Van Gorp emphasises the connection frames have with culture. He argues that shared culture, and thus shared cultural frames, play an important part in determining meaning and what counts as knowledge within a group. Accordingly, it is claimed that frames can be used to craft a particular meaning for a media story:

> [B]ecause frames are part of culture, the actual frame is not encompassed in media content. The text and the frame must be seen as independent from one another. Both the attribution of meaning to media content and the connection with certain frames are part of the reading process… The receivers connect the framing devices in a news story with cultural phenomena because they are already familiar with them. By implicitly suggesting a cultural theme, the frames can determine which meaning the receiver attaches to an issue… Because these frames often are unnoticed and implicit, their impact is by stealth.
>
> Van Gorp (2007: 63)

So, media providers can influence the way content is interpreted by implicitly associating it with particular cultural frames which shape the meaning attributed to the content by the 'receiver' (this means the reader, watcher, or listener). This is done through the use of 'reasoning' or 'framing' devices, including "metaphors, exemplars, descriptions, arguments, and visual images." (Van Gorp [2007: 64])

Aside from the frames used by journalists themselves, others who are featured in media content (such as politicians, charity spokespeople, and so on) can seek to apply 'advocate frames' in order to shape the way their messages are interpreted. All those involved in the production of media content may wield some influence over the framing, and thus meaning, of that content.

The power of media providers to direct public opinion is, however, limited. Journalists can only work with the cultural frames in existence: they do not invent novel frames. It is up to the receiver to associate media content with cultural frames, which are only indicated and hinted at by the media providers themselves. Moreover, although frames (as they exist in culture) are fairly stable, the application process is dynamic, and subject to contestation and negotiation on the side of the media providers and the receivers. Further, it also seems that media providers themselves may not always be so self-aware as to consciously adopt a particular frame. Journalists are subject to the same external influences as the rest of us, and may implicitly grasp at particular cultural reference points, and thus encourage the use of a particular frame, without realising what they are doing.

Meaning?

In both the *mirror* and *mould* theories, it is assumed that there are two separate entities: public opinion as it exists 'out there' in the world, unmeasured and independent of attempts to assess it; and public opinion as derived from media content, opinion polls, and whatever other means we have of gauging it. The former I will call Public Opinion, the latter Media Opinion. As such, the project of the mirror / mould theorists is to explain how Public Opinion and Media Opinion come to look so much alike (as it is claimed they do). The discussion of framing also serves to describe the processes acting to bring Public Opinion and Media Opinion into alignment.

However, it is not clear that we know what the Public Opinion is that the Media Opinion is claimed to resemble: what is our 'gold standard' of Public Opinion, against which we may judge the similarity of Media Opinion? Since we cannot know Public Opinion without measuring it, all we can know of Public Opinion is what we are able to glean through media analyses, opinion polls, and so on: those very methods whose application we are seeking to assess.

It is argued by some that the creation, identification, and definition of Public Opinion is determined by our tools of measurement. Osborne and Rose (1999) provide a socio-historical account of how the practice of measuring Public Opinion developed, and how this practice can be understood to have created the phenomenon of Public Opinion.

> [W]e have to acknowledge that the notion of opinion is the product of the particular procedures by which opinion is elicited... For clearly *without* surveys and forms of measurement we would not know of public opinion at all; we would have no knowledge of what there is to measure without procedures of measurement... More interestingly, we can observe that public opinion is something that is demanded by the very activity of asking questions in surveys. That is, the existence of questionnaires and surveys *themselves* promote the idea that there is a public opinion 'out there' to be had and measured.
>
> Osborne and Rose (1999: 387) (emphasis in original)

For Osborne and Rose (1999: 371) "public opinion exists, at least in the sense that it has a reality and an efficacy in the world in which we live", and it is *this* 'public opinion' that is of significance; 'public opinion' produced through a well-established research discipline, and which has a pervasive influence on politics, advertising, debate, and so forth.

In this sense, Public Opinion and Media Opinion are effectively the same thing. The phenomenon we are looking for is created by the procedures we use to measure it, and so the two become synonymous. If we were to accept this then we might take it that media analyses, as one of the conventional methods contributing to the development of Public Opinion, will be an excellent way of measuring it, and anyone interested in Public Opinion would be wise to pay them significant attention.

It seems sensible to say that there is at least *something* that exists, pre-measurement, but the reality of what this is and how we interpret it is heavily influenced by the systems of measurement we have in place. A telling distinction is that drawn between an 'opinion' and an 'attitude.' The former is "public by definition, collective by design… opinions are things that people, so to speak, project into the public arena", and moreover "opinions always themselves relate *to* specifically public matters". Attitudes, in contrast, are personal things, held by an individual agent, perhaps even to be considered 'reflexes' (Osborne and Rose [1999: 386]).

Through practices of Public Opinion measurement individuals become proficient in 'having opinions,' and become amenable subjects for public opinion researchers. In this sense, Public Opinion is something that exists only after these practices of measurement have become established. Although dependent upon various social institutions for their existence, opinions are often not themselves made public. In order to have opinions, we may need these systems of measurement in place, but it does not follow that those systems will then always succeed in capturing our opinions. Attitudes are something that people can hold in the absence of the practice of Public Opinion measurement, but will clearly be closely related to opinions.

We must thus be careful about the terminology we use when discussing these concepts. The phenomenon of Public Opinion, as described by Osborne and Rose (1999) will be very close to Media Opinion. However, it will not be synonymous with the aggregate of all the opinions held by members of the public, and nor will it accurately reflect the attitudes of all those individuals. Depending on our reasons for wanting to know what opinions people hold, we may be more or less satisfied with the output of media analyses and similar tools.

WHY IS PUBLIC OPINION OF INTEREST?

There are numerous reasons why policy makers might be interested in public opinion. Two such reasons are based on *pragmatic* and *democratic* grounds. First, it will be pragmatically useful to know how a new policy or intervention is likely to be received by the citizenry. If a great deal of hostility is probable, perhaps the policy ought not to be implemented at all. Knowledge of public

attitudes may help politicians to alter policies so that they are more palatable, or introduce them in ways that cause the least upset.

Second, a democratic government's legitimacy is based on the capacity for the electorate to express their preferences through voting for those they feel will best represent their interests. Public involvement in policy is important to ensure that elected governments remain sensitive to the attitudes of the public, and responsive to (some of) their demands. The practice of voting in an election is itself a method of assessing Public Opinion, but public involvement in the government of the country can go much further than this.[6]

Media Opinion will thus not only provide a useful source of arguments, but, to the extent it tracks public opinion, it may be of use to policy-makers on these pragmatic and democratic grounds. Media Opinion will, however, have its limitations. As discussed earlier, very particular processes are involved in producing news media, which shape the discussions taking place and determine what arguments relating to incentives will be featured, and how they will be reported.

Concluding Remarks

One way of surveying the arguments relating to incentives is to look at those arising in the news media. I have suggested that the media can act as a 'pool' of arguments relating to incentives, but that very particular forces act to determine what arguments make it into media reporting and correspondingly, what arguments are left out. Thus, media analyses which describe the contents of press reporting of incentive schemes will give a somewhat skewed representation of the possible arguments that might be made about health incentives.

This 'skew' might have a particular use, however. I summarised some of the literature which proposes that Public Opinion and Media Opinion are very similar, and on some accounts, the same. Thus, extracting Media Opinion can help understand Public Opinion. There are limitations on the extent to which Media Opinion might be thought to track Public Opinion, assuming we do not adopt a fully constructivist approach which takes the two to be equivalent.

---

[6] For an interesting set of discussions on the role of experts and publics in policy-making, see Doubleday and Wilsdon (2013).

Imperfect distinctions between 'publics' must be drawn, media must be processed and interpreted, opinions must be simplified and aggregated into more basic viewpoints.

Despite imperfections resulting from the processes by which we extract the opinions of individuals from the media, such practices can still help us to understand the 'general mood' of the public, and provide useful information as to how likely it is that an intervention will be met with enthusiasm or hostility. Public Opinion expression through media outlets is also a means through which democratic processes can operate, and citizens can judge the actions of those in power. Media production may also provide an opportunity for those involved in creating media content to influence the views of the public, and determine the sorts of debates that take place.

Thus, the media analysis from Parke et al. (2011) is interesting both as a source of arguments, and as an indication of Public Opinion. In the next chapter, I will summarise some of the findings from the media analysis, and describe the kinds of arguments that arise regarding incentives. I do not structure the remainder of the thesis around these media arguments strictly, but rather, at the end of Chapter Two, I suggest how these arguments might be grouped into more general 'themes,' and it is these themes which I then use as the basis for chapters Three to Six.

TWO

THE MEDIA PART II: ARGUMENTS AND HEALTH INCENTIVES

INTRODUCTION: THE ANALYSIS ANALYSED

This chapter will look at the arguments extracted from the media by Parke et al. (2011). A copy of the full media analysis is included in Appendix B for reference. First I will introduce the analysis and some of its findings, before looking more closely at the different arguments identified in the media reporting on health incentives. The aim of this consideration of arguments in the media is to provide a basis for the consideration of ethical objections to the use of incentives to encourage healthy behaviour. As discussed in the previous chapter, with some caveats, the media can act as a pool of arguments, as well as a proxy for Public Opinion. In both these roles, it may provide a useful starting point for this discussion.

Parke et al. (2011) analysed 210 articles which reported on some aspect or other of the use of financial incentives to alter health behaviour. All of these articles appeared either in the print media or online between January 2005 and February 2010. Articles for inclusion in the analysis were identified using a search strategy that involved inputting different combinations of key words into various databases for print and online media. A full description of the search strategy can be found in Parke et al. (2011: Appendix 1 [see p. 270 of this document]). The articles selected for inclusion were then coded according to a number of criteria: the overall tone adopted toward incentives (i.e. favourable / unfavourable); the sources referenced (for example, politicians, charity

spokespeople); and the arguments presented in the article (both 'for' and 'against' incentives).

As reported in the analysis, most of the articles (58%) adopted a mixed attitude towards incentives (were neither overly negative nor overly positive). 13% of the articles tended to give a more negative picture of incentives, whilst the remaining 29% tended to be more positive. The overall positivity / negativity of the articles varied with regards to the behaviour being targeted by the incentive scheme: schemes incentivising weight loss and illicit drug use reduction received more unfavourable reporting, whilst schemes intended to improve the health of pregnant women received significantly more favourable reporting. In all, six different types of health behaviour were discussed in the articles reporting on incentive schemes: weight control, health promoting behaviour in pregnancy, antipsychotic medication adherence, illicit drug use reduction, smoking cessation, and sexually transmitted infection (STI) testing.

Sources quoted in articles were more often favourable in their discussion of incentives. Such sources tended to be government spokespeople, public health representatives, services allied to medicine, academics and doctors. Sources that were typically more critical of incentives tended to be opposition party politicians and lobbyist groups (such as the Taxpayer's Alliance). Charity spokespeople were also quoted in media articles, and tended to be fairly evenly split between support for and opposition to incentives.

A coding scheme was developed to identify the different 'for' and 'against' arguments featuring in the media articles. Each time an argument on the coding scheme appeared in an article it was noted down, and each time a novel argument was found it was added to the coding scheme. This way a scheme was built that included the different arguments arising in the articles and their frequency. In total, 34 different arguments were identified: 15 'for' and 19 'against.'

The articles were also categorised according to which type(s) of health behaviour they discussed (out of weight control, behaviour in pregnancy, antipsychotic medication adherence, drug use reduction, smoking cessation, and STI testing, plus some articles which discussed incentive schemes in general). The frequency of different arguments and the sources included can be cross-referenced with the health behaviour being targeted by incentives in each case.

THE ARGUMENTS

The full list of 34 arguments, with definitions and examples, is provided in the appendix of Parke et al. (2011: Appendix 2 [see pp. 271-275 of this document]). The different arguments, along with the number of times they feature, and how many different articles they appear in, are given below:

'For' Arguments:

Positive Evidence (*130 times; 82 articles*)	Financial Burden (*50 times; 34 articles*)	Not Novel (*10 times; 10 articles*)
Best Interests (*124 times; 94 articles*)	Last Resort (*45 times; 32 articles*)	Improves Inequality (*9 times; 6 articles*)
Prevalence (*117 times; 78 articles*)	Benefits Society (*26 times; 19 articles*)	Positive (*8 times; 8 articles*)
Money Saving (*79 times; 54 articles*)	Innovative (*19 times; 15 articles*)	No Harm (*3 times; 2 articles*)
Disease (*63 times; 47 articles*)	Other Organisations (*11 times; 9 articles*)	Praise (*2 times; 2 articles*)

'Against' Arguments:

Better Spent (*111 times; 61 articles*)	Wrong Message (*19 times; 16 articles*)	Health Harm (*12 times; 7 articles*)
Not Long Term (*46 times; 24 articles*)	Encourage (*18 times; 14 articles*)	Dr-Patient Relationship (*9 times; 9 articles*)
Misuse (*39 times; 29 articles*)	Unethical (*18 times; 13 articles*)	Cash Won't Work (*6 times; 4 articles*)
Bribe (*33 times; 27 articles*)	Responsibility (*16 times; 11 articles*)	Political (*4 times; 3 articles*)
Rewards Unhealthy (*30 times; 20 articles*)	Nanny State (*12 times; 8 articles*)	Universal (*4 times; 2 articles*)
Negative Evidence (*29 times; 19 articles*)	Giveaway (*10 times; 7 articles*)	Stigma (*1 time; 1 article*)
Criticism (*21 times; 18 articles*)		

Parke et al. propose that each of these refers to a distinct argument made in relation to health incentives. I will briefly go through each of the arguments in turn to give a little detail of what each is about. Some of these arguments look like they share connected themes. For instance, Negative Evidence and Cash Won't Work both seem to be making a claim about the efficacy of incentive schemes (or lack thereof). Other arguments are concerned with a variety of other features, including justice, or with the autonomy of participants in incentive schemes.

I will seek to draw out some of the themes running through and connecting the different arguments. Often, the discrete arguments as they appear in the media, as identified in the media analysis, can be interpreted in different ways. As such, a particular phrase could be interpreted as criticising (or praising) incentives on more than one basis. In discussing more general themes, the significance of close interpretation of each argument is diminished. It also removes some of the repetition that may result from discussing each and all of the (sometimes overlapping) media arguments. The effect is also to lessen the significance of the provenance of the arguments (i.e. as appearing in the media), and so too, the extent to which the benefits of looking at such arguments (the link to Public Opinion in particular) are seen. However, the gains in terms of philosophical coherence, unity of discussion, and space saving I think will compensate for these losses.

In subsequent chapters, I will expand upon these themes and structure the ethical analysis of incentives around them. In this chapter, I mean only to outline the media arguments, rather than to offer an opinion on the strength or otherwise of particular arguments. Thus, the main ambition of this chapter is mostly descriptive (though it also involves a fair amount of interpretation), rather than normative.

In the following, each argument heading is followed by the definition and the press quote as given by Parke et al. (2011: Appendix 2) to illustrate what is meant by that argument.

'FOR' ARGUMENTS

Positive Evidence

> Theory or empirical evidence which suggests the effectiveness of incentives, also personal stories of success and 'experts' who believe the schemes will work.
> *'Such schemes have been used in the U.S. with research showing participants stay drug-free for twice as long as those not taking part in incentive schemes'* BBC (website), June 2008.

The most prominent claim made in this argument relates to the efficacy of incentives: that incentives actually work. This can be extended to say something about efficiency, which will depend upon how much it costs to achieve a particular outcome through the use of an intervention. If incentives are efficacious, then they are potentially an efficient means of improving health.

This argument could also involve an implicit claim about well-being: incentives may be judged effective if they are shown to produce improvements in the well-being of the participants. This is going further than is implied in this argument, which only needs some way of judging efficacy and need not relate it to well-being (for instance, it could be weight loss or life expectancy or similar). However, improvements in well-being are often the implicit, downstream goal of health interventions, and measures such as weight loss and life expectancy are often used as a proxy for well-being.

Best Interests

> Incentives help people to do what is in their best interests, including mentions of the health benefits to the individual.
> *'The one-off payment is intended to help pregnant mums stay fit and healthy in the run-up to the birth'* South Wales Echo, April 2009.

The well-being-related theme is more apparent in this argument. Support for the Best Interests argument derives from the claim that people who receive health incentives will be made 'better off,' which is roughly the same as identifying an increase in well-being.

Health benefits are mentioned explicitly as a way in which people may be made better off by incentives, and good health is typically thought of as an

important contributor to well-being (health may also have intrinsic value distinct from its contribution to well-being). However, incentives could also contribute to 'best interests' in other ways. Acting in one's best interests implies acting rationally and with an appropriate awareness of the likely outcomes of one's behaviour. As such, this argument could be praising interventions which help people behave rationally, and follow goal-directed, long-term, well-considered behavioural choices. Such features are often thought important for autonomous action.

Prevalence

> The prevalence of the health issue is cited.
> *'The UK is in the grip of an obesity epidemic with a quarter of all adults and one in five children now obese'* Daily Mail, January 2008.

Although this argument does not make reference to why incentives are preferable to other sorts of intervention directed towards similar ends, it emphasises the significance of those ends. One claim being made here is that there are certain diseases that have a huge impact on population health, and the implicit claim is that such diseases are deserving of attention. Incentives are one possible way of combatting these diseases.

The significance of the target disease is also relevant to the potential impact of successfully treating it, in terms of efficiency and economics. Targeting diseases of high prevalence will potentially be an efficient way of allocating scarce resources (assuming that by combatting such diseases the disease burden of society is reduced more effectively than by targeting less prevalent ones).

Money Saving

> Incentive schemes will save money in the long run, they are cost effective.
> *'The trust says the plan could save the NHS thousands of pounds by preventing conditions such as heart disease'* BBC (website), October 2008.

This argument is explicitly proposing that incentives are cost-effective. This means that incentives will be effective at reducing disease for a sufficiently small

cost (measured in relation to the best alternative intervention, or the cost of not treating the disease, or suchlike). In particular, this argument mentions the long term benefits incentives may convey, presumably because they typically target chronic disease resulting from lifestyle-related behaviour. Thus, incentives can prevent people from developing diseases that would be costly to treat.

Disease

> Disease related to the health behaviour is mentioned, or more general health problems relating to the health behaviour are mentioned.
> *'Low birthweight babies and premature deliveries are much more common in mothers who smoke'* BBC (website), February 2008.

Once again, this argument is centred around the premise that some disease are highly influential, both in their direct and their indirect effects on individuals. This does not specifically support the use of incentives to treat those diseases, but rather promotes the use of interventions generally that will combat those diseases.

These claims also relate to efficiency and well-being: if the diseases targeted are highly significant then the potential impact of interventions targeting them is increased, as is the potential to improve population well-being.

Financial Burden

> Incentives will help relieve some of the financial pressures on the recipient.
> *'The financial stress associated with impending motherhood can now be eased by a £190 cash boost for women due to give birth on or after April 6'* Belfast Telegraph, January 2009.

This argument seems to relate to well-being in that it proposes that providing incentives to people will make them better off by furnishing them with extra resources. However, it speaks specifically to the capacity of incentives to make those who are particularly unfortunate better off. This has significance for ideals of equality, fairness, and distributive justice. There may be some additional 'good' to be derived from incentives if they are capable of ameliorating the

suffering of the worst off in society, and in doing so reduce the gap between the wealthiest and the poorest.

Last Resort

> Ideal as a last resort, for the hard to engage, disadvantaged and vulnerable.
> *'Financial incentives might be a treatment option for a high-risk group of non-adherent patients with whom all other interventions to achieve adherence have failed'* BBC (website), January 2007.

This argument could also relate to fairness and equality, in that it suggests incentives could help a specific group who are often left unaided by other interventions. It could also involve a claim about well-being: that incentives have the capacity to further improve well-being by helping a particular population of individuals for whom other approaches have failed.

This argument also makes a negative claim, which is that other interventions have not worked. Part of the support for incentives here is derived from the promise they may hold as an alternative to other, ineffective interventions. It is not necessarily that incentives have been shown to be effective, but that other interventions have proven not to be.

Benefits Society

> Benefits to the rest of society, not just the participant in the incentive scheme.
> *'If it works to keep people in treatment there would be considerable benefits to the public'* The Daily Telegraph, January 2007.

This argument involves claims about well-being and efficacy. First, using incentives to help treat some individuals may not provide only them with benefits, but also others. Making people 'better off' is effectively improving their well-being. Thus incentives could raise overall levels of well-being in society. There could also be benefits distinct from well-being that are achieved through incentives. Community values, such as solidarity, may be hinted at here. Such things are not easy to quantify in terms of well-being.

As with the Prevalence and Disease arguments, Benefits Society emphasises the scale of the problem and the impact a solution could have. Once again, this does not speak directly to incentives as an intervention, but to all effective interventions targeting these same problems.

Innovative

> Incentive schemes are a new and innovative idea.
> *'This is a very important aspect of Scotland's battle for better public health and it is worth being brave and innovative'* The Herald, June 2006.

In and of itself, it is not clear why an innovative, different approach should be superior to more conventional approaches. However, if it is the case that previous approaches to tackling the same problems have failed then it may speak in favour of an intervention that it is novel, rather than a tweak on previous, failed methods. This incorporates something of the negative-style argument that arose in Last Resort: incentives may not be positively supported, but their alternatives can be actively criticised.

From a political, public relations point of view, innovation may represent a positive, imaginative approach to problem-solving. Thus, distinct from the virtues of an intervention with regards to its efficacy at improving health, apparently innovative interventions may be particularly appealing to healthcare providers.

Other Organisation

> Other organisations are looking to follow suit with health incentive schemes, or other organisations are advocating the use of incentive schemes.
> *'The Welsh Office said it was interested in pursuing a similar scheme'* BBC (website), June 2008.

The simple statement that others are interested in incentive schemes does not provide a supportive argument for the use of incentives, but it might indicate that those organisations have been convinced by other arguments. One might trust the judgement of these other organisations and choose to rely on their

judgement of incentives, and thus to follow them in choosing to support the intervention.

Not Novel

> Incentives are not a new and controversial idea, but an already established practice.
> *'Professor Priebe argued that financial incentives to influence healthy behaviour already existed, such as higher taxes on cigarettes and alcohol'* BMJ, October 2009.

It is interesting that both novelty and lack of novelty are presented as favourable features of incentive schemes in the media. This nicely illustrates how important framing can be to the presentation and interpretation of media reporting. One can argue both that incentives are a new idea and that they are not, and that this is a good thing and that it isn't.

Not Novel could be an argument that seeks to reassure those who think incentives have the potential to create harmful effects that are normally absent from more conventional interventions. In particular, stressing the prevalence of incentives in everyday life could mitigate the suggestion that incentives are overly interfering with individual free choice or that they are destructive of autonomy in any new or unusual way.

This argument could also seek to show that the extensive use of incentives in other instances should be taken as evidence of their success. This is similar to how Other Organisations sought to show that the adoption of incentive schemes by others implied that those others were convinced by the arguments in favour of incentive schemes. If we trust the judgement of those who use incentives, we should ourselves use incentives.

Improves Inequality

> Incentives may help to close the gap in health inequalities in our society.
> *'The scheme by Tayside Health Board aims to break the link between low income and high levels of nicotine dependency'* The Herald, June 2008.

This argument is directed towards ideals of fairness and equality, suggesting that it is a good feature of a health intervention if it is able to counteract forces that exaggerate the gap between the wealthiest and poorest in society. The summary description from Parke et al. repeated above refers to *health* inequalities, but it seems the Improves Inequality argument could also refer to inequality more generally, given that there is such a well-established link between socioeconomic factors and health (see Chapter Five). If incentives can reduce the gap in health outcomes between the most and least deprived, they may also reduce inequalities in other areas.

Positive

> A more positive way of relating to patients, it is more honest and can improve the doctor-patient relationship.
> *'It provides a much better and positive way of relating to drug users than sometimes we have done in the past'* BBC (website), January 2007.

This argument could be making a claim about well-being, namely, that there are benefits to recipients that derive indirectly from the method of the intervention, rather than just from its intended outcome effects. Improvements in the doctor-patient relationship could also be given currency as a means by which well-being can be boosted.

Alternatively, incentives could promote values other than well-being. Perhaps the provision of incentives involves treating people in ways that are respectful and which augment rather than detract from agent autonomy. The reference to the doctor-patient relationship is also interesting, and there may be values associated with this interaction that are worth promoting in their own regard, such as trust, professionalism, and loyalty. These values may not be captured in terms of improvements in well-being.

No Harm

> No harm can come to people involved in incentive schemes.
> *'There is no harm intended or caused - the service users can revoke the offer at any time'* The Metro, July 2007.

This is another argument that seems to make a negative, rather than a positive claim. No Harm does not give us reason to think that incentives will necessarily be beneficial, but reassures us that at least they won't make people worse off. This could be a defence given against criticisms that people could be harmed by an intervention such as this. No Harm could be a defence against claims that incentives would have detrimental effects on well-being, or that they are are coercive, freedom-limiting, destructive of autonomy, or a variety of other criticisms, including those featured in the media and discussed later in this chapter.

Praise

> Praise for incentives without any reasoning given.
> *'Others feel it's an effective way to cut the problem'* The Mirror, July 2009.

There is little to be said where the claim that incentives are a good idea is not supported by any form of reasoning! However, the example quote given above to illustrate Praise does reference efficacy, and so arguments of this sort may be involved in generalised praise of health incentives.

'Against' Arguments

Better Spent

> The money being spent on incentive schemes could be better spent elsewhere, includes mention of the cost of the schemes and references to taxpayers' money.
> *'Is NHS cash going to be channelled into dance lessons and vouchers for fatties when people need cancer drugs and better end-of-life care?'* Belfast Telegraph, November 2008.

The most obvious claim in the Better Spent argument is one relating to efficiency: this claim places money central to the discussion of the appropriateness of health interventions. Explicitly, the claim is that incentive schemes are not an effective way of achieving the outcomes we desire

(primarily, better health). As such, there might be other ways of spending the same money that will achieve greater results.

There might be other motivations behind such a claim. The 'better' in the description is ambiguous. There may also be a claim about what we *should* care about. For instance, it could be that although spending money on incentives to tackle lifestyle related disease would reduce the overall disease burden on the nation, or produce the greatest gains in well-being, it would actually be 'better' to target other (causes of) disease.

The literature on attitudes towards risk is enlightening here. There are a huge number of factors influencing how people perceive risk (see Slovic [2000]), and some of these could be at work here. For instance, people tend to be more averse to small risks of catastrophic events than larger risks of less bad events. Also, people are keener to avoid risks over which they have no control: this is thought to partially explain why people tend to think more money should be spent on rail safety than road safety (Wolff [2005]). The main point to take from this literature is that the probability and severity of risk does not have a linear relation to people's desire to avoid risk. Thus, it is perfectly plausible (likely, in fact) that similar attitudes will be at work in the case of avoiding risk of disease: some risks will be avoided with more urgency than others, out of proportion to their likely effects on well-being.

Some disparities between expected utility and preferences for funding particular interventions could relate to the perception of the recipients, as well as the nature of the disease or risk. If we consider that those targeted by financial incentives are somehow blameworthy for the diseases they suffer then we may consider them less deserving of help. Thus, it may be thought 'better' to spend money on treating a patient with childhood leukaemia than on incentives to encourage a lifelong smoker to quit (consider the quote above which sneers at the idea of 'vouchers for fatties,' in comparison to treating cancer patients).

Not Long Term

> The scheme won't offer any long term solutions, it does not address the roots of the problem.

> *'This is no kind of long-term solution - a temporary financial incentive won't stop people putting the weight back on once they have got the cash'* Daily Mail, January 2009.

Again, the Not Long Term argument seems to be primarily a claim about efficiency. This time, it is proposed that incentives will not have the kind of sustained effects that will lead to significant health improvements. If this is the case, then incentives may not be economically efficient in the long run.

There could also be criticisms against short-termism, regardless of the economic effects. So, it could be seen as uncaring to implement healthcare interventions that one knows will only work for a brief period. For instance, some incentive schemes target pregnant smokers. Even if these schemes only have a fairly short period of efficacy, they may have defined health (and economic) benefits (in particular, for the unborn child). However, it may be seen as cynical to offer an intervention to help women quit smoking, when really one is aware that the recipients will probably start again once the incentive is removed.

Misuse

> People may abuse the system, fraud may be a problem. Or, the rewards may be misused by the recipient.
> *'But they are expected to be free to spend the cash as they see fit, even on unhealthy products like cigarettes and alcohol'* Birmingham Post, September 2007.

Some sort of efficiency claim could be at work here once again. If people cheat in order to get rewards without any simultaneous behaviour change then money will be lost. Moreover, there could be a secondary efficiency effect, whereby those who earn rewards spend them on 'unhealthy' things, meaning any good achieved by the intervention is undermined, and the incentive scheme fails to achieve health improvements and thus economic benefits.

If individuals end up indulging in unhealthy activities as a result of the rewards received through incentive schemes, there will not only be a negative effect on the efficacy of such schemes to improve health, but on the well-being

of the participants. One might worry that recipients of health incentives are not really made better off if they are likely to put the rewards to 'bad' use.

There also seems to be a claim relating to desert. Incentive schemes could provide an opportunity for undeserving people (those willing to cheat and steal from the NHS) to gain some extra money, at the cost of taxpayers. Such an effect would run contrary to our usual ideals of people getting what they deserve.

Bribe

> Incentives coerce patients into making decisions they may not otherwise have made.
> *'The option of being paid to take a drug treatment could unduly influence people's decision making over whether the treatment is right for them'* BMJ, October 2009.

The primary concern involved in this argument seems to relate to the well-being of participants. If incentives distract people from thinking about the important aspects of a treatment (such as side-effects), then those individuals may be more likely to make poor decisions, or ones they will later regret. This tends to assume that individuals themselves know what is in their best interests, and that interference is (at least generally) undesirable.

A related worry could be one regarding the freedom or autonomy of the participant. Incentives may interfere with an individual's capacity to make a decision based on preferences she holds that are consistent with her self-identity, general outlook on life, long-term plans, 'better self', and so on. One aspect of this is freedom from a coercive, external influence (feeling pressured by those offering incentives to abide by their wishes to comply). This, in turn, may detract from well-being (one is made worse-off if subjected to coercive influence). Autonomy may contribute to well-being as well, though it is also often thought to be valuable distinct from any contribution it makes to well-being. Thus, coercing people to act in ways that will improve their health (and well-being) may simultaneously detract from their well-being by smothering their freedom and eroding their autonomy. Damage to these values may have further evils associated, distinct from well-being losses.

Finally, the Bribe argument could relate to social norms and the corruption of things we value. It might be considered inappropriate to use money in this way, to encourage people to be healthy. Many people have a 'yuck' response to the idea of exchanging money for health, with some notion that these are incommensurable goods. Financial incentives may thus be thought to destroy something we value about being healthy, being cared for by medical professionals, social norms, or similar.

Rewards Unhealthy

> The unhealthy or undeserving are rewarded whilst the healthy miss out.
> *'Why is this society so hell bent on rewarding the least deserving?'* Aberdeen Evening Express, January 2007.

Desert, again, is important to this argument. The claim seems to be that those who are candidates for health incentives are blameworthy in some way for their predicament, and thus undeserving (or less deserving) of help, at least in the form of financial incentives. This desert claim may be extended to become a more general claim about justice: the payment of incentives to a particular group(s) of people could create a disparity at the population level which conflicts with our ideals of justice.

There could also be efficiency-related consequences from interventions that 'reward the unhealthy' if such interventions encourage unhealthy behaviour, rather than encourage the opposite.

Negative Evidence

> A lack of evidence to support the effectiveness of the incentive scheme, or evidence that shows the schemes to be ineffective.
> *'There is however little research that shows that a financial incentive, combined with nutritional advice, is enough to persuade mothers from the most deprived areas to change their lifestyle'* News-Medical (website), September 2007

Efficiency seems to be a feature of this argument, with the claim that there is not clear evidence to show that incentive schemes will be significantly effective.

There may also be a degree of scepticism about the evidence that does exist. Although some schemes appear to show an effect, these may not be sufficiently reliable indicators to make one think incentives really would be successful at improving health. Often, incentive schemes are pragmatic interventions set up by healthcare providers, with the aim of improving health, rather than of testing the efficacy of incentives. As such, it can be difficult to infer conclusions about how effective incentives are from the results of these schemes.[7]

Criticism

> An incentive scheme is criticised without any reasoning provided.
> *'Drug workers described the proposals as 'ridiculous"* Daily Mail, January 2007.

There is little that can be said to extend upon this 'argument' as no specific claims are made. This perhaps indicates a general hostility towards health incentives, which may be based on any number of different reasons, none of which are mentioned.

Wrong Message

> Rewarding healthy behaviour sends out the wrong message. Being healthy should be its own reward, incentives shouldn't be the motivation. Rewards undermine intrinsic motivation.
> *'Staying healthy should be enough of an incentive for people to come in for testing, they shouldn't need to be bribed by the opportunity to win high-end electrical goods'* Milton Keynes Citizen, July 2009.

Efficiency could be a motivator for this argument once more. The suggestion that 'rewards undermine intrinsic motivation' implies that, whilst incentives might provide some motivation for behaviour change, any benefit they have could be reversed by a loss of 'intrinsic' motivation. A secondary efficiency-related effect could arise if social norms are changed such that people come to expect incentives in exchange for healthy behaviour. In time, this could be very

[7] See House of Lords (2011: 45-51) for a discussion of some of the considerations and difficulties associated with evaluating behaviour change policy interventions.

costly if incentives become more and more relied upon as a means of encouraging healthy behaviour.

It may also be implied that one form of motivation is better than another. So, even if intrinsic motivation is lost when rewards are introduced, the overall degree to which people are motivated to change their behaviour could still be sufficient to justify incentives on economic efficiency grounds. However, one might claim that it is *better* to be motivated by intrinsic desires (to be healthy, say) than extrinsic rewards (incentives). There may be different reasons for thinking this: perhaps it is more satisfying for individuals to act on intrinsically derived desires than extrinsically planted ones, or it could be thought more compatible with autonomous agency.

The concern about corruption seems to arise here as well: the idea that money is an inappropriate motivator for healthy behaviour, and people should want to be healthy just because there is value in being healthy itself, not because there will be associated rewards.

Encourage

> A perverse effect of incentive schemes may be that they encourage people to adopt the unwanted behaviour.
> *'We have to be careful we don't appear to be encouraging people to take up smoking in order to reward them for giving up'* The Metro, February 2009.

Again, efficiency is an obvious concern in the Encourage argument. As mentioned in relation to the Rewards Unhealthy argument, if incentives have the perverse effect of encouraging unhealthy behaviour, perhaps because individuals are tempted to actually take up unhealthy behaviours in order to earn rewards for subsequently quitting those behaviours, the system will lose money. Incentives may thus fail to be an effective way of improving health.

In addition to this, one might also worry about individual well-being. Incentives that encourage unhealthy behaviours could do more harm than good in terms of affecting people's health and well-being.

Unethical

> Where the ethics of the scheme are raised without any further explanation.

> *'Three quarters of respondents said they had concerns about using financial incentives, most of whom said the practice would be unethical'* BBC (website), January 2007.

As with the Criticism argument, it is not really possible to say much more about this criticism. 'Unethical' is too broad an argument to prove a fruitful source of further analysis. Clearly, articles using this argument are referring to something to do with the ethics of using incentives, but this could relate to any number of things.

Responsibility

> People should take responsibility for their own health, it is within their own control.
> *'I question very much whether the NHS should be directing its resources to weight loss - something within people's control'* Daily Mail, January 2009.

The Responsibility argument could be pointing to an efficiency-based criticism of health incentives. If it is within people's power to change their health behaviour without the action of rewards then this would presumably be cheaper and more efficient. Thus, one might criticise the use of health incentives on the grounds that they are unnecessary, and instead we should focus on getting people to live more healthily without the aid of incentives.

Some form of autonomy-related claim could be involved here as well. Perhaps there are certain aspects of a person's life that should be determined by the individual herself, and something like the lifestyle she chooses to adopt could be one such thing. It might be better that people are left to take responsibility for their own health, even if this means the outcomes, in terms of disease, are less good.

Another issue that arises here is one regarding the role of the state and the extent to which a government can legitimately interfere in the lives of its citizens. The kind of interference involved in the introduction of health incentives might be a step too far in terms of the state influencing how people live their lives. The role of the state is clearly related to the need to respect the liberty and autonomy of individual agents.

Finally, an important aspect of the Responsibility argument is apparent if we think that people may be more or less deserving of healthcare. Some argue that those who are responsible for their poor health are less deserving of the help they might need to improve it. The claim that many receiving incentives suffer illnesses it is within their capacity to avoid could be the basis for asserting that such interventions are less deserved than other interventions that could be funded instead.

Nanny State

> Incentives cited as another excess of the nanny state, the government's aim to right every wrong.
> *'Time and again this molly-coddling government thinks it can right what is fundamentally wrong with our lives'* Sunday Mirror, January 2008.

Concerns relating to both the role of the state and autonomy seem apparent in the Nanny State argument. In this instance, the focus is more on health incentives as an example of the state having too great an involvement in people's lives and decisions about how to live. There need not be something specific about the way health incentives influence behaviour: it is perhaps simply offensive that the government thinks it is within its remit to influence behaviour at all.

Within this argument might also be a question about the intentions and motivations underlying the state's interference with the lifestyles of citizens. Perhaps there are questionable perfectionist values behind the government wishing to encourage exercise and discourage smoking, ones that make a claim about the best way of living, and which could contradict liberal assumptions that there are many different things one may value as part of a 'good life', and this should be left to the individual to determine.

Giveaway

> Rewards are handed out too easily.
> *'What is this great country coming to? Free gifts and handouts for junkies and failed asylum seekers'* The Sun, July 2007.

The issue of how deserving recipients of incentives are arises again here. The claim seems to be both that people are not required to do anything sufficiently deserving in order to be rewarded, and that people who are likely to be rewarded tend to already be less deserving than the average person (this certainly seems to be the case in the quote provided by *The Sun*).

Efficiency could also be a worry here: if rewards are provided without sufficient demands being met, this could result in a loss to the NHS, or the economy more widely. The argument is not particularly apparent or clearly explicated in the Giveaway criticism, but the suggestion that rewards are provided 'too easily' could imply that it is acceptable to reward people, but that more needs to be required from them in exchange for the provision of incentives.

Health Harm

> The behaviour being encouraged may have negative health consequences on the participant, including drug side-effects.
> *'The mental health charity MIND says that paying people could coerce people into taking drugs that are known to have serious side effects'* BMJ, October 2009.

The main concern expressed by the Health Harm argument seems to be one about the well-being of the recipient of incentives. There will sometimes be potential harmful effects with any treatment programme, and sometimes these risks will be realised. Those seeking to change people's behaviour would have to be very cautious not to bring about harmful consequences. In particular, there may be unintended consequences of incentives which result in harm to the recipient.

There may also be a worry that incentives could cause people to fail to fully appreciate the potential harms of a treatment or behaviour change (if they are so focused on the reward that they are distracted from considering all of the potential effects). Informed consent is an important procedure in medical ethics, and incentives might seem to hinder this, if consent is based more upon a desire for incentives rather than a good understanding of the likely outcomes of the intervention. Apart from having an impact on the participant's well-being, there could also be more subtle effects relating to feelings of regret or loss of

autonomy, where individuals suffer harm as a result of participating in an incentive scheme.

Doctor-Patient Relationship

> Incentives may have a detrimental effect on the doctor-patient relationship, or other relationships between clients and professionals. The relationship should be about joint decision making.
> *'It undermines the therapeutic alliance the doctor and patient have - something crucial for long-term health care'* Medical News Today (website), August 2007.

Corruption appears to be a concern here: that the inappropriate involvement of money in a situation where it ought not to be a factor is destructive of certain things we value. One might argue that money should have no role in the interaction between a doctor and her patient, and that the introduction of explicit rewards turns this interaction into a 'transaction' and damages something in the quality of the relationship. For instance, trust may be an important feature of these kind of relationships, and money could undermine this.

One might also worry that the offering of incentives by healthcare professionals could feel like an imposition on the recipient, as if the doctor is pressuring the patient to behave in a particular way. It might be objected that such an influence of a doctor over her patient is inappropriate. Although such an effect could occur in the absence of incentives, it could be that incentives exaggerate the feeling that a physician is trying to steer the patient's reasoning in some way.

The relationship between doctors and patients may be valuable for a number of reasons, and it is fundamental to a well-functioning healthcare system. The destruction of the doctor-patient relationship could have detrimental effects on the ability of the NHS to provide good healthcare. As such, there are efficiency-based reasons for wanting to preserve and protect the doctor-patient relationship from destructive forces.

Cash Won't Work

> Money won't cause a behaviour change in people.
> *'If it were just a question of money, they would have stopped eating years ago. After all, if they ate less, they would be richer'* The Daily Telegraph, January 2008.

This argument seems to express scepticism about the reliability of the theory and empirical evidence motivating the use of health incentives. Distinct from the Negative Evidence argument, Cash Won't Work does not point to evidence that undermines claims that incentives are effective. Rather, it seems to indicate an intuitive disbelief that incentives could possibly bring about behaviour change.

There may also be a concern that incentives simply won't be an efficient way to improve health. This would be the case if incentives occasionally bring about the desired behaviour change, but on a population level are not a solution to the problems of obesity, smoking, and so on.

Political

> Just another political exercise.
> *'We could be concerned if incentives were used by poor-performing treatment services to mask problems and hit government targets'* BBC (website), January 2008.

The motivations and intentions of those offering incentives come into question here. One might be concerned with the reliability of evidence suggesting incentives to be effective if one suspected those seeking to promote incentives had ulterior motives. Similarly, one might be concerned for the well-being of recipients if those providing incentives were not doing so with the best interests of the recipients in mind, but rather with the intention of appearing to be doing something of worth.

Recall the 'Innovative' argument in favour of incentives. This implied that a positive feature of incentives could be that they appear new and imaginative, and thus have public relations benefits, even if they are actually ineffective. The Political criticism seems targeted at these sorts of superficial motivations.

Universal

> The drawbacks of universal benefits as opposed to targeted ones.
> *'Why wasn't it aimed at those women more in need, rather than being given to everyone, irrespective of their income?'* The Times, April 2009.

The Universal argument points to worries about fairness: a desert-related concern that it is problematic to treat people as they do not deserve to be treated. In this case, the suggestion is that some of those receiving incentives are not deserving of them (perhaps when compared to others). This also relates to concerns about distributive justice, whereby we want to ensure that resources are distributed fairly at the population level. Providing incentives to everybody, regardless of their current wealth, could allow distributive gaps to widen even further, compounding injustices.

Another point of contention related to Universal is the efficiency-based concern that providing incentives to everybody, rather than targeting them towards the most needy, might not be the most efficient way of improving health.

Stigma

> Negative effects on the individual, including stigmatisation, loss of autonomy and loss of dignity.
> *'As well as risking further stigma of people suffering from mental illness'* The Times, September 2007.

The argument that incentives could stigmatise those in receipt of them seems mostly to be a worry about well-being: that those receiving incentives will suffer harm as a result. The description also explicitly mentions dignity and autonomy, and these could relate to well-being as well, or might have intrinsic value in their own right (which could be destroyed through stigmatisation).

A further justice worry could come into play here if it turns out that those already in an unfortunate position (the most deprived in society) are the very individuals whose misfortunes are compounded by the additional effects of stigmatisation.

General Themes and Principles

In the forthcoming analysis of incentives, my focus will be on the *criticisms* of incentives, rather than on those arguments which support their use. This is partly in order to restrict discussion to a manageable size, but it also seems more pressing to consider those claims that propose reasons as to why incentives must not be used. A primary consideration for any health intervention will be its likely efficacy. Thus, if an intervention is efficacious and affordable, that will provide the bulk of the justification needed in order for its implementation. It will then be a question of whether there are good reasons for *not* implementing it.

In order to restrict the scope of this analysis, I shall set aside the question as to whether there is sufficient reason to implement incentives. Much of the work needed here relates to the efficacy and implementability of incentives, and requires expertise that I lack. Hence, my discussion focuses on those ethical arguments that suggest incentives should not be used.

I now wish to identify certain themes that run through the Media Arguments. Often, these arguments are made in response to specific health incentive schemes, and arguments relating to similar underlying themes, or deriving from concerns about the same moral principles, are often repeated in different forms. So, a particular argument may make an explicit claim about the likely efficacy of incentive schemes, but could also incorporate assumptions about the fairness of such schemes. In this way, the different arguments tend to incorporate numerous, overlapping claims.

If we consider the three most frequently made arguments against the use of incentives, Better Spent, Not Long Term, and Misuse, they all make, or can be interpreted as including, some reference to the efficiency (or efficacy) of incentive schemes. As do the arguments Negative Evidence, Wrong Message, Encourage, Responsibility, Giveaway, Doctor-Patient Relationship, Cash Won't Work, and Universal. Unsurprisingly, the (cost-) effectiveness of health incentives is something which arises frequently in the media. Those criticising incentives often directly reference their (supposed) inability to effectively change behaviour, or they reference features of incentive schemes that will lead to downstream inefficiency.

Related to the importance of efficiency is the capacity of health incentives to promote well-being. Well-being generally refers to whatever it is that makes one's life go well, and is often used as a measure of the contribution a healthcare intervention can make. As such, well-being (or some proxy for well-being) will be crucial to the efficiency of an incentive scheme. The arguments Misuse, Encourage, Health Harm, Doctor-Patient Relationship, Political, Stigma, and Bribe can be interpreted as making the claim that incentives will either fail to bring about improvements in the well-being of the recipient, or that they will be detrimental to the well-being of recipients or the wider public, either through their direct or indirect effects.

The next theme that can be identified relates to ideas about justice and fairness. Often these claims relate to the desert of the individual being offered incentives. The argument Better Spent could involve claims about fairness, if 'better' is taken to refer to those who are more deserving, for example. Misuse, Rewards Unhealthy, Giveaway, Universal, and Stigma also relate to concerns about justice.

Another theme which a number of arguments draw upon is that which references harms resulting from limits to freedom imposed by incentives. The clearest example of this would be the case of coercion, where recipients are effectively forced to act in a particular way. Milder forms of freedom-limitation that fall short of full-blown coercion could still be undesirable, as will other harmful effects on the recipient that could result from incentives. Arguments referencing these sorts of issues include Bribe, Responsibility, Nanny State, Health Harm, Doctor-Patient Relationship, Political, Stigma, and Wrong Message.

Whereas the previous theme references harms to the recipient that could arise from incentives, another theme arising from the Media Arguments are harms to society in general, arising through the degradation of values we hold. Often this is described in terms of 'corruption,' whereby the introduction of money can 'corrupt' the values and norms previously operating in a given context. The kinds of values of concern include trust, dignity, solidarity, and respect. Arguments that seem to make reference to this theme include Bribe, Wrong Message, Health Harm, and Doctor-Patient Relationship.

There will undoubtedly be further themes that could be extracted from the Media Arguments described, and it is quite likely that some of the

arguments which I have not linked to a particular theme *could* be so linked, even if this is fairly tenuous. It is the themes I have described here that will form the basis of the chapters to follow, and which will make up the substantive ethical analysis of incentives.

Concluding Remarks

I have provided a few paragraphs for each of the arguments extracted from the media to give a little explanation of the nature of those arguments. In so doing, I have sought to extend slightly the analysis of these arguments provided by Parke et al., and explore what kinds of claims are involved in each of the discrete arguments identified in the media analysis. There seem to be a number of recurring themes, relating to things that we value such as efficiency, fairness, and well-being.

I propose to structure the remainder of the discussion of the ethics of health incentives around these more general themes, rather than the particular Media Arguments. These arguments require interpretation and, as is apparent already, they refer to similar, overlapping concerns about incentives. It will hopefully be more efficient and clearer to consider the slightly abstracted, more general themes, than working through each argument in turn, and inevitably repeating what has previously been said.

I hope this analysis will maintain some connection to the concerns arising from the media. As discussed in the previous chapter, Media Opinion is of interest not only as a source of arguments, but as an indication of the kinds of concerns that are of particular relevance to members of the public. It is worth recalling from the previous chapter the importance of *framing*, which can influence how media is received by the public, and which may also be influenced by public attitudes. This can be very important in affecting how acceptable a novel health intervention turns out to be.

In particular, incentives can be seen as either payment for desirable behaviour, or as a psychological aid to help people adopt healthier behaviours. As described in the introduction, this trade-type / aid-type distinction can influence how incentives are assessed along different ethical dimensions. Some arguments appearing in the media, and those arising in the discussion to come, will be more applicable to one or other of these conceptions of incentives.

The trade- / aid-type distinction is an example of a case where the particular presentation of the discussion of incentives in the media is important, and not just the ideas behind it. A disadvantage of abstracting from the media in the way I intend is that contextual information about how the arguments are presented is lost. For instance, in the brief discussion of the arguments above, I hope to have shown that some arguments which seem to be making claims about efficiency could also involve ideas about moral desert and fairness. It is interesting that claims about desert are sometimes smuggled into what appear to be arguments about efficiency, and this is perhaps indicative of the sorts of arguments that are thought to be appropriate to make.

Despite these disadvantages, I believe it will be more productive to structure this discussion in terms of the more general themes, rather than sticking strictly to the arguments extracted from the media. I shall, however, refer back to the Media Arguments where appropriate. The themes explored in the chapters to follow (summarised below) are derived from the claims raised in the Media Arguments.

The remaining chapters will consider:

i) Notions of efficiency, and how healthcare interventions might be assessed in terms of their capacity to produce desirable outcomes at little cost. In this chapter I will also discuss well-being, and how this is often used as a 'currency' for assessing the success of healthcare interventions. Many of the arguments mentioned above made reference either explicitly or implicitly to the efficacy of incentives, and to claims about additions to or detractions from well-being that would result from the use of incentives;

ii) Coercion, and the pressure potentially placed on recipients' wills by the offer of incentives. Coercion, and other methods of influencing an agent's behaviour, can harm the subject by limiting her freedom and capacity for self-determination. This chapter will thus cover some of the concerns raised about the recipient of incentives: that her interests must be protected;

iii) Fairness concerns, specifically, desert arose in a number of arguments relating to incentives, and shall form the basis of Chapter Five. In the process of considering desert, I also provide some discussion of responsibility and the process of praising and blaming agents for their actions;

iv) Corruption of values. This relates to the claim that the use of incentives, particularly money, degrades things that we value in the context of health and healthcare provision. Trust, solidarity, health itself, dignity and respect arise in the Media Arguments and are the sorts of things that are typically thought to be threatened by the use of health incentives.

Three

Healthcare and the State: Efficiency, Perfectionism, and Distribution

Introduction

In the previous chapter, I introduced a number of arguments relating to incentives that have featured in the media and were identified in an analysis by Parke et al. (2011). I then grouped these concerns together into broader themes of criticism which relate some of the individual arguments to one another at a more general level. In this chapter, I focus on those concerns that tend to come under the sphere of political philosophy: issues to do with how the state treats its citizens; what objects ought to be promoted through state action; what level of interference in citizens' lives by the state is acceptable, and so on. Such issues are sometimes backgrounded in discussions of medical ethics. However, these kinds of political issues are crucial to public health ethics and the debates about healthcare provision (Coggon [2012]).

The issue of efficiency is forefront in debates about public spending on healthcare. Indeed, the way the state uses the resources it receives from the taxpayer is generally of great interest to the public, and any perception of wastefulness on the part of the state is unlikely to be tolerated. It is unsurprising, then, that efficiency-related arguments are the most frequently arising arguments in the media analysis concerning health incentives. That incentives will save the NHS money is one of the best things to be said in their

favour; that incentives are a hopeless waste of money is the most damning criticism.

I do not focus solely on efficiency, narrowly construed, here. I argue that 'efficiency' is a composite notion, involving three broad and overlapping elements. I think this is true of complex calculations of efficiency in general, but is particularly the case when considering healthcare. First, there is a conceptual element: what is the scheme we should use to value healthcare interventions? When should healthcare be considered 'effective'? Second is the calculative element: how much good outcome does a health intervention produce? What costs are associated with running it? How does the cost-effectiveness of one intervention compare to another? Finally, there is the element involving judgements which situates all of this prior theorising into the context of state action more generally: does this health intervention promote the aims of the state? How much funding should be provided to different interventions?

Elements of Efficiency

Conceptual - specifying the kind of outcomes which will indicate 'good' and 'bad' effects.

Calculative - measurements of the outcomes of an intervention, generally quantified and translated into a common unit.

Contextual Judgements - considerations of the significance of the efficiency of different schemes and other values in the context of the legitimate sphere of state action.

These different elements often get run together when talking about 'efficiency.' Further, it may be difficult, as well as unhelpful, to always seek to sharply distinguish these elements of concern regarding efficiency. For consider, in developing a concept of effectiveness we will need to have in mind what is of value to us (why we should care about this sort of 'effectiveness'). This concept must then feed into what is to be measured, so it will be important to understand the constraints of our powers of measurement. We may need to accept that we cannot directly measure the effectiveness of an intervention, but we can adopt some proximate measures. At this stage it becomes important to recall the gap between what is referred to by the conceptual version of 'effectiveness,' and what we are actually able to measure.

Similarly, when considering the value of efficiency we should consider what it will be desirable that the state promote, and how our concept of efficiency can usefully guide policy. This assumes that efficiency will be a desirable quality in state action, meaning that, when we develop a concept of efficiency and methods of measuring it, we must ensure we are developing something we believe the state ought to be in the business of promoting.

I cannot answer the question of whether incentives are, or one day will be, an efficient method of improving health, and I do not attempt to do so in this chapter (nor this thesis). I provide some discussion of the conceptual and calculative elements of efficiency in the first sections of this chapter, considering the uses and drawbacks of well-being as a way of valuing health, and the translation of healthcare effects into QALYs (Quality Adjusted Life Years) as a method of quantifying efficacy. Methods of calculation and quantification (whether or not QALYs are used) generally require specific expertise suited to those trained in economics and the social sciences and is not something that I attempt here.

After introducing these elements of efficiency, I consider how criticisms (in particular, those in the media), which are apparently directed at the 'inefficiency' of incentives, relate to concepts and calculations of efficiency. Further research would be needed to uncover exactly what those whose criticisms of incentives appear in the media actually mean when they say 'cash won't work' or that money spent on incentives could be 'better spent.' In the absence of this research, I provide some suggestions as to what these kinds of criticism might be referring to. I argue that much of the controversy around whether any given healthcare intervention is efficient or not refers least of all to the calculative question, and more to the conceptual and wider question of what the state ought to do.

I then move on to discuss this broader concern of what the state ought to be doing when it comes to healthcare. This question becomes increasingly pertinent as the ethical issues surrounding healthcare provision shift from an emphasis on the clinical setting to concerns relating to public health ethics, preventative healthcare, the promotion of health at the community level and interventions outside 'medicalised' settings of hospitals and GP (General Practice) surgeries.

The role of the state in terms of healthcare provision may seem rather separate from the concern with efficiency that motivated this discussion initially. I do not think this is the case: the conceptual and normative elements of efficiency are at least as important as the calculative process. When trying to assess whether an intervention is efficient, and what this means as to whether we should use it or not, much of the answer will depend on questions about how we should value the outcomes of the intervention and how we should balance these values against others which state action may impact upon.

I thus broaden the discussion at this stage to consider some of the debates from political philosophy as they relate to healthcare provision. This is a vast topic, and not one I am able to consider in full here. I restrict my discussion to two issues. First, the question of liberty, which I flesh out in terms of whether the state ought to adopt a perfectionist or neutral approach to promoting conceptions of the good in society. Second, allocative issues regarding how the state ought to distribute resources between different individuals and groups under its care.

Both these issues seem particularly central to an ethical analysis of health incentives, and to assessing whether this form of intervention falls within the state's legitimate sphere of action, with acceptable consequences for society. Initially at least, incentives look like a way of promoting a particular way of life - a 'conception of the good' - which may be controversial and fall outside the remit of state operation. Incentives also involve the provision of monetary resources to recipients, and so their distributive effects go further than more conventional healthcare interventions.

I argue that health incentives are not precluded on an account of state action which assumes moderate perfectionism to be justified. Health incentives, insofar as they operate as a form of healthcare intervention, promote an uncontroversial good (health) and do so without placing unacceptable pressure on citizens to accept that good. I suggest that the extent of perfectionism required to legitimate health incentives is no greater than would be required for many other interventions aimed at encouraging healthy behaviour, including those using information provision and persuasive arguments to influence people's behaviour.

In discussing distributive issues, I suggest how incentives might look on a number of different accounts of distributive justice. In the absence of empirical evidence as to the real world effects of incentives on resource distribution, I suggest that incentives need not be incompatible with the distributive principles discussed. These principles do not cover all of those proposed within differing accounts of political theory, but they represent a sample likely to be influential in the operations of western democracies.

EFFICIENCY

> Efficiency is like motherhood and apple pie; no one can admit to being against it.
>
> Harris (1997: 671)

Efficiency is often taken as an unalloyed 'good thing,' which we should always aim to maximise. Indeed, generally, the streamlining of a process to make it more efficient and to increase the outputs achieved from a set of inputs will be preferable to the alternative of failing to do so, and thus wasting resources. Yet it is not difficult to think of instances where an obsessive focus on efficiency and maximisation has resulted in undesirable consequences. Consider the production of animals and animal products for the food industry, where the pressure to be economically efficient has lead to the increasing intensification of farming and deplorable living conditions for some livestock: battery chickens provide a vivid example of productivity trumping animal welfare.

In the case of egg production, the pressure driving efficiency is, roughly, to produce as many eggs as possible. Clearly, efficiency needs to be balanced against other things of importance, such as animal welfare and product quality. It is also important to ensure the measure of efficiency (the outcome, such as number of eggs in this example) is the right one. For instance, simply maximizing egg production with little thought to housing conditions could result in poor animal hygiene and potentially poor quality eggs. In this case productivity, profitability, and thus 'efficiency' may fall.

The requirements of efficiency will depend upon the context: the aims of stakeholders, the market, the harms and benefits resulting from externalities, and so on. I may find my friend's method of making tea infuriating because she

always waits until the kettle has boiled before getting a mug and placing a tea bag in it (a job which could have been performed whilst the kettle was heating up, saving valuable seconds). Clearly, my concern is with minimising the time it takes to make the tea. Yet, if one of my friend's primary goals in this activity is to irritate me, then her actions are highly efficient.

Thus, the value of efficiency resides in the value of its net output. The value (benefit) of outputs must be balanced against the cost of inputs and cost (or disvalue) of other outputs (externalities). Efficiency is judged where this balance is optimal (no more valuable output could be achieved without suffering disproportionate disvalue elsewhere). Obviously, the correct balance will sometimes be fiendishly difficult to identify and disagreement is inevitable in some cases.

Healthcare is one such case. In part, this is because it is very difficult to measure the values and disvalues of inputs and outputs of healthcare, which has such far reaching effects, often obscured from view. Yet, more fundamentally, it is because we do not always know how to value these effects: how to quantify the good of health, the bad of disease, the worth of life, and so on. There is no obvious answer to this, and I do not propose to provide one here. Instead, I will provide a brief discussion of what is currently the most influential theory: that which proposes the value of healthcare lies in its ability to promote well-being. In order to permit measurement of value and comparisons between different interventions, the 'QALY' (Quality Adjusted Life Year) has been developed. This is a method of quantifying the value of healthcare and helping policy makers decide how healthcare resources should be rationed.

Well-Being and the Value of Health

The diverse benefits and harms associated with healthcare, health, and disease make it difficult to cross compare different actions and interventions. Those tasked with making such evaluations seek to discipline these disparate values and disvalues, converting them into a more manageable format. Pragmatically, then, it is useful to conceive of a single 'master value' which can act as a universal currency and the single outcome of interest. Some propose that well-being meets these criteria. Philosophical theories which propose that well-being is the only non-instrumental holder of value are described as *welfarist*.

Well-being, on its most general characterisation, is whatever it is that makes an individual's life 'go well.' This may involve pleasurable experiences, having one's desires satisfied, achieving certain objective goods, flourishing in life, or other objects, experiences, and relationships. A great many theories have been developed, which place different emphases on potential contributors to well-being (recent notables include Griffin [1986]; Sumner [1996]; Kraut [2007]; and Haybron [2008]).

Some version or other of well-being is increasingly becoming the dominant value in healthcare. At least, well-being is generally taken as the most *distal* value, which may still be promoted through the promotion of other, more proximal values (pain relief, life preservation, and so on). Healthcare is no longer (if it has ever been) only about curing disease and alleviating pain, but rather, is about facilitating 'health and wellness,' such that agents can optimise their well-being.

In this case, well-being will obviously stretch beyond healthcare, as many other factors will influence how well one's life goes (safety and security, wealth, social relations, to name a handful). Thus, well-being lays claim to the attentions of state action across all departments.

Yet, there is plenty of scepticism directed towards this appealing vision of well-being as a master value. Some object that the very notion of well-being does not make sense, and that it is mistaken to think there is only one non-instrumental good. Following from this, the exclusive pursuit of well-being may be undesirable. It is often argued that there are many things of value which cannot be captured in terms of their contribution to well-being, including loyalty, trust, liberty, aesthetic beauty, and treating people justly. Even if we do *sometimes* act with the promotion of well-being in mind, it seems that most of the time, we do not. Scanlon (1998) argues that, when thinking about ourselves, we rarely act with the express intention of promoting our own well-being. It is only when we think about how we should benefit others that we sometimes use something like well-being as the primary goal.

Even if a single value like well-being is adopted, measurement may still be very tricky. For example, where well-being is said to reside in having one's desires satisfied (or one's *informed* or *rational* desires), it will be difficult to judge when this occurs successfully. Further, it will not be plausible for policy initiatives to closely track the desires of individuals in order to promote

subjective well-being, as collecting accurate information about this would be vastly labour consuming, as would finely tailoring policies to promote a whole range of individual interests. Policies must, to some extent, be broader than this, though they may have sensitivity for personal desires built into their structure.

Knowing what objective contributors to well-being will generally include will be useful. It is also reasonable to assume that a number of these things will be the same across different individuals. In fact, at the practical level, it may not matter too greatly whether we assume a welfarist or non-welfarist position about value. It looks like there will be significant overlap between those objective goods thought valuable because they instrumentally contribute to well-being, and those objective goods thought intrinsically valuable on non-welfarist accounts. In the case of healthcare, for instance, there is reason to think that well-being motivated and value pluralist policies will promote similar outcomes. Both are likely to prioritise healthcare that prolongs life, relieves pain, restores mobility, diminishes distress, and so on. Assumptions that such effects will be 'good' in some way need not rest on the truth of welfarism but on common sense understandings about what is valuable to human agents.

We might know roughly, then, what sort of things we want to promote through healthcare: what the outcomes of efficient action should be. The difficulty, of course, comes with the 'hard cases,' when trading off between different good and bad outcomes must happen, and where opinions about what is good and how good it is vary widely.

QALYs

Once we have decided what sort of outcomes efficient healthcare will provide, we still need a methodology for specifying the magnitude of the values of these outcomes to facilitate decision making. To this end, processes of translation and quantification may be adopted. In the UK, the organisation central to this process in healthcare is the National Institute for Health and Care Excellence (NICE).

NICE plays an advisory role in determining what treatments are provided by the NHS. It is worth saying a little about the processes of valuing healthcare at this stage to inform later discussion, although I shall try to keep this summary brief.

> Our evidence-based guidance and other products help resolve uncertainty about which medicines, treatments, procedures and devices represent the best quality care and which offer the best value for money for the NHS.
>
> NICE (2012)

One of the tools NICE uses in determining what will represent the 'best value for money' is the Quality Adjusted Life Year (QALY). Different treatments can be evaluated in terms of how many QALYs they are likely to contribute; that is, they are assessed in terms of both the number of years they will add to someone's life, and the quality of life the individual will experience during those years. This is an effort to capture what we value in terms of both the quality and the length of life.

However, the notion of a QALY, how they are calculated, and how they are put to work in determining healthcare provision is widely debated (Nord [1999]; Nord et al [2009]; Moller [2011]; Dolan and Kahneman [2007]; Drummond et al [2009]). Quantifying the value of healthcare interventions may involve 'objective' judgements about the potential for treatments to extend life and relieve symptoms, as well as procedures for eliciting people's preferences regarding living with certain conditions. These methods are effectively a way of calculating the well-being gains produced by an intervention. As we have seen, though, it is not clear that well-being is the only thing of value, and it may be inappropriate to take it as the sole aim of healthcare provision or government action more generally. Moreover, even if well-being were taken as the justified end of all state action, methods of quantification (and 'QALYfication') may not be good ways of measuring how much well-being an intervention creates.

Hausman (2006: 254) points out that asking people to choose between different health states is unlikely to generate well-being values: "Since people care about other things than their own well-being, and since their expectations are often incorrect, evaluating health states by preferences among health states will not coincide with an evaluation in terms of well-being."

We may discard well-being and argue simply for healthcare to follow preferences. Preference elicitation, however, is problematic for further reasons. Hausman argues that preferences about health states are nonbasic, and "reflect

fundamental values as well as judgments about the character and consequences of health states in particular environments… In measuring preferences among health states, health economists are eliciting complex value judgments, which may be correct or incorrect. They are not measuring gut feelings." (2006: 273)

Menzel et al. (1999) provide a useful overview of some further difficulties with QALYs. For instance, patients suffering from a particular disease tend to rate their quality of life as higher than those asked to predict what their quality of life would be were they to suffer from that disease (1999: 13). Further weaknesses include failure to properly incorporate the unique value of treatments which save life; the limited potential for actual patients to have increased health prospects (because of age and other features); the feeling that treatment should be prioritised towards those who are worse off generally; problems with discounting for future health gains and losses; and the maintenance of hope that treatment will be provided for rare illnesses or varieties resistant to the most commonly effective treatments (Menzel at al. [1999]).

Far from being a purely objective exercise of measurement, calculating QALYs (or any procedure that seeks to inform efficient healthcare provision) involves numerous value judgements. I described three elements of efficiency: specifying a conception of efficiency (what will count as 'effective'); calculating the contributions to efficacy made by different interventions; and judging how the state should act within a general requirement for efficiency, which must be balanced against other goods. I have provided some discussion relating mostly to these first two aspects. I hope that it is clear that there is much overlap between the debates involved in the construction of the notion of 'efficiency' and the methods of calculating the efficiency of healthcare interventions. Neither of these elements are benign and value neutral, nor incontestable in nature.

I do not have much by way of a constructive theory to offer here. QALYs seem unsatisfactory for some of the reasons mentioned above, as do efforts to value everything in terms of well-being. Perhaps a strategy, such as that hinted at by Hausman (2006), where the reasons supporting agents' preferences (rather than the 'raw' preferences themselves) are used to inform valuations of health states. It still seems hard to get away from the problem that it is difficult to predict how we will experience health states, and to take account of the very different situations of the individuals who will actually be in need of healthcare.

Interventions, such as health incentives, with wide-ranging effects will be difficult to evaluate, whatever the outcome measure is taken to be. Benefits will take the form of health (though these will be time-delayed, uncertain and variable) as well as contributing to other aspects of life such as ability to work, raise a family, enjoy leisure pursuits, and so on. Capturing all the effects resulting from a single intervention can be extremely tricky.

JUDGEMENTS OF 'INEFFICIENCY'

As described, 'efficiency,' in some form or another, arises frequently in the media criticisms of health incentives. It is worth noting that efficiency-based reasoning is also one of the most frequently employed arguments in *support* of health incentives. The argument of the critics is that incentives are a 'waste of money,' or that money spent on health incentives could be 'better spent' on alternative interventions. The contrasting, supportive argument is that health incentives represent 'good value of money.' It is important to consider what such claims may relate to. My intention here is a philosophical one of providing a rough framework that judgements about the efficiency of incentives could fit into, rather than a sociological one of analysing what judgements people *actually* are making about incentives.

Incentives may be judged as (in)efficient in relation to interventions aimed at combatting the same sorts of disease, or in relation to healthcare interventions targeted at different diseases, or in relation to government spending more generally. Efficiency judgements in each of these cases will involve very different sorts of reasons.

In the simplest case, where comparisons and judgements of efficiency are made most easily, one incentive scheme may be judged as more or less efficient than another scheme. A scheme that offers cash rewards may cost the same to run but result in greater uptake than a scheme that offers grocery vouchers, for example.[8] In this case, the efficiency claim may be relatively straightforward: controversies about how cost-effectiveness analysis is conducted (such as QALY

[8] This is imagined, and further factors than simply cost of running the scheme and number of people taking it would would need to be considered to judge efficiency even in this simple case.

calculation) are less sharp, because the disease and intervention are very similar, and the outcomes can be judged in the same way.

Slightly more complex would be the comparison between an incentive scheme and a different sort of intervention targeting the same disease. For instance, an incentive scheme offering rewards to individuals who receive the HPV vaccination may be compared to an informational intervention which sends letters to individuals in the target population, outlining the risks of HPV and encouraging them to receive the vaccination. Here, one might measure the uptake of the vaccination and the cost of running the intervention in each case, and make a judgement about which is more cost-effective. Still, the calculation and comparison seems reasonably straightforward, though the informational campaign may offer benefits aside from increased uptake of vaccinations, such as increased understanding of sexually transmitted infections like HPV.

When the outcomes being assessed begin to differ, efficiency calculations become much harder and more controversial. When the benefits resulting from kidney transplantations must be compared to the benefits of providing a particular drug to manage type II diabetes, or the emergency treatment of a child hit by a car, different values begin to conflict. To describe one of these treatments as 'more efficient' than another is not a straightforward claim in the least. Presumably, few people would forgo providing lifesaving treatment to a child in order to divert more money into slightly cheaper diabetes medication, even though it would be plausible for a well-being maximising strategy (such as by calculating QALYs) to require the latter treatment be favoured.

Arguments that incentives are 'inefficient' could represent a comparison being made between health incentives and a closely related healthcare intervention. This 'like with like' comparison may enable reasonable efficiency-related judgements. I cannot provide further evidence which could inform this (largely calculative) claim. At this stage, it is not established whether or not incentives represent a viable form of health intervention. Some research shows effectiveness, but the effects are sometimes small and the long-term effects are not known (Sutherland et al. [2008]; Ranganathan and Lagarde [2012]; Burton et al. [2010]; Lussier et al. [2006]; Prendergast et al. [2006]; Cahill and Perera [2009]; Paul-Ebhohimhen and Avenell [2007]).

Yet, as efficiency judgements move away from comparing very similar interventions, there is more scope for debate about the concept of efficiency being used. Different interventions will have different outcomes, both intended and unintended, and these will impact health, well-being, and other things in diverse and often unpredictable ways. Debates about how we should value these different outcomes become highly involved and a greater understanding of the issues is required to judge relative efficiency.

There is a great deal of variety in the format incentive schemes may take, relating to the behaviour targeted for change, size of incentive, type of incentive, frequency of payments, length of scheme, provider of incentives, and so on. These can all affect the costs and outcomes of a scheme and so the scheme's efficiency. Non-comparative claims about 'inefficiency' will be difficult to support unless there is clear evidence that incentives generally have little or no impact on behaviour, or tend to have negative effects on health (not supported by current evidence).

What, then, should we make of claims in the press that incentives are 'inefficient' and a 'waste of money'? It seems probable that most such claims will not be based upon an excellent understanding of the outcomes of various sorts of incentive schemes, how such outcomes should be valued, and how incentive schemes then compare to other healthcare interventions or examples of state intervention. This is not because of an inability to grasp the importance of such features, but rather, because when people express their opinions they are rarely well-developed in this way.

More likely, heuristics will play an important role in these kind of judgements. This could be as simple as some form of backward reasoning: asked what I think of the state paying incentives to encourage healthy behaviour I may think 'I don't like the sound of that.' Rather than expressing this in terms of the intuitive dislike which it is, I may grasp at an explanation: saying 'I think they're a waste of money' comes easily to mind.

The process may be more considered: I may think of the sorts of things I associate with healthcare interventions (things that seem likely to change behaviour). Of particular importance may be what I think is likely to change *my* behaviour. Because I think of myself as a sensible, rational human agent I expect that reasons and information about the risks of unhealthy behaviour will

influence me, even though evidence suggests it has little impact on lifestyle (WHO [2008]). Given the medical context I may expect 'medicalised' interventions to be effective (pills, injections, operations). I may also think the right attitude is very important: wanting to change one's behaviour, caring about one's health, and having the willpower to resist temptation.

None of these heuristics I have imagined are likely to lead one to think that incentives will be effective. Other heuristics may provide more optimism about incentives: the idea that people frequently act in order to gain money, for instance. It does not matter too much for my point what heuristics act, and in which directions, but merely that it is plausible that heuristics are influential here, rather than a good understanding of the forces contributing to (in)efficiency. Thus, many of these judgements may not be based upon good reasoning and familiarity with the evidence.

I think there is also another important factor to consider when trying to understand the claims about inefficiency in the media. I have suggested an effort is made to say something about calculative efficiency, and that this will often be a poor estimate because people do not typically have access to good information about the effects of healthcare interventions, and because poor markers of efficiency may be substituted in the absence of this evidence. It also looks like people could substitute judgements about the third element of efficiency - how it should be used to guide action - into the place of evidence. Where the response to incentives is 'I don't like the sound of that,' the next logical move is 'I don't think they should be used.' Thinking that incentives should not be used may be adopted as a heuristic for their being inefficient or ineffective in some way.

Many off-putting and salient images exist around incentives (often printed alongside media articles): greedy fat people eating too many hamburgers and promiscuous youths too lazy to take sexual health tests. Certain diseases have historically been wrapped up with stigma and a lack of understanding, particularly when they tend to affect certain (already stigmatised) minority groups, or when they are associated with activities that have a negative image. Value judgements about how we should behave, what obligations the state has towards its citizens, and what kind of healthcare interventions are appropriate, can be important considerations when assessing incentives and healthcare provision more generally. However, the fact that such

judgements ought to play a role in evaluating efficiency does not mean that instinctive judgements, based on little understanding of the relevant issues and evidence, will be helpful.

Next, I want to discuss the extent to which it is appropriate that the state interferes with people's lifestyles in an effort to improve health, and I frame this in terms of debates about perfectionism and state neutrality. This begins to move away from what we think of as 'efficiency,' but I include this discussion here as health incentives typically seek to influence aspects of people's lives that operate 'outside of medicine,' as well as targeting people who are currently healthy, in order to prevent ill health. I think the questions of how we should value health and healthcare, and what the government ought to promote in providing healthcare for its citizens are intimately related.

The Bounds of Healthcare Provision and the State

To consider efficiency fully requires looking beyond a narrow focus on the calculative aspects of cost and effectiveness. Once we try to understand what makes efficiency desirable we begin to grasp how important a theory of the value of health and healthcare is. Further, we need to know how the state ought to promote these values, and what the limits of state action are. In the UK, the majority of healthcare is provided by the NHS, and so the question of the legitimacy of health incentives in part *just is* the issue of whether it is legitimate for the state to provide health incentives. The accepted aims of the state will inform the role of a state-provided healthcare system, and will define the limits of state action.

The limits of legitimate state action are constantly up for debate, with more libertarian thinkers pushing for a smaller government with less involvement in people's lives, whilst proponents of a more interventionist role for government arguing that sometimes the liberty of individuals needs to be limited to ensure a better life for others (often, the worst off). Concepts of liberty vary and incorporate sufficient complexity so as to allow similar interventions to be considered both liberty-constraining and promoting. Thus, simply adopting a broadly libertarian or liberal position does not fully determine one's position on particular interventions. This can be helpful, as Sunstein (1998) points out, as it allows those with apparently incompatible positions at the theoretical level to nonetheless agree on practical actions.

Individuals may or may not have a 'right' to minimal healthcare provision, though such provision may be justifiable on other grounds (see Buchanan's [1984] discussion). In the UK, healthcare has been provided free at the point of use to citizens through the NHS since 1948, initially funded by National Insurance payments, later by all tax paying citizens through general taxation. A system of state provided healthcare allows for nationally coordinated healthcare provision which enables strategies such as countrywide immunisation programs to be effected and other large scale public health interventions. Such programmes would probably be more difficult to achieve were healthcare provided by independent private organisations.

Socialised healthcare also protects the worst off in society, providing access to healthcare according to need rather than ability to pay. The existence of the NHS, and the provision of some degree of healthcare is generally well accepted in the UK, and the NHS is often treated as a central component of British cultural identity. However, the amount of money that should be spent on the NHS, the extent and range of healthcare provided, and its management and structure is much contested. Maintenance and reform of the NHS is often a hot topic for party political debate.

Healthcare providers are increasingly having to confront the harm posed by diseases resulting from lifestyle choices. What someone eats for breakfast; whether she cycles to work or plays sport; how many cigarettes she smokes each week; if she practices safe sex; decisions she makes about getting vaccinated or having her children vaccinated; where she lives; what job she does; are just a few of the things that can impact on an individual's health outcomes.

Once somebody is suffering ill health (partly) as a result of choices she has made about her lifestyle - for instance, someone who develops type II diabetes due to obesity or injures her leg in a skiing accident - then healthcare will be provided to restore her to health so far as possible. Healthcare interventions at this stage are relatively uncontroversial (although there will be plenty of debate about the best sort of healthcare to be provided, as well as debate over care given to individuals 'blameworthy' for their ill health; see Chapter Five). It is often preferable to prevent ill health in the first place, both for the individual, those her health impacts upon, and the healthcare providers.

The nature of a lot of the risk factors for chronic disease means that they are bound up with what appear to be personal decisions an individual makes about how to live her life. It is often considered inappropriate for the government to interfere with such areas of people's personal lives. And yet, altering the choices people make about lifestyle is a logical point at which healthcare could operate. The current government's push for more localism extends to the provision of healthcare, including measures targeted at reducing preventable disease. Some of this responsibility will be devolved to "a new, dedicated and professional public health service": Public Health England. Due to begin operating from April 2013, Public Health England will oversee and provide guidance for local governments' efforts to improve public health. (HM Government [2010: 52])

It seems naive to suppose that 'healthcare' should be restricted to hospitals and GP surgeries, when some of the most important healthcare interventions may take effect as part of citizens' everyday behaviour. The assumption that the state should stay out of people's decisions as to how to lead their lives looks difficult to reconcile with healthcare interventions intended to alter those very lifestyles.

There are two significant objections to interventions seeking to alter lifestyle in order to improve public health. The first of these is the threat to individual liberty posed by intrusive interventions; the second is the charge of perfectionism that might be leveled at a state which seeks to promote a particular form of lifestyle over others. These concerns are connected, and I shall discuss them both below.

LIBERTY

In considering liberty, we might consider both the imperative that liberty as an *intrinsic* good be safeguarded, and that liberty might be *instrumentally* valuable in other ways. This could alter the way in which the promotion of liberty is approached. If liberty is valuable as an instrumental contributor to well-being, self-expression, respect, and so forth, it will be wise to accept some trade in liberty when doing so will better promote those other goods, and it will not be worth sacrificing things like well-being in the name of liberty. If, however, liberty is an important good on its own then trading it for well-being and other goods may be counter productive.

Unless we adopt a strong welfarist position or similar, it seems most sensible to assume that liberty is both instrumentally and intrinsically valuable. Having the freedom to act as we like (for the most part) is ordinarily something we strive for and would wish to protect, even where that means we may experience less pleasure, poorer health, and reduced well-being overall. Liberty is also valuable instrumentally, as it will often contribute to our ability to experience pleasure and so on. Neither treating liberty as *merely* an instrumental contributor to well-being, nor as a non-tradable (perhaps master) value, will provide a satisfactory account of how we ordinarily value liberty.

It is also worth noting the important distinction drawn between 'positive' and 'negative' liberty, as described by Isaiah Berlin (1969). Whilst negative liberty covers freedom from interference, positive liberty is a richer notion that describes the ability of an agent to control her own life course and act in line with her considered reasons. This latter, positive form of liberty, overlaps with autonomy, although distinctions can be drawn between the two concepts. In this section, I will be more concerned with the former, negative type of liberty that is enabled when individuals are left free from external interference from other agents. Positive liberty will be more important in the next section in which I discuss the importance of debates about perfectionism and neutrality in political philosophy to evaluating state healthcare interventions such as incentives. The two forms of liberty are clearly linked, however, and so whilst being aware of the different ways in which liberty may be important (and in which its importance may be limited), I shall not press this distinction too hard.

The extent to which a particular healthcare intervention affects liberty requires careful consideration, as it may simultaneously promote and detract from it. Further, because liberty contributes instrumentally to many other goods, impacts on liberty can have wide-ranging effects: sacrifices in liberty may be acceptable in order to protect or promote some other good.

The Nuffield Council on Bioethics has designed an 'intervention ladder' to illustrate progressively more intrusive healthcare interventions. The ladder begins with "Do nothing or simply monitor the situation" and progresses up to "Eliminate choice" at the top (Nuffield Council on Bioethics [2007: 42]). For each 'rung' on the ladder, interventions become increasingly intrusive, and more impactful on individual liberty.

Using incentives comes about halfway up the ladder, implying that a greater degree of benefits to the individual / public must be conferred in order to justify incentives as opposed to doing nothing or providing information, but not as much good is required as would be needed to justify interventions that would restrict or eliminate choice.

This seems, however, a rather simplistic way of considering the relationship between intervention, liberty, choice and justification. Intrusiveness, and specifically impact on liberty, may not map cleanly onto the different categories presented here. Consider: a pharmaceutical company may reduce production on a drug that has been surpassed in effectiveness by another drug, meaning the option of taking the first drug is less available to me. Restricting or eliminating this choice (provided the new drug works as well or better for me) is not intrusive or harmful to my liberty, as I am left with an equally valid alternative action. Further, if the old drug was found to have dangerous side effects then removing it from the market would protect me from something that could harm my liberty.

Restricting choice may also just appear unimportant. Whilst it matters very much that I am able to access drugs that work, reducing the number of types of shampoo from which I can choose seems less important in terms of freedom just because the choice is less significant. It will matter more if I end up with a very restricted choice, but so long as there are still plenty of shampoos available to me it doesn't seem to matter a great deal (since I don't have particularly strong feelings about shampoo). Thus, the relationship between freedom of choice and valuable freedom is complex. More choice does not always mean more (valuable) freedom.

One worry about incentives is that they might coerce the agent by making her an 'offer she can't refuse.' This could infringe upon liberty if the state was strongly pressuring agents to act in a particular way through the use of incentives. A slightly more subtle infringement of liberty could arise if incentives somehow cause the agent to act in ways not consistent with her view of herself as self-determining, or in accordance with her preferred preferences, or in some other way that causes cracks to appear between the agent's will and her actions. Methods that manipulate or seduce agents to behaviour in particular ways may have these effects.

In the next chapter I consider the capacity for incentives to influence agents in these ways. I consider coercion, persuasion, manipulation, and 'nudging' as ways of influencing behaviour and discuss the extent to which incentives might employ these methods, and how problematic that is in relation to individual liberty and autonomy. Liberty, in the sense currently under discussion, is most under threat where the agent is coerced (as coercion severely restricts the agent's capacity for free choice). Yet, as I argue, it is unlikely that incentives would coerce, and there are steps which could be taken to avoid the risk of coercion.

If incentives do not force, coerce, unduly pressure, or make unavailable in some other way, valuable alternative choices for the agent, then they should not be regarded as freedom limiting in a strong sense. In a weaker sense, if incentives are effective in changing behaviour then they have altered something about the agent, and so are at least mildly interfering with the agent's actions (and / or her psychological state). Again, these affects on individual psychology, preferences and behaviour are taken up in the next chapter.

I have already discussed how efficiency will only be a suitable goal of state action if an appropriate concept of what will count as 'efficient action' is adopted, desirable outcomes are carefully selected, and useful ways of measuring the value (and disvalue) of outcomes can be developed. Liberty, in both its political and personal senses will play an important role in determining state action: what the duties of the state are within its commitment to provide an efficient healthcare system to its citizens, and what the limits of its actions should be.

The next sections will consider how incentives fit into debates about perfectionism, neutrality and state action, and the distributive effects of using incentives to encourage healthy behaviour.

Perfectionism and State Neutrality

Perfectionism in politics involves the development of an account of the human good, and making the promotion of this the legitimate aim of state activity. Within this general description of a political theory, perfectionism has many variants. These correspond to different conceptions of the human good, the way

it might best be promoted, the appropriate balance between the good aimed for and other important principles, and so on.

Perfectionist politics has been criticised as elitist, as there is a tradition within perfectionism which seeks to develop just a few of the most extraordinary examples of human achievement, to the neglect of lesser degrees of 'perfection' in a greater number of individuals. Further, goods taken to be worthy of promotion tend to be those virtues or ways of life exhibited often by the minority of individuals who have enjoyed relatively fortunate circumstances in life. Hence, perfectionism is sometimes perceived as deeply inegalitarian. However, perfectionism need not require resources to be directed towards the greatest achievements of human good; satisficing, egalitarian, and prioritarian accounts of perfectionism may represent opportunities for perfectionist political theory to preferentially direct resources towards the worst off.

Those opposing perfectionist state policies uphold that states ought to obey a principal of neutrality, and not favour specific conceptions of the good over others. This may require states to intervene to protect those whose particular conceptions of the good are disadvantaged by the way things are, or to refrain from interventions that would promote certain other conceptions of the good. Neutrality is defended as the only way of appropriately allowing citizens to be respected as free agents, and to avoid unfairly advantaging or disadvantaging agents on the basis of diverging conceptions of the good life.

Perfectionism and neutrality indicate legitimate spheres of action for a state. 'Lockean libertarians,' such as Nozick (1974) (as described by Scheffler [1992]), insist that this is very small, allowing only for the provision of basic rights and protection from harms. Defenders of neutrality could allow a more interventionist state, if it is considered appropriate that the state intervene to ensure fair opportunities are available to those with differing conceptions of the good life. So, the adoption of a perfectionist / neutral position on state action does not necessitate a more / less interventionist state.

Different grounds have been offered for the principle of state neutrality, including but not limited to the need to show respect for persons, to facilitate personal responsibility, to protect and promote autonomy, to create a neutral ground for debate, and as the only sensible way to behave given our ignorance about what form of the good life is best. Opinion is still divided: many philosophers have discussed the necessity for some degree of state neutrality for

a liberal society (see, for instance, Dworkin [1985]; Larmore [1987]; Rawls [1988]; Ackerman [1980]), whilst others argue neutrality is not appropriate (Wall [1998]; Raz [1986]; Sher [1997]; Mason [1990]).

The difficulty for supporters of health incentives could arise from either neutrality *or* perfectionism. If neutrality is defensible and the use of incentives to encourage healthy behaviour represents an objectionable instance of non-neutral state intervention, then incentives will be illegitimate. However, a theory of perfectionism may object to incentives on the grounds that they do not promote the correct conception of the good life.

I find convincing those arguments which reject highly demanding formulations of state neutrality and endorse moderate versions of perfectionism. Such theories tend to point out a number of problems with strong notions of neutrality, whilst insisting that a highly interventionist, freedom-limiting state that promoted controversial conceptions of the good would be extremely undesirable and illegitimate (Wall [2001]; Mason [1990]; Chan [2000]).

One problem with neutrality is that of justification: if neutrality is grounded in a principle of liberty, or autonomy, or respect for persons, as it often is, then it seems that the neutralist must acknowledge these are 'goods' worth promoting. In acknowledging this, the neutralist has already adopted a minimal degree of perfectionism. This does not seem wrong, and the moderate perfectionist would point out that things like liberty, autonomy, and respect are uncontroversial goods, and that it is right for the state to promote them.

Further, one can argue that the duties of the state go beyond the mere positive responsibility for any harm it causes, but it is also negatively responsible for harm it allows through not acting when it might easily have done so (Wall [2001]). This deserves a much more nuanced discussion than I can provide here, but it seems that the neutralist would have to make some argument for why the prevailing environmental conditions should generally be left unaltered, given that freedom of agents might just as easily be undermined through non-intervention as through intervention.

The difficulty, then, is to select those goods which contribute uncontroversially to all conceptions of the good life, and then to promote them

to the exact extent that people are free to pursue their own specific conceptions of the good, aided but never hindered by state action.

This is probably impossible, but we can at least aim for an approximation of such a state. Given that there will always be disagreement between reasonable people about the specific features of a good life, it is vital that the state not adopt anything that looks like a 'comprehensive doctrine,' which specifies and ranks all of the different contributors to a good life in terms of importance. Rather, it may promote certain agency and prudential goods, whilst not dictating what individuals' preferences ought to be when making comparative valuations between these goods (Chan [2000: 11-14).

> *Prudential* goods contribute to the good life, and include things like having a supportive relationship with one's family or deriving pleasure from sporting activities.
> *Agency* goods are features of the agent, virtues and dispositions, which constitute the good life, such as honesty, respect, courage, and so on. I believe we can add being in good health to the list of prudential goods.

The provision of health incentives fits with this moderate version of perfectionism: health is an uncontroversial good which contributes to (almost) all of the many and varied conceptions of the good life; the state is both positively responsible for the impact on health of its actions, as well as negatively responsible for certain health-related consequences of its failure to act. It will, however, be up to individuals to determine how much weight to assign to different goods, and how highly to value health in relation to other prudential and agency goods in determining what sort of a life they wish to lead.

At this level of generality, health incentives (along with other non-coercive healthcare interventions) seem consistent with the state fulfilling its responsibility to protect and promote the interests of its citizens, whilst remaining sufficiently neutral as to respect all individuals properly as self-determining, independent agents. But what about when we consider the specifics of incentives: offering money to people who quit smoking; grocery vouchers to those who lose weight; lottery tickets to everyone who receives a certain vaccination? Perhaps these schemes overstep the mark and promote

specific ways in which people can achieve good health that deny them the respect to determine such things for themselves.

I suspect, in discussing behavioural interventions such as health incentives, there is a tendency to overestimate the significance of some of the lifestyle behaviours which incentives seek to alter. That is, there is a danger of exaggerating how significant to someone's sense of self-identity the composition of her typical breakfast is, or that she tends not to get tested for STIs. Although many of these behaviours (or decisions about these behaviours) happen in the privacy of our own homes, it is not necessarily the case that they are therefore fundamental to who we are as people, and that altering them is to alter something important about ourselves.

This will not always be true: there will be examples of gastrophiles, dogmatic smokers, principled objectors to vaccinations, and so on. Such people could perceive certain of their lifestyle behaviours as fundamentally important to who they are and their own efforts at pursuing their conceptions of the good life. In these cases, perhaps the state ought not to encourage alternative ways of living.

Health incentives have two significantly redeeming features here: first, they are almost certainly non-coercive, nor influential in liberty or autonomy-incompatible ways (as I shall argue in the next chapter). Thus, incentives can be refused with relative ease, and those for whom certain (incentive-promoted) behaviours are incompatible with their conceptions of the good life are unlikely to be much disadvantaged. Second, incentives which take the aid-type rather than trade-type form are supportive of individuals' own conceptions of the good life (or at least, with how agents want to behave, even where this is not a significant aspect of their conceptions of the good life). If it is possible for incentives to be restricted to this aid-type form, and targeted at those who would welcome them, incentives are unlikely to undermine individuals' conceptions of the good life.

These features mitigate the negative impact offering health incentives could have on individuals whose lifestyles aren't compatible with the conception of the good life favoured. Living in a society which values a conception of the good different from one's own may, yet, be disadvantageous. As such, it is important that perfectionism is *moderate* so that the imposition on those individuals whose conception of the good differs is not too great. The

uncontroversial nature of health as a valuable good also helps to deflect such concerns.

Some will disagree with the position of moderate perfectionism which I have adopted, and used as a reference by which to evaluate incentives. There are arguments to be made that the state ought to remain more neutral between conceptions of the good life than health incentives allow, or that perfectionism is acceptable (demanded, even), but incentives promote the wrong conception of the good life. I do not think either of these arguments will be very promising unless a radically different structure of the state is adopted to the one we currently accept, and a number of ordinary intuitions about what counts as an acceptable healthcare intervention are rejected.

In relation to the claim that incentives are not neutral enough, consider the difference between incentives and information provision and medical advice. Both information provision and medical advice can and are used to encourage healthy behaviour. Insofar as there is some 'healthy behaviour' which is promoted, clearly, the state must adopt a non-neutral position. Taxation on 'unhealthy' products and activities also acts so as to disadvantage those who engage in them, and benefit those who avoid them. It would seem that, were we to endorse a stronger version of neutrality that insisted healthy behaviours (as identified by the state in conference with healthcare professionals) must not be encouraged, then we would have to reject a whole swathe of other interventions.

Alternatively, it might be argued that incentives are not in line with certain forms of perfectionism. I am thinking here particularly of those more traditional, Aristotelian theories which place an emphasis on the development of a virtuous character. Incentives, if they bypass mechanisms of development whereby individuals exercise willpower, restraint, commitment, and effort in applying themselves to the task of making healthier their lifestyles, may preclude agents from developing properly virtuous characters. This kind of criticism relates to particular conceptions of how one develops a moral character (see Burnyeat [1980]).

I think suspicion is due where arguments assume 'doing it the hard way' is preferable, and where the process is presented as more important than the successful outcome. First, this kind of perfectionism seems open to the objection that it is promoting a more controversial good than mere health,

specifying how one ought to act and what motivations one must adopt in seeking to achieve good health. Further, it is likely that such an argument is seeking to promote behaviour that many individuals will simply not be able to achieve: should we deny those who wish to lose weight access to support groups, low fat foods, discounted gym membership, and fitness apps that might help them to succeed on the basis that it will be more beneficial for willpower development if these aids are absent? We must accept our own psychological structure which means that the vast majority of people will be more successful in achieving lifestyle changes if they focus on short term, achievable goals, and receive immediate rewarding feedback when they meet these goals.

Aid-type incentives, targeting as they do those who wish to alter their behaviour but cannot do so without assistance, do not obstruct people from developing good characters. Individuals who *can* (and wish to) adopt healthy behaviours without incentives will do so, whereas those who would fail are supported by incentives (where they are effective) to succeed. It is also worth noting here that 'self-efficacy' - the extent to which an agent believes she is able to alter her behaviour - is important in whether or not people succeed in behaviour change attempts (Bandura [1978]; Bernier & Avard [1986]). Thus, success (whether aided by incentives or otherwise) could help to develop an agent's sense of self-efficacy, and the cumulative effects of succeeding in smaller goals can enhance the likelihood she will succeed in larger, future goals.

The provision of incentives is not a 'quick fix' for lifestyle behaviour change. A good deal of effort will still be required for agents who have developed unhealthy habits to alter them. We should not describe the use of health incentives as 'bypassing human agency,' 'denying the opportunity for moral maturation,' or 'infantalising' people, when such criticisms are based on a mistaken conception of the influence of incentives and human psychology.

A final response to charges relating to perfectionism and neutrality might be proposed: that offering health incentives does not qualify as promoting a particular conception of the good life.

If this claim is supportable, then incentives cannot be criticised on the grounds that they are offensively perfectionist. However, I shall not discuss this here. I believe there is enough to be said in defence of health incentives as part of a moderately perfectionist state to render this argument unnecessary. Let it

suffice to say that the definition of 'conceptions of the good' and the identification of policies which would count as promoting such conceptions is disputable. Health incentives may, for instance, be considered as promoting a particular conception of social justice (if healthcare is considered an area of state action governed by justice requirements). States are neither able nor required to remain neutral on issues of social justice, even according to liberal theories espousing neutrality with regards to the good life (Nagel [1991]). Such a move rather shifts the problem to the realm of what state policy is justified in order to establish a just society, and objections to incentives and the promotion of healthy lifestyles might simply be reformulated so as to attack incentives within this domain.

Distribution of Healthcare Resources

As discussed, healthcare provision that seeks 'efficiency' by maximising a single outcome (such as QALYs) may go very wrong. I have discussed how state action must be sensitive to particular limitations and requirements, and that we must engage with questions about the extent of appropriate state interference in order to know what 'efficient healthcare' will involve. Liberty in both its psychological and political senses is important here, and in this chapter I have focused on the debate relating to perfectionism and neutrality.

Another centrally important feature of political philosophy as relates to healthcare is the just distribution of resources. Maximising strategies will fall prey to the distributive criticism of utilitarianism, as such theories ignore the personal aspect of value: it matters not only that there *is* good, but it matters *who* has that good. The ability for health incentives to combat health inequalities will be an important consideration in terms of their overall desirability (Health England [2009]).

It is not obvious what a fair distribution of healthcare will involve. Some might argue that it is important that agents take responsibility for their own behaviour, and that healthcare should only be provided where it is deserved in some sense. This is a commonly raised criticism of incentives, which are perceived as 'rewarding bad behaviour.' I consider this concern for fairness in the next chapter, where I also discuss some relevant evidence relating to the link between socio-economic factors and health outcomes. There are a number of other principles that may guide distributive justice, for instance, the Difference

Principle proposed by Rawls (1971/1999), principles proposed by various formulations of Egalitarianism, Utilitarian (maximising) strategies, Libertarian (non-interventionist) principles, and others.

The framework of the healthcare system, its methods of delivery, and the sorts of interventions provided will determine the allocation of healthcare (and health) amongst the population, as well as factors about recipients that influence their uptake of healthcare resources. Health incentives may have specific effects on this distribution, particularly as they not only affect the distribution of health, and the economic benefits that good health may provide, but they also directly deliver economic benefits to agents in the form of cash, vouchers, lottery tickets, and so on.

Rather than discussing in depth the virtues and drawbacks of different principles of distributive justice, I will discuss how incentives fit with a selection of different principles of distribution. This discussion will inevitably be limited in scope, as I do not wish to spend too long introducing each principle. Further, it will be limited by evidence: there is not sufficient information to understand exactly (or even approximately) how incentives will affect the distribution of healthcare and health outcomes in the population. I will thus focus on the general aims of providing incentives, to consider whether these are compatible with certain distributive principles, rather than the outcomes of incentive scheme provision. My discussion will thus be necessarily rather speculative and general. What I aim to show is that incentives could actually fit with a variety of accounts guiding the distribution of healthcare. Thus, incentives should not be considered excluded by most of these accounts, although that could change if they proved to have effects that contradict those principles.

Before beginning this discussion, it is worth pointing out that the problem of identifying the appropriate outcome and finding a good way of measuring that outcome extends into distributive debates. It will be necessary to carefully select the good that is to be distributed fairly, and to develop a method for assessing how this good is distributed, in the same way as these processes are required in order to judge effectiveness and efficiency.

Rawlsian Liberalism

The key feature of distributive justice in Rawlsian Liberalism is the Difference Principle (Rawls [1971/1999: 65-66]), which is intended to ensure that those

worst off in society (with the fewest resources) are prioritised when it comes to distributing resources. This principle asserts that inequality is permissible only where it is to the material benefit of the least advantaged. Thus, Rawlsian Liberalism expressly contradicts maximising principles like welfarism which permit the most advantaged to benefit further, even at the cost of the least advantaged (so long as there is an overall gain in well-being).

It is important to note that Rawlsian justice is a feature of *institutions*. The institution that treats people fairly (according to the Difference Principles, as well as Rawls's other principles of justice) will by definition produce a just distribution. The goods Rawls is concerned with are the 'primary goods'; "things that every rational man is presumed to want… the chief primary goods at the disposition of society are rights, liberties, and opportunities, and income and wealth." Rawls also includes self-respect as a primary good, along with "health and vigor, intelligence and imagination". These last four are, Rawls states, "natural goods; although their possession is influenced by the basic structure, they are not so directly under its control." (Rawls [1971/1999: 54])

Those institutions providing healthcare will impact on a number of these primary goods and hence will have a significant impact on distributive justice. Health incentives form one practice within the institution providing healthcare, so we must question whether their affect on the distribution of primary goods follows Rawls's principles of justice, in particular the Difference Principle.

Health incentives need not be explicitly directed at the worst off to the exclusion of the better off: people who have plentiful access to primary goods may still be eligible to receive incentives in return for quitting smoking, getting vaccinated, or any of the numerous other behaviours that could be incentivised.

However, there are reasons for thinking incentives would sometimes be consistent with a broadly Rawlsian distributive justice. First, the kind of unhealthy lifestyles incentives are deployed to change are more prevalent in deprived populations, and so healthcare is likely to be prioritised towards the less advantaged generally. Second, insofar as incentives target people lacking in health (or likely to lack health in the future) then they are compensating for a (real or potential) deficiency in that primary good. Third, incentives also provide monetary or quasi-monetary redistribution, and so where the first two points hold, the effect of incentives to direct more primary goods towards the least advantaged will be exaggerated. There may be further benefits resulting

from improved health, in that people may have greater opportunities, liberties, improved self-respect, and so on.

The extent to which such effects of redistribution and prioritisation of healthcare toward the least advantaged will result from the provision of incentives is unclear: there are many complex and interacting factors to take into account and a lack of empirical evidence (Voigt [2012]). It is possible that incentives could have a deleterious impact on inequalities. For instance, wealthier individuals typically make more use of healthcare services, and so may take advantage of health incentives to a greater extent than poorer individuals (Bains et al. [2000]). Such an effect is likely to be an issue for any healthcare intervention, however, and incentives are not exceptional here.

Incentives targeted toward poorer populations will be more likely to be consistent with Rawlsian justice. This may be achieved by explicitly excluding more privileged individuals from access to incentive schemes; by selectively promoting schemes in more deprived areas; or by incentivising behaviour change that is more likely to be relevant (and beneficial) to more deprived individuals. As mentioned, many incentive schemes will tend to selectively target more deprived individuals or those lacking in primary goods, because the unhealthy behaviours targeted are more prevalent in those populations (see Chapter Five for more discussion of these trends).

Those incentives which target behaviours which are only weakly linked to socio-economic factors may not help combat inequality. Also, there could be cases where incentives become available for a behaviour that more advantaged individuals would perform anyway (receiving a vaccination or disease screening, for example). Introducing incentives for such behaviours could benefit more advantaged individuals even without changing how they would behave anyway.

Luck Egalitarianism

Distinct from 'strict' egalitarianism, 'luck' egalitarianism seeks to equalise differences in *access* to resources, rather than between outcomes. Luck egalitarians argue that inequalities arising from preferences and choices of individuals are acceptable, whereas those arising from circumstances and luck are not. A variety of theories may be described as broadly 'luck egalitarian,' though many of the authors of these theories do not identify themselves as such

(prominent examples being Dworkin [1981a, 1981b, 2000] and G. A. Cohen [1989, 2001]).

On the luck egalitarian account, the state ought to compensate individuals for (economic) disadvantages they suffer as a result of bad luck, but individuals should bear responsibility for any disadvantages (or advantages) resulting from their own choices. Dworkin distinguishes between 'ambitions' and 'endowments' here, the former referring to the choices we make, the latter to things over which we have no control.

Considerable theoretical work has gone into the development of various accounts of luck egalitarianism: it is decidedly tricky to distinguish natural talents for which individuals are not responsible, and those which result from choices and application for which individuals are responsible, and for which they can lay claim to the resultant rewards. It is also difficult to see how a practical scheme of compensation could be developed. Despite these problems, luck egalitarianism is still appealing to many due to its efforts to track the intuition that disadvantage ought not to reflect brute bad luck. I will not go into further detail on these complexities here, but there is a plentiful and growing literature, some of which is directed specifically at the distribution of healthcare resources and health (see also Arneson [1990, 2000]; Sher [1997]; Segall [2007]; Dworkin [2003]; Cohen [1989].

How, then, do incentives look on a luck egalitarian account? The question of pertinence here is the relative contributions or mitigations incentives make to the distribution of resources according to endowments (or luck; things outside the agent's control) and ambitions (or choices; things within the agent's control).

It might initially seem that health incentives are doomed to fail according to the luck egalitarian project: they explicitly target individuals who are likely to suffer the negative consequences of their choices (the lifestyles they have adopted). Insofar as incentives direct more resources towards those who have made poor choices (and fewer towards those making prudent choices) then they are unlikely to result in a luck egalitarian distribution where an individual's resources reflect her ambitions but not her endowments.

In fact, as mentioned, it is sometimes very difficult to separate the consequences of luck from those of choice, and so to determine what should be compensated for and what should not. Luck egalitarianism is also a difficult

position to reconcile with many of our intuitions about healthcare. What, for instance, should we do about the woman who chooses freely to ski, aware of the risks, and then suffers an injury? Should we refuse to set her leg on the grounds that she 'chose freely' and must bear the consequences?

It is unlikely a luck egalitarian would insist on the neglect of victims of such 'option luck' (those who suffer the unfortunate, though foreseeable, outcomes of their choices). Segall (2007) provides a defence of criticisms based on this unintuitive aspect of luck egalitarianism, pointing out that luck egalitarianism can incorporate other principles to guide right action, such as solidarity, which will command that we empathise with the plight of the skier and help her out.

If incentives *do* sometimes benefit those who are only suffering because of their choices we might still, it seems, think this compatible with some varieties of luck egalitarianism. Moreover, individuals making bad choices may not be making free choices in the way required for them to be considered responsible (where responsibility would render them no longer entitled to compensation for the disadvantages they suffer as a result). I discuss the connection between free choice, responsibility, and desert in Chapter Five.

The Capability Approach

This account identifies valuable outcomes in the core concepts of 'capabilities' and 'functionings.' The latter are 'beings and doings,' including, for example, being in good health. Capabilities, in contrast, refer to the opportunities for these functionings, so having access to basic healthcare would be a capability for the functioning of being in good health. The motivating force behind the capabilities account arose from a particular concern with global justice and the conditions of deprived populations in developing countries, with the two principle architects being Nussbaum and Sen (1993) (see also Sen [1990,1999, 2009] and Nussbaum [2000, 2011]).

As with the luck egalitarian account, capabilities theory focusses on seeking to neutralise unequal influences on people's *access* to resources rather than on the outcomes (the actual distribution of resources). As such, the emphasis is placed on capabilities rather than functionings: what matters is that people have equal access to the different capabilities that may make their lives go well. On the luck egalitarian account, a concern with responsibility leads to

the requirement that the actual distribution of resources be controlled, as far as possible, by choice. On the capabilities account, the concern is with well-being and freedom: individuals must be allowed to follow their own life-plans to pursue functionings that will contribute most to their well-being.

Another important feature of capabilities theory is the prioritisation of those functionings which enable primary, or basic, capabilities. This makes sense, as certain capabilities are necessary for a whole range of other functionings. For example, lacking access to basic healthcare will make a whole range of other functionings inaccessible, such as access to work, opportunities to raise a family, and so on.

I have focused on the UK context of health incentives, mostly for pragmatic reasons, and the particular healthcare system of the UK has specific implications for the structure and delivery of incentives. This also narrows focus with regard to the kind of behaviour changes likely to be useful in the UK context, and the kinds of ethical concerns that arise in western democracies. Yet health incentives can and are used elsewhere in the world (see, for example Nglazi et al [2012]; Ranganathan and Lagarde [2012]; Giuffrida and Torgerson [1997]) Many of the issues from the UK context are transferable (obesity is of global significance, for instance). Yet, in the context of global justice, the provision of incentives to encourage healthy behaviour to relatively wealthy British nationals might be a poor use of resources, considering the benefits such resources could bring to those far more disadvantaged elsewhere in the world.

On a more localised reading, this theory demands that the most fundamental capabilities be ensured (invariably this will also mean targeting disadvantaged groups, as those are the populations most likely to lack such capabilities). As discussed, without a careful analysis of the empirical evidence (more of which is needed) it is difficult to judge to what extent incentives do preferentially provide healthcare to the more deprived.

An interesting feature of incentives, in terms of the capabilities account, is their disputed impact on freedom: incentives may be seen as consistent with ensuring agents are free to choose how to act, or they may be seen as compromising this freedom. This links into the discussion of liberty, perfectionism and neutrality earlier in this chapter, as well as the discussion of coercion and other influences on behaviour in the next chapter. Where incentives restrict choice and pressure individuals to act in ways inconsistent

with their own preferences, they will look undesirable on the capabilities account. In contrast, where they have the effect of supporting the agent in making well-reasoned decisions, and bringing her actions into line with her stable preferences, incentives will look much more appealing.

These opposing effects of incentives will depend not just on the theoretical account provided (how incentives can be 'spun'), but on features of specific schemes. It looks like incentive schemes which target individuals who already have a wish to alter their behaviour (for instance, smokers who wish to quit) will be more likely to act in this supportive role (in the manner of aid-type incentives). Incentives which are broadly targeted will be more likely to push some individuals in ways inconsistent with their wills (particularly in the case of trade-type incentives). Further, where incentives are at risk of having this latter effect, those which move beyond the limits of moderate perfectionism that I have described, and pressure people to adopt controversial conceptions of the good life, will be the least acceptable on the capabilities account.

Communitarianism

Communitarian values emphasise the situational ties of the individual to other members of her community, in contrast to the typically individualistic focus on much liberal political theory. Without dismissing influential values such as well-being, communitarianism generally proposes that seeking to promote individual well-being by reinforcing individual human rights will not achieve success. Rather than abstracting the interests of the individual from those of the social group within which she exists, communitarianism emphasises the importance of those values which contribute to the good of the group as a whole. This line of thinking encourages us to foster those values which encourage interdependence and fellow-feeling.

Often regarded as a response to the domination of Rawlsian liberalism, and the preoccupation with autonomy of much political and moral theory (see for example MacIntyre [1981, 1988]; Sandel [1998, 2012]; Walzer [1983]) Communitarianism focuses instead on values such as solidarity, mutuality, and reciprocity (Prainsack et al. [2011]). This move away from individual preference satisfaction towards the promotion of society's interests as a whole requires some shared understanding and agreement about what is adopted as the common good.

The welfare state is conceptually underpinned by these kind of communitarian values: Weale (1990: 477) argues that "the immediate occasion for the development of welfare states was not a concern with equality but, rather, with… social solidarity". Distributive forces that follow these initial motivations for the welfare state ought to fall into line with communitarianism.

There is a sense in which incentives might be open to criticism from the communitarian standpoint. This relates to the worry that those receiving incentives are acting in their own best interests, selfishly, with little thought to what impact it will have on society. If people who adopt unhealthy lifestyles do so with a callous disregard for the significant burden this will place on the NHS then responding to their behaviour with rewards and incentives may be inappropriate. This is clearly linked to criticisms relating to desert and responsibility (see Chapter Five).

However, evidence from medical sociology and health psychology suggests that social, environmental and economic factors have a large influence on lifestyle choices, and thus, on health (see Chapter Five for a fuller discussion; Marmot et al. [2010]; Marteau [2010]). The healthcare problem created by the prevalence of unhealthy lifestyles may thus be framed as a *social* problem, and not one of individual responsibility. Rather than increasing the pressure on the individual to 'make the right decision' and 'take responsibility for her actions,' the communitarian approach would be to seek to redirect these socio-economic influences away from facilitating unhealthy behaviours and towards healthier lifestyles.

Aid-type incentives start from a position of seeking to correct for internal and external influences that encourage unhealthy behaviours, helping agents who wish to alter their lifestyles to successfully do so. So far as this counteracts the pernicious effects of deprivation, whilst acknowledging the power of these effects (and, similarly, not expecting agents to demonstrate unrealistic levels of willpower), incentives may be compatible with communitarian thought.

For the most part, however, I suspect communitarians would be hostile to incentives. Sandel (2012), for instance, shows a good deal of suspicion towards incentives of various sorts. Much of Sandel's criticism uses a rhetoric of corruption and coercion (discussed in Chapter Six), but this can, in part, be attributed to a communitarian vision where incentives shouldn't be needed: individuals who are properly on board with a solidaristic system of healthcare

ought to remain healthy for reasons other than the prospect of personal gain. The problem with incentives, on this account, is that they imply the motivation is the financial reward itself, rather than the other benefits that will arise from adopting healthier habits.

In the case of trade-type incentives, where the individual is moved to alter her behaviour for the most part by the offer of an incentive, this objection may have some purchase. Yet the supportive, aid-type incentives need not be seen in this way, and may operate in a way perfectly compatible with a concern for the social good. Further, as we are considering the distributive effects of incentives, it is not clear that those who do not require incentives to adopt healthier habits (those with reasonably healthy lifestyles in the first place) do so out of feelings of solidarity or similar anyway. Thus, it is not clear that incentives distribute resources away from individuals with solidaristic, communitarian values, nor that they promote an individualistic ideal of healthcare.

A nuanced account of incentives can, I think, fit with a communitarian conception of how the state ought to distribute healthcare resources. However, I also think it would be overly optimistic to expect communitarian thinkers, on the whole, to accept incentives as consistent with communitarianism.

This section has sought to briefly outline a number of principles of distribution, informed by different political theories, and to consider health incentives in the light of these theories. My aim was twofold: first, I have explored some of the distributive effects incentives could have. Second, I have indicated how incentives might be consistent or inconsistent with the normative recommendations of these different theories.

For the most part, I have argued that incentives can be made compatible with all of the theories discussed here. Perhaps the greatest stretch comes when we consider the ideals of communitarianism, but even in this case I have suggested that a more nuanced understanding of incentives shows them to be not inconsistent with valuing solidaristic relations within a community. I do not think this result is a product of philosophical trickery, and nor do I propose that it is enabled by the lack of differences between these different political theories (they are quite distinct). Rather, it is because health is recognised as such an

uncontroversial value, and the state provision of healthcare is justified on all of these accounts.

There will be some political theories which deny that the state ought to interfere to provide healthcare to its citizens. In that I have neglected to discuss such theories I have left out a number of arguments that could oppose the distributive effects of incentives and position this form of healthcare as outside the sphere of legitimate state action. Given a starting position of accepting, in principle, state provision of healthcare, it is more productive to consider how those principles that would permit some healthcare interventions would respond to the case of incentives.

Another reason why incentives may be made to fit with a variety of principles is that they are not yet fully understood, both in terms of their mechanisms of action and their distributive effects. This discussion has been mostly speculative, and so it remains to be seen if incentives do, in fact, combat inequalities in health, well-being, capabilities, and so on, or if they reward individualistic, self-regarding action. Insofar as incentives can take many different forms, and may operate both as aids to psychological efforts to change behaviour and as trades to tempt behaviour change, it seems there is scope for different varieties of incentives to have different distributive effects, and so too have differing degrees of consistency with principles of distributive justice.

Concluding Remarks

In this chapter, I have sought to explore a range of political considerations relating to health incentives. Much of this has served more to indicate the degree of complexity involved in seeking to measure or predict the impact of healthcare interventions, rather than to confidently argue for an understanding of how incentives fit into this picture.

The initial motivation for this kind of analysis was the concern with 'efficiency,' both as a recognised concern of state action in general, and as a criticism frequently arising in the media analysis. I presented efficiency as involving different elements, all of which contribute to an analysis of any healthcare intervention. These were the development of a concept of efficiency (including what 'effectiveness' will mean, what outcomes will be desirable); a calculative method (whereby inputs and outputs can be assessed and cross-compared); and a judgement which situates these concerns in the wider political

context and acknowledges the kinds of action that are permissible for a state to undertake.

Tverdek (2004) argues that efficiency can be thought of in a stripped back, prudential sense, where it is concerned only with the question of *how* a particular end might be maximised. It seems to me that much of the discussion about the 'efficiency' or otherwise of health incentives (as well as healthcare more generally) is not concerned with such calculative questions, but rather with the 'what' and 'why' questions which prescribe the appropriate ends of action and tell us what efficient action will look like. My intention in this chapter has been to look at those concerns, rather than to address the calculative question.

I discussed the notion of well-being, and the way that such a concept is often taken as fundamental to calculating the effectiveness, and thus efficiency, of healthcare interventions. In particular, QALYs are often used as a method for quantifying and comparing the well-being produced by different interventions. I sought to indicate a few of the ways in which well-being and QALYs are imperfect tools for measuring the value of healthcare.

Much of the trouble with trying to evaluate the worth of healthcare interventions stems from the difficulty inherent in conceptualising just how health should be valued, rather than with the calculative problem. Although arguments expressed as 'efficiency' concerns may trigger associations with mathematical questions and solutions, it seems that the conceptual level presents us with at least as much work to do.

I argued that where criticisms directed towards incentives make reference to 'inefficiency,' this will rarely refer to uncontentious claims about efficacy and cost. Rather, more likely, such criticisms could be prompted by heuristics and will often be informed by moralised judgements about the value of such interventions. I do not intend to suggest such judgements are worthless: quite the opposite, I have argued that the negotiation of values when considering efficiency is vital. However, the 'gut instincts' expressed by individuals in the media and elsewhere, when relatively uninformed about the realities of incentives, and when apparently masquerading as pronouncements about calculative efficiency, will not be particularly helpful.

It seems that many prescriptive arguments about the appropriate ends of healthcare and state intervention are presented as prudential questions of how

to achieve maximum efficiency. Criticisms that incentives are a 'waste of money' or 'won't work' or that they won't convey lasting benefits appear to be empirical claims about the effectiveness and efficiency of this intervention. Yet, where these claims are made by lay members of the public in the media, and are not supported by empirical research into the effectiveness of incentive schemes, and where there is no clear framework for assessing 'efficiency' judgements, these arguments begin to look less like 'strict' efficiency claims and more like moralised arguments about the appropriate ends of healthcare and state intervention.

Efficiency is of importance when assessing healthcare interventions, and it is likely that the involvement of money in the case of incentives lends this issue greater salience when we are asked to consider the worth of such an intervention. Yet, often, concerns may be best expressed through discussion of things other than economic efficiency, and the resolution of these debates would result in greater progress than ongoing head scratching about efficiency is likely to achieve.

Hence, much of this chapter has been concerned with questions relating to what ought to be promoted through healthcare; what ought not to be interfered with; what the limitations on state intervention in people's lives to promote health should be; and what distributive effects the state should aim for. Well-being is of great significance when considering how far the state ought to go in promoting health, and some would be inclined to argue that it is the *only* thing that matters. However, concern for other goods - justice and liberty in particular - might constrain efforts to promote well-being.

In the context of the extent of legitimate state intervention, I have discussed how healthcare interventions, and incentives in particular, might be considered threatening to liberty. The next chapter will be of further relevance to this issue as it takes up the question of whether incentives might exert a coercive influence on people's decision-making capacity. Thus, the next chapter considers whether incentives can limit the psychological liberty of agents.

In this chapter, I focused on the issue of political liberty, and whether a government that promotes certain (healthy) lifestyles, especially through the use of incentives and rewards, might be considered as not acting according to the principles of a fair democracy. The debate here links to the ongoing

discussion within political philosophy surrounding perfectionism and state neutrality.

I suggested that a position of moderate perfectionism looks to be the most plausible, both in that it fits with commonsense attitudes towards allowing the state *some* degree of interference, but whilst protecting citizens from an overly enthusiastically interventionist state. Such a position permits the promotion of relatively uncontroversial components of the good life, such as health, but is cautious in its use of pressure in doing so. I argued that health incentives, along with many other forms of health behaviour change interventions, will be permissible on such an account, so long as individual liberty is respected. An understanding of the value of liberty is important here, as it is not a straightforward case of 'the more, the better.' Restriction of choice may, in some cases, not represent a loss in valuable liberty for the individual, and at times (where it helps agents achieve their long-term goals) it may promote some aspect of liberty.

Aid-type incentives, then, are likely to be supported by a conception of political legitimacy that permits moderate perfectionism. Not all feasible designs of incentive schemes will be so benign: incentives could be offered that strongly promote highly controversial goods in forceful ways, and these may be unacceptable. For the most part, however, it looks like health incentives will remain within these parameters.

There will also be other reasons for thinking that the state ought not to use incentives to encourage healthy behaviour - reasons which do not relate to liberty in the way that the neutrality principle does. Most accounts of political theory identify the state as responsible for ensuring a just patterning in the distribution of resources amongst its citizens (a notable exception being libertarianism, which focuses on the procedural justice of the free market, which will inherently produce a just distribution). On most accounts, then, healthcare and health are important goods, the fair allocation of which the state ought to be concerned with.

I outlined a few influential approaches to distributive justice from different theories within the liberal tradition. These were Rawlsian liberalism (in particular, the Difference Principle), luck egalitarianism, the capability approach, and communitarianism. Absent a good understanding of how,

exactly, health incentives are likely to impact upon distributive patterns, I sought only to show that they do not look immediately unacceptable on this range of accounts. Once again, the variety of potential incentive schemes means that there will always be a broad range of effects that could result, and so, in some cases, health incentives could have objectionable allocative tendencies.

Features of incentive schemes which will make them more appealing to a variety of distributive principles will include prioritising resources and benefits to those worse off so as to reduce inequalities between different socio-economic groups; supporting people to act in line with their (considered, rational) preferences without overwhelming the agents' wills to act freely, such that all strata of society have the necessary opportunities to direct their lives as they wish; encouraging prosocial behaviour and solidarity.

Not all incentive schemes will be efficient, fit within the constraints of a moderately perfectionist state, and have desirable distributive effects. However, it is plausible that some amongst the multifarious types of incentive scheme possible, will fit such criteria, and both promote those values which we think it right for the state to promote without overstepping important boundaries of acceptable state intervention. In the following chapters I will consider some further objections to the use of health incentives which might yet deliver some restrictions on the kinds of schemes that will be permissible, or even provide reasons for thinking that incentives on the whole will be undesirable.

FOUR

THE INFLUENCE OF INCENTIVES: COERCION AND ITS COMPANIONS

INTRODUCTION

A number of the arguments arising in media discussions of health incentives concerned the potential for health incentives to pressure the recipient into changing her behaviour.[9] The worry is that this form of pressure harms the recipient, perhaps by coercing her into behaving in ways she doesn't want to behave. Any interventions intended to influence decision-making can raise issues of this sort because it is often assumed that free, autonomous, independent decision-making is preferable to decision-making that is in some way determined, influenced, manipulated, or controlled by another. Actions that are perceived as controlling are often met with hostility, and there is a particular tradition within medical ethics procedures which involve influencing agent behaviour.

Related to issues of influencing behaviour is that of paternalism. By definition, this identifies an action that is intended as benevolent by the perpetrator, but which goes against the subject's wishes, or is performed in ignorance of those wishes (Feinberg [1986]; Dworkin [1988]). Paternalistic action may be justified when there is good reason to believe that an agent is not capable of behaving rationally or is at risk of causing herself significant harm: for instance, individuals who are suffering from severe psychosis, are

---

9 For instance, Bribe, Responsibility, Nanny State, Health Harm, Doctor-Patient Relationship, Political, Stigma, Wrong Message

particularly young, or are unaware of all the relevant facts. In such cases the paternalist may have a privileged capacity to identify what is best for the agent. Where no factors that undermine the agent's capacity to choose for herself are present, then paternalistic action denies her the opportunity to do so.

Coercion, force, compulsion, or extreme pressure is generally thought undesirable because it denies agents the opportunity to act freely. Paternalism is generally objected to on the grounds that it undermines agent autonomy (related to freedom, but not equivalent). My concern in this chapter is to consider if incentives will ever, never, or sometimes result in objectionable liberty and / or autonomy undermining outcomes. In particular, I am interested here in the *effects* of incentives on the agent, rather than the *intentions* of the incentive provider. Paternalistic and non-paternalistic interventions may be enacted through a number of means of influence, and it is these (rather than the presence or absence of paternalism) that I wish to discuss.

It is plausible (indeed probable) that health incentives can be described as paternalistic: generally, they aim to promote the best interests ('healthy choices') of the agent, and need not take into account what the agent's preferences about her behaviour are. However, whilst interesting, I do not think the question of paternalism should be the primary concern. As discussed elsewhere (see Wilson [2011]) many health interventions (particularly in the domain of public health) will be paternalistic, and this should not be assumed to be problematic. Rather, the concern ought to focus on the opportunity for agents to act freely and maintain a degree of autonomy (where feasible). Thus, the means of paternalism, rather than the paternalism itself, may be considered the primary issue.

In the previous chapter I discussed liberty, and in particular, whether governments must remain neutral with regards to conceptions of the good life so as not to impose alien values on citizens. This chapter will be concerned with liberty as well, but here I am more interested in the psychological freedom of the agent, and whether she is put under unacceptable pressure to act in particular ways (perhaps against her will). Sufficient psychological and political liberty are necessary conditions for autonomous agents, as well as being important in their own right.

This chapter is split into three parts: Part One considers the concept of coercion and outlines some of the key ideas developed in philosophical theories of coercion. Part Two seeks to apply this discussion to the case of incentives, assessing the extent to which incentives might be thought to have a coercive influence over recipients. Of the methods of behavioural change discussed in this chapter - persuasion, coercion, manipulation, nudging - the most threatening to liberty appears to be coercion, as it is often taken to utterly subjugate the agent's will to that of the coercer. Hence, I lend more discussion to this concept. Part Three considers these other means of influencing behaviour, and how they play a role in the way incentives encourage healthy behaviour.

My aim in this chapter is to be able to make some comments about how health incentives impact upon liberty and autonomy. I seek to do this by building up relatively abstract accounts of different ways of influencing behaviour, with a particular focus on how these influences are experienced by the agent (how they affect her psychological state). I then compare these forms of behaviour to what looks to be happening when incentives are used to encourage healthy behaviour, in order to establish whether incentives can be thought to employ these means of influence. The final task is to consider the relevance of this for liberty and autonomy, and the overall ethical acceptability of health incentives.

First, by way of framing the discussion, I will make some introductory remarks about autonomy, and the importance of having a realistic conception of the human agent in identifying undesirable deviations from the ideal. This will serve to inform the later discussion about the extent to which incentives should be considered to negatively (or positively) affect liberty and autonomy. I describe some of the misconceptions of autonomy, often involving excessively demanding conditions to be met for autonomy to be achieved. I also run through some areas of discussion which have relevance for autonomy and (health-related) behaviour, including situationism, nudging, and research into the social determinants of health.

'Independent' Behaviour and Autonomy

It is generally assumed that 'free,' 'autonomous' action is the most worthwhile, and action lacking in these qualities will be undesirable. Yet there is a tendency

to talk about such attributes as if all human agents are capable of attaining 'perfect' freedom or autonomy, and that this should be the ultimate (downstream) aim of all self- or other-regarding action (or at least, one of its primary aims).

Even where ambition does not stretch quite so far, idealised models of autonomous agents may be far removed from reality. This can result in the bar being set too high, and unrealistic assumptions being made about the extent to which agents can (and should) make robust, self-conscious, character-driven, internally-motivated, independent, rational, reasons-sensitive, future-aware decisions, and cleanly execute these decisions in all of their actions.

This is clearly pie in the sky: none of us lives in isolation, and all of our actions are influenced at some level by other agents and the environment. Further, we are constantly subject to failings of cognitive function. I need not surrender my perception of myself as an autonomous agent simply because, once again, I poured coffee on my cereal and filled my mug with milk, or made another 'accidental' purchase from Amazon. There is a leniency in everyday life for behaviour to be less than fully consistent with rational, considered preferences, and yet for the agent to still be considered autonomous. We must adopt a reasonable level of 'autonomy' that is both achievable and desirable.

When it comes to incentives, we must consider how far the external influence disrupts the agent's capacity to make decisions freely, and her status as an autonomous agent. It is not at all clear that simply being subject to more external influence will be undesirable. As social creatures, being aware of and responsive to our environment is key to successful survival, and the alternative, of being isolated and unresponsive to others in one's behaviour, will not be a worthwhile enterprise.

The philosophical literature on the concept of autonomy has developed to reflect the debates about just what is valuable about different forms of political, moral and psychological forms of autonomy. Hence, autonomy discussed in different domains may stipulate different conditions. A widely influential discussion of moral autonomy, however, is provided by Kant (1785/1993). The Kantian tradition places autonomy central to practical reason, and uses it to

derive the Categorical Imperative and the basis of morality.[10] This fundamental significance of autonomy is maintained in the dominant political philosophy of our time, Liberalism.[11]

I cannot do justice to the rich and involved literature on autonomy that already exists, even within the disciplinary confines of bioethics.[12] Historically, autonomy has played a substantial role in moral and political thought, and continues to hold importance for contemporary philosophical theory, especially within bioethical debates (Beauchamp and Childress [2001]; Gillon [2003]). Despite more recent scepticism as to the centrality liberty and autonomy ought to play in moral thinking in bioethics (Dawson [2010]; Wilson and Dawson [2010]) the concept still clearly holds intuitive appeal as a valuable feature of human agency.

For the purposes of this discussion, I take it that some conception of autonomy can uncontroversially be assumed a desirable 'good,' though the specifics of this claim I leave largely underdetermined. As this chapter develops, the kind of features that I take to contribute to valuable autonomy, and factors that detract from it, should become clear.

Of relevance to the discussion of autonomy in the context of influencing health behaviour is the role the environment plays in determining preferences and decision-making. A better understanding of human psychology may alter how we conceive of (the importance of) autonomy. Some, including 'situationist' social psychologists, argue that it is the situations a person finds herself in, rather than her robust character traits, that predict how she will behave (Doris [2002]; Harman [2000]). Thus, situationists claim, "across a range of

---

[10] Kant offers the Categorical Imperative as the supreme principle of practical reason, a commonly used formulation being "Act only according to that maxim whereby you can at the same time will that it should become a universal law" (Kant [1785/1993: 30]). Once again, much sophisticated analysis of Kantian autonomy, including the Categorical Imperative, has been undertaken, but I do not intend to engage with this discussion here (see O'Neill [1975]; O'Neill [1989]; Guyer [2000]; Korsgaard [1996])

[11] In particular, it is central to much of Rawls's work, including 1971's *A Theory of Justice* and the proposal for establishing a society built upon just institutions.

[12] A handful of relatively recent influential discussions directly addressing or relating to autonomy include Mill (1859/2006); Raz (1986); Berlin (1969); Dworkin (1988); Dworkin (2000); Feinberg (1986); O'Neill (2002).

situations, the person's behavior tends to converge on the behavioral norm for those situations." (Kamtekar [2004]) Various experiments have been conducted by those within moral psychology and experimental ethics, seeming to show interesting examples of inconsistent, contradictory and often surprising moral reasoning in experimental subjects (Appiah [2008]; Strack and Deutsch [2004]; Haidt et al. [1993]).

Such a theory proposes to challenge the significance of the Aristotelian conception of the "firm and unchangeable" virtuous character, expressed through right actions (Aristotle [1984: 1105b1]). A debate has developed between the situationists on the one hand who argue that the importance of environmental influences on behaviour mean we should discard notions of virtue and character as attaching to particular individuals, and the virtue theorists on the other, who argue that the empirical evidence poses no such challenge to virtue theory (for proponents of situationism, see Doris [1998, 2002]; and Harman [1999, 2000]; for responses see Kupperman [2001]; and Kamtekar [2004].

Another example where it seems that behaviour is surprisingly responsive to unnoticed external influences is described by the 'libertarian paternalists' Thaler and Sunstein (2008) whose popular and oft debated book *Nudge: Improving Decisions about Health, Wealth, and Happiness* describes how subtle changes to the way a decision situation is set up (the 'choice architecture') can have a significant impact on the way people behave. They provide what has become the classic example of nudging: arranging food items in a particular way in a cafeteria in order to influence what people eat. Supposedly, by positioning healthy items in more salient positions, one can make it more likely that diners will select those items over less healthy alternatives (Thaler and Sunstein 2008: 1).

Again, whereas in the situationism/virtue ethics debate the claim is that it is environmental influences rather than moral character that best predict behaviour, here the libertarian paternalists claim that it is choice architecture rather than personal preferences which predict dietary decisions.

A final domain of research worth mentioning relates to lifestyle choices and socio-economic status. In the next chapter, I discuss further some of the empirical literature which looks at the relationship between level of deprivation and health outcomes. Briefly, there is a correlation between social and

environmental factors and health which may be at least partly explained by non-personal factors, calling into question the idea that lifestyle choices are the result of agent-specific, independent decisions (Marmot et al. [2010]).

Empirical evidence of the sort mentioned above is open to debate as to the reliability and interpretation of results, and, importantly, how they should shape our theories about human agency. However, the assumption that people ordinarily act according to stable preferences seems weakened by such evidence, and psychological research relating directly to the mechanisms of health behaviour suggests that much behaviour is not consciously willed.[13]

As mentioned, it will be important to have a realistic understanding of what kind of 'autonomy' human agents are ordinarily capable of attaining. When considering the danger posed by health interventions, therefore, we should focus on the general capacity for self-determination: to direct one's life in the manner in which one chooses. In order for agents to have this control over their own lives, they must be capable of informed deliberation between a reasonable choice set, and they must have the power to implement those choices and influence how their life goes.

This is a fairly minimal notion of autonomy, but I adopt this in part to indicate a version of autonomy that could be considered valuable to agents on the majority of accounts.[14] This rather modest autonomy will, however, be frustrated where an agent's actions do not reflect her own plans and projects, but rather those of another, or of some overwhelming compulsion within herself. Coercion, manipulation, and 'nudges' by others (however benevolent their intentions) which circumvent the agent's deliberative capacities as a means to behaviour change may be deleterious to agent autonomy. The aim of this chapter will be to explore if and when health incentives might employ such means of behaviour change, and whether in doing so agent autonomy is undermined in a troubling way.

---

[13] I discuss this further in the section on 'Reflective and Impulsive Control Over Behaviour' towards the end of this chapter.

[14] Inevitably, some will be dissatisfied and would insist a more robust notion of autonomy is an essential criterion to any permissible health intervention.

THE CONCEPT OF 'COERCION': TWO APPROACHES

There is something paradoxical about the concept of 'coercion.' Whilst it seems to offend against a number of uncontroversial 'goods' (choice, freedom, voluntary action, autonomy, and so forth) it is also widely accepted as being essential to the government of society. Most illustrative is the need for a law enforcement system with the power to coerce (would-be) criminals with (threats of) fines, incarceration, and other punishments. Some philosophers, such as Robert Paul Wolff (1970/1998) bite the bullet here, insisting that state coercion, where this is at the expense of individual autonomy, is never justified. Most, however, uphold that to coerce an individual will generally be wrongful, but is in some instances justified.

There are two key requirements, then, for an account of coercion. First, it must specify under what circumstances we should identify an influence as being coercive. Second, it must instruct us as to when a coercive influence is morally permissible, and when it is not. Such an account will provide a descriptive and normative framework for talking about coercion in a useful way – a way that allows us to evaluate potentially coercive actions according to this framework.

A useful distinction can be drawn between two approaches to developing an account of coercion which I will call the *situation-centric* and *agent-centric* approaches. The former is more concerned with looking for coercion in facts (or assumptions) about the situation in general, whilst the latter looks for coercion in facts (or assumptions) about the agent. One might phrase this (roughly) in terms of causes and effects: the *situation-centric* approach focuses on the *causes* of coercion, and what items need to be present in a situation in order for it to qualify as coercive. Contrastingly, the *agent-centric* approach focuses more on the *effects* of a proposal, and what consequences would indicate the occurrence of coercion.[15]

---

[15] Imagine how one might identify what is a cake and what is not. The situation-centric approach will look at the ingredients list, and how it was cooked. If it contains eggs and flour and butter and is baked in an oven, it is probably a cake; if it contains none of these ingredients then it could not be a cake. The agent-centric approach, in contrast, will look at the finished product and try to decide if the final item meets some 'cake' standard.

The ***situation-centric*** approach focuses on the actions of the proposer and the situation the proposee subsequently finds herself in. This approach tends to be tied to discussions of the distinction between threats and offers, and the significance of baselines by which to assess proposals. Such items provide standards to judge the coerciveness of a proposal against.

The ***agent-centric*** approach focuses on the psychology of the proposee, and the impact the making of the proposal has on her. Of less interest to this account is the position of the proposal relative to some baseline (and thus, whether it is a threat or an offer). All influences on the proposee's psychology are evaluated in the same way.

Of course, most sensible accounts seek to capture both these aspects of coercion, acknowledging that both agential and non-agential factors contribute to whether or not a proposal is coercive. However, the *two approaches* distinction I have described here is rarely explicitly acknowledged. Rough and ready though the situation-centric / agent-centric distinction may be, it is worth making and bearing in mind during the discussion of different accounts of coercion. Many accounts will emphasise one or other of the situational and agential factors in assessing coercion, and there also tends to be more within-group consistency along this distinction.

THREATS, OFFERS, AND BASELINES

Much of the literature on coercion dwells on the distinction between threats and offers, and whether offers, as well as threats, may be coercive. Threats and offers can broadly be distinguished thus:

A proposal is a ***threat*** if the recipient will be made worse off if they reject the proposal.

A proposal is an ***offer*** if the recipient will be made no worse off by rejecting the proposal.

The threat / offer distinction provides a platform for thinking about proposals in terms of how they affect the recipient relative to some baseline. The baseline provides the reference point for the 'better off' and 'worse off' used to distinguish threats from offers. This approach tends to be more situation-

centric, because it looks to features of the pre- and post-proposal situation to identify whether the proposal should be considered a threat or an offer.

Most theories that place an emphasis on threats, offers, and baselines assume that these distinctions are significant to identifying coercion. That is, whether a proposal is coercive or not will depend in part on how it affects the recipient *relative to some baseline*. Agent-centric approaches place more emphasis on the psychology of the recipient and tend to be less interested in the threat / offer distinction, instead emphasising the effect a proposal has on the recipient rather than the form of the proposal itself. On such accounts, all influences on an agent's psychology (threats and offers alike) are treated in the same way. If there is a qualitative difference between how threats and offers are experienced by recipients (with one being capable of coercion and the other not) then the threat / offer distinction would hold more significance on the agent-centric account.

Baselines are used to judge whether an individual has been made 'better' or 'worse' off by a proposal, and thus whether that proposal is a threat or an offer. Often, baselines appear obvious and need not be explicitly identified in order to distinguish threats from offers. In the following two cases, Case I seems obviously to be a threat, and Case II obviously an offer.

Case I: Dick holds a loaded pistol up to Ambrose's head, demanding that Ambrose hand over his purse, or suffer being shot in the head.

Case II: Boris asks David if he would like to buy his rhinestone-encrusted bicycle for £500.

In Case I, Dick's proposal (to take Ambrose's money in exchange for not shooting him) is a threat because, once the proposal has been made, Ambrose is worse off than before the proposal was made. In Case II, before the proposal is made, David has all of his money but no rhinestone-encrusted bicycle. After Boris has made his proposal, David may choose to either part with £500 in exchange for a sparkly bicycle, or keep his money and continue to drive to work. David is no worse off for receiving Boris's proposal, as he can safely reject it without suffering any loss from his baseline (pre-proposal) condition.

Roughly, in these cases we compare the situation before the proposal was made (the pre-proposal situation) to the situation after the proposal was made (the post-proposal situation). The pre-proposal situation acts as a baseline by which to judge if the proposal was coercive.

Sometimes, however, peoples' baselines place them in already vulnerable positions. Nozick (1997: 26-27) provides an example of a drowning man who is offered rescue in exchange for $10,000. Clearly the man is better off for having received the offer of rescue, yet the actions of the greedy rescuer still seem coercive. One response of the baseline-theorist is to change the stipulation of the baseline: instead of being the 'normally expected' course of events, the baseline is now identified as the 'morally expected' course of events. The baseline is now set as what the proposer *ought* to do (in this case, the proposer clearly ought to save the drowning man for no money) (Nozick [1997: 27]). This creates a moralised (baseline) account of coercion.

Wertheimer (1987) adopts a broadly Nozickian account of coercion, but whilst Nozick does not insist that *only* threats (and never offers) can coerce, Wertheimer's view differs:

> I have argued, in effect, that the coerciveness of proposals is all in the baseline. And relative to that baseline, only threats are coercive.
>
> Wertheimer (1987: 222)

There are others who agree with Wertheimer and argue that only threats, and never offers, can be coercive. Gorr (1986) and Berman (2002) both argue that proposals should be seen as 'biconditionals' – that is, they have two parts to them, the offer and the 'anti-offer'. For instance, I might propose to give you £20 if you give me your watch. The first conditional here is that which is explicitly made: I give you £20, you give me your watch. The alternative conditional is implicit: I do not give you £20, you do not give me your watch. This can be read as "unless you give me your watch, I will not give you £20." According to Gorr and Berman, when proposals are coercive, it is always by virtue of one of their constitutive conditionals being a threat.

Zimmerman (1981) is one of the most prominent advocates of the view that offers, as well as threats, can coerce. According to Zimmerman, "By and large, threats involve coercion and offers do not: mainly because people do not

like to be threatened whereas they do like to receive offers." (1981: 125) In explaining his position, Zimmerman refers to another example of Nozick's, the Slave Case, where a regularly beaten slave is told he will not be beaten on a particular occasion if he performs some specified (unpleasant) task:

> [A]n offer is coercive only if [the recipient] would prefer to move from the normally expected pre-proposal situation to the proposal situation, *but he would strongly prefer even more to move from the actual proposal situation to some alternative pre-proposal situation*. The slave, for example, would strongly prefer not being a slave to having a choice between being beaten and being spared a beating for performing a disagreeable task.
>
> Zimmerman (1981: 132, emphasis in original)

The slave would much prefer to be in an alternative situation (where he is neither a slave nor beaten). In this situation, the 'slave' would be made worse off by the 'slave owner's' proposal ("perform action *A* or I will beat you"), thus it would be coercive. It is through keeping the slave in his enslaved situation, and denying him access to a preferred situation, that the slave owner's 'coercive offer' derives its coerciveness.

Zimmerman stipulates conditions to restrict the kinds of alternative pre-proposal situations that might be referenced. First, the alternative situation must be *feasible* for the 'coercee,' and second, the 'coercer' must have *actively prevented* her from achieving it (Zimmerman [1981: 144-145]). In the Slave Case, it is feasible that the slave not be a slave, and also the case that the slave owner actively prevents the slave from accessing this preferred situation, thus the slave owner's proposal is coercive. In the Drowning Man Case, however, although it is feasible for the man not to be drowning the greedy rescuer is not actively preventing him from accessing this situation. On Zimmerman's account, then, the greedy rescuer does not coerce.[16]

In summary, the requirements for coercion to occur according to the different baseline theories are:

[16] Instead, Zimmerman identifies this as a case of exploitation.

- One is better off in the preproposal situation than the postproposal situation (it would be better not to receive the proposal). The proposal is a *coercive threat* (according to both moralised and non-moralised accounts), or;
- One is better off in the postproposal situation than the preproposal situation (it is better to receive the proposal). However, there is an alternative (morally acceptable) preproposal situation that is better than one's actual (morally unacceptable) preproposal situation, and also better than the postproposal situation. The proposal is a *coercive threat* (according to the moralised account), or;
- One is better off in the postproposal situation than the preproposal situation (it is better to receive the proposal). However, there is an alternative preproposal situation that is both feasible and made unavailable by the proposer. This alternative preproposal situation is better than one's actual preproposal situation, and also better than the postproposal situation. The proposal is a *coercive offer* (according to the non-moralised account).

NON-BASELINE APPROACHES

Another group of philosophers take a different approach to identifying coercion. Frankfurt (1988/2007) and Feinberg (1986) both look to the affect a proposal has on the recipient's will and their ability to choose between courses of action to distinguish coercive from non-coercive proposals. Neither of these accounts require a baseline to be identified.

Frankfurt takes coercion as a very robust notion, whereby one who is coerced is absolutely compelled to act in the way she does - her will is utterly overwhelmed by the proposal:

> A coercive threat arouses in its victim a desire – i.e., to avoid the penalty – so powerful that it will move him to perform the required action regardless of whether he wants to perform it or considers that it would be reasonable for him to do so. Now an offer may also arouse in the person who receives it a desire – i.e., to acquire the benefit – which is similarly irresistible. This suggests that a person may be coerced by an offer as well as by a threat.
>
> Frankfurt (1988/2007: 41)

Frankfurt relates coercion to moral responsibility: one is free from moral responsibility when one is coerced. Hence, Frankfurt stipulates such demanding conditions for agents to be considered as having acted under coercion.

Feinberg (1986) similarly identifies coercion as placing an extraordinary degree of pressure on the subject's will, such that her choice is not truly 'free' or 'voluntary.' Feinberg draws a line between compulsion and coercion, however. In the former case, all actions apart from that which the compeller wishes to be performed are closed off, so that it is not possible for the compelled agent to act in any way other than how she does. In the case of coercion, all actions other than that action the coercer wishes to be performed become associated with an extremely high cost. The coerced agent *could* still choose to act other than the coercer wishes her to act, but it would be to her vast detriment (she may be shot in the leg if she chooses not to hand over her money, or suffer some other terrible fate) (Feinberg [1986: 190-193]). Depending on the degree of coercive pressure placed on the recipient by a proposal, the voluntariness of her subsequent actions may be utterly compromised, or may be less affected.

As with Frankfurt, Feinberg accepts that offers may coerce as well as threats (Feinberg 1986: 229). Part of Feinberg's criteria for identifying *coercion proper* (where coercion is judged to have occurred) distinct from *coercive pressure* (where 'coercive' proposals have been made but the recipient is not coerced) is that the recipient accede to the coercive proposal. This gives Feinberg a way of distinguishing coercive from non-coercive threats: the former are successful in getting the recipient to act as the coercer wishes, whilst the latter are not. People will only succumb to a threat (and thus act in a manner that makes them worse off) when the consequences of the alternative action are even more undesirable.

This raises an issue when trying to distinguish coercive from non-coercive offers. People will often act in accordance with offers, as they make the recipient better off, yet we should not wish to say that on these occasions the recipient is coerced. To resolve this issue, Feinberg identifies two criteria that are significant to identifying coercion: the *polarity* and *proximity* conditions (Feinberg [1986: 235]). In the *polarity* condition, Feinberg stipulates that only when one part of a proposal (half of the biconditional) promises to make the

recipient worse off can that proposal be coercive.[17] The *proximity* condition refers to the degree of coerciveness brought by a proposal, and thus how likely it is that the recipient will be coerced.[18]

These approaches to coercion are agent-centric, as the focus is not on the proposal itself, but on how it influences the recipient. Where the proposal sufficiently impacts upon the will or decision-making ability of the recipient so as to render her actions unfree, involuntary, compelled, or 'not her own' in some way, the agent is considered to have been coerced. The non-baseline accounts need not strictly distinguish between a threat and an offer because they are concerned with the way a proposal affects the recipient's psychology and her decision-making ability, rather than with the nature of the proposal itself.

Summary Remarks

I have sought to outline some different forms an account of coercion could take. In doing so, I have indicated how differences in these theories relate to different evaluations of potentially coercive proposals. The main distinction I have drawn has been between situation-centric and agent-centric accounts, which typically focus on baselines in the first case, and psychological pressure and voluntariness in the second.

As I have described it, the agent-centric, non-moralised account will probably be more permissive when it comes to identifying coercion (although all accounts will vary depending on what other necessary conditions they include). Unless it is made a precondition of coercion that the coercer acts wrongfully, it is possible that trivial threats and proposals not intended to put significant pressure on the recipient will qualify as coercive because they have an unusually large impact on the recipient. For instance, the threat "I will stuck my tongue out at you unless you give me some of your popcorn" could be intended as a

---

[17] It cannot be the case that choosing between two options that both make me much better off, or choosing between my current (comfortable) status quo and a better situation can be at all coercive.

[18] Two minimally bad outcomes will make for a 'hard' decision because there is little to choose between them. Contrastingly, choosing between a bad outcome and a completely disastrous one may be a more harrowing decision to make, but is actually quite 'easy' in this sense. The latter decision will prove more coercive.

joke. However, if the recipient of such a proposal happened to have a mortal fear of tongues then such a proposal could coerce her.

If we care about the moral status of the actions of the proposer and, in particular, the recipient, we should adopt an agent-centric account, where the key to identifying coercion is the impact a proposal has on the recipient's will. Abstracted from the effect on the recipient, a coercive (or non-coercive) proposal carries no moral weight. Focus on the legitimacy of the coercer's actions might be judged better by measures of her intentions (whether to coerce or not), and this might be accessible by considering the reasonableness of the expectation that the recipient will be coerced by a proposal (rather than whether the recipient *actually was* coerced). I am, however, more concerned with the recipient than the proposer, and the extent to which the recipient's liberty and autonomy is threatened by a particular proposal (in this case an incentive). Hence, an agent-centric account seems to identify the right features here.

That said, baseline (and situation-centric) criteria probably provide a handy way of judging when an agent is likely to be coerced, but seem to miss something of the importance of coercive influence (that the agent is not able to act freely and voluntarily). It is a matter of psychological fact rather than philosophical speculation, but it seems much more likely that the threat of being made worse off will pressure an agent and limit her liberty than the offer of being made much better off. Where baselines present a helpful way of assessing the influences on an agent, they may be a good way of structuring the consideration of what kind of pressure her will is placed under.

In order to avoid (the charge of) coercing someone, the proposer must make herself aware of the recipient's psychological state, as far as is feasible. This would require one to avoid, where possible, making proposals likely to overwhelm an agent's will and prevent her from acting voluntarily. There may also be vulnerabilities to particular threats and offers that affect people quite generally, not just those with limited capacity to resist coercive proposals. The proposer must therefore make it her business to be aware of just how much pressure a proposal is likely to place on the recipient's will.

In Part Two, I will consider how this account of coercion relates to the provision of health incentives. It will be important to bear in mind that incentives may be targeted at particular groups of people, and such groups may be likely to share particular psychological features that make them more

susceptible to incentives and thus more vulnerable to coercion from such interventions. I say this to pre-empt the likely (and completely appropriate) complaint that it is rather too large a burden on the proposer to make herself aware of all of the psychological foibles of the recipient. Indeed, one could hardly consider someone to have acted wrongfully by making a well-intentioned offer which happens to coerce. As mentioned, we may adopt a standard of reasonableness here which will firstly indicate what kinds of proposals could potentially coerce, and further, to what extent the proposer must inform herself of the specific vulnerabilities of her recipients.

Part Two: Coercion and Health Incentives

Introduction

I now want to assess health incentives in relation to this discussion of coercion. I will follow roughly the structure of Part One. First, I consider the significance of the 'two approaches' to theories of coercion proposed in the previous Part. I then discuss how incentives fit into the earlier discussion of baseline accounts of coercion, providing some examples of how different incentive scheme designs might look on these baseline accounts. Next I go on to look at non-baseline accounts, those tending to relate coercion to degree of choice and voluntariness of action: to what extent the recipient feels compelled to act in a certain way by a proposal.

I conclude with a brief discussion of how an agent-centric account of coercion relates to those factors significant to making normative judgements about the use of incentives. I gesture at how notions of choice, voluntariness, responsibility and autonomy will make a coercive act undesirable, and how, too, coercion might be justified where it benefits similar phenomena, or improves welfare or suchlike. My main point here is to emphasise that the relationship between the coerciveness of an act and the moral permissibility of that act is non-linear. The extent to which coercion is experienced, the manner of coercion, the benefits produced through coercion, and numerous other factors will contribute to the overall assessment of the rightness or wrongness of a potentially coercive act.

Two Approaches

In Part One, I suggested that a useful distinction could be drawn between theories of coercion that place an emphasis on the *situation* and those that place an emphasis on the *agent*. These different approaches to identifying coercion could have different relevancies for the use of incentives in altering health behaviour. To recap:

The ***situation-centric*** approach focuses on the actions of the proposer and the situation the proposee subsequently finds herself in. This approach tends to be tied to discussions of the distinction between threats and offers, and the significance of baselines by which to assess proposals. Such items provide standards to judge the coerciveness of a proposal against.

The ***agent-centric*** approach focuses on the psychology of the proposee, and the impact the making of the proposal has on her. Of less interest to this account is the position of the proposal relative to some baseline (and thus,

On the situation-centric account, the pertinent question when considering the coerciveness of incentives is whether or not there are features of incentive schemes that could make the proposals involved within them coercive. Contrastingly, if we take an agent-centric approach, we must consider what features of human psychology might mean that incentives could have a coercive impact on the recipient.

My preference is for a more agent-centric account, as these foreground the importance of the agent's experience, and whether or not she *feels* coerced (and is able to act voluntarily). However, both situational and agential factors will be helpful for predicting and judging coercion. Another division within theories of coercion was between baseline and non-baseline accounts. Again, I think both have something to offer to the analysis of health incentives, and provide different ways of framing the issue: does this proposal make the recipient worse off (baseline); does this proposal place excessive pressure on the recipient's will (non-baseline). Both seem pertinent to assessing the permissibility of behavioural interventions such as health incentives.

THREATS, OFFERS, BASELINES, AND LOSS AVERSION

My aim in this section is to relate the earlier discussion of the threat / offer distinction and baselines to the coercive potential of incentives. Baseline accounts of coercion identify some relevant starting point, against which the issuing of a proposal or the effect of that proposal is judged. Some accounts stipulate that only proposals that count as threats relative to the baseline can be coercive whilst others allow that offers may coerce as well. Different accounts also suggest different strategies for selecting the relevant baseline. In the earlier discussion, I suggested the following conditions under which a proposal can be considered coercive, according to the baseline accounts:

- One is better off in the preproposal situation than the postproposal situation (it would be better not to receive the proposal). The proposal is a *coercive threat* (according to both moralised and non-moralised accounts), or;
- One is better off in the postproposal situation than the preproposal situation (it is better to receive the proposal). However, there is an alternative (morally acceptable) preproposal situation that is better than one's actual (morally unacceptable) preproposal situation, and also better than the postproposal situation. The proposal is a *coercive threat* (according to the moralised account), or;
- One is better off in the postproposal situation than the preproposal situation (it is better to receive the proposal). However, there is an alternative preproposal situation that is both feasible and made unavailable by the proposer. This alternative preproposal situation is better than one's actual preproposal situation, and also better than the postproposal situation. The proposal is a *coercive offer* (according to the non-moralised account).

Although generally oriented towards situational factors, these baseline approaches can be more or less agent-centric as well. They focus on the effect the proposal has on the recipient, but judge this against some discernible baseline (which may be composed of elements of agent psychology and situational factors). Ordinarily, however, baseline approaches tend to emphasise situational factors in judging the valence of a proposal. In applying this aspect of

coercive theory to incentives, I will focus on the key consideration of whether the proposal makes the recipient worse off. Here, the emphasis is on the identification of the appropriate baseline.

Take the most straightforward use of incentives, where the proposal is considered both welcome and an offer. In this case, the recipient is no worse off for receiving the proposal of an incentive. This incentive could be something in the form of "as a one off, if you do not smoke any cigarettes today, I will give you three pounds." In this case, receiving the proposal makes you no worse off as you can turn it down without suffering any harm. This situation accords with none of the requirements listed above for a proposal to qualify as coercive: it is not a threat relative to the normal baseline; it is not a threat relative to the moral baseline; and let us assume there is no feasible, preferred pre-proposal situation that I have actively prevented you from accessing. In no way, according to the baseline accounts, should this proposal be considered coercive.

If this were the only way in which incentives could be used then, on the baseline account, there would be little need for further discussion. However, we can further complicate the picture. An important consideration for determining whether incentive schemes would be an appropriate means of changing behaviour and improving health is how efficacious they are. Some incentive schemes could involve straightforward mechanisms, such as the case where you do not smoke for a period of time and receive a cash payment in return. There might, however, be more effective ways of implementing the same incentive. Information relating to the psychology of decision-making and behaviour can help those designing incentive schemes to make them as effective as possible. In the next section, I will discuss how efforts to design effective incentive schemes may influence their coerciveness.

INCENTIVE SCHEME DESIGN: THE *CUSTODIAN* AND *HABITUATION* CASES

Economists and psychologists have long been aware of the phenomenon of *loss aversion*, developed as a component of 'Prospect Theory' by Kahneman and Tversky (1979). Prospect Theory seeks to explain why individuals do not always behave so as to maximise 'expected utility' (i.e. people do not choose the option with the greatest likely payoff). Initially, Prospect Theory was used to explain the decisions people made under conditions of 'risk,' but components of

Prospect Theory (such as 'loss aversion') are observed in riskless choice situations as well (see Tversky and Kahneman [1986: S258]; [1991]).

Loss aversion explains why people sometimes do not act in a manner most likely to maximise their gains, in preference for acting to avoid losses. This is because individuals are more sensitive to losses from the status quo than they are to gains (Tversky and Kahneman [1991: 1047]). Roughly, people weight some monetary loss about twice as heavily as the same monetary gain (Tversky and Kahneman [1991: 1053-1054]).[19]

Loss aversion can manifest itself in a number of empirically observed phenomena, such as the 'endowment effect' (Tversky and Kahneman [1991: 1041]; Thaler [1980]; and Kahneman, Knetsch, and Thaler [1991]). The endowment effect is identified as causing the asymmetry between the valuation of items made by buyers and those made by sellers. In such cases, "the loss of utility associated with giving up a valued good is greater than the utility gain associated with receiving it." (Tversky and Kahneman [1991: 1041])

Another important aspect of the theory relating to loss aversion is *reference dependence*, whereby "the carriers of value are gains and losses defined relative to a reference point." (Tversky and Kahneman [1991: 1039]) This references dependence is important in the case of incentive schemes and coercion, because it plays a fundamental role in determining the baseline.[20]

In designing incentive schemes, as with any healthcare intervention, efficiency will be advantageous. Exploiting phenomena such as loss aversion may provide a way of achieving greater cost-effectiveness. One way of utilising loss aversion in order to make incentive schemes more effective would be to make recipients experience non-gains as losses, and gains as non-losses. This can be achieved through *framing*. Depending on how a situation is framed – how it is described – logically equivalent occurrences can be interpreted differently, and affect a person's behaviour differently.

---

[19] I.e. to lose five pounds is about twice as bad as gaining five pounds is good. So, I might be expected to go to the same effort to avoid losing five pounds as I would in order to gain ten pounds.

[20] However, the reference point is not equivalent to the baseline. The reference point is that which losses and gains are judged relative to, whereas the baseline incorporates additional facts about the situation (for instance, Ambrose has all of his money *and* no threat to his life; David has £500 *and* no sparkly bicycle).

I propose that framing will, in part, depend upon a person's expectations: actual events will be evaluated in the light of what the agent expects to happen. In terms of incentives, for example, someone who does not expect to receive five pounds will experience the receipt of five pounds as a gain, and will experience the absence of the five pounds as the maintenance of the status quo. In contrast, someone expecting to receive five pounds will experience receiving the money as maintenance of the status quo, and not receiving the money as a loss.

There are at least two ways in which the lack of an incentive payment could be experienced as a loss from the status quo. One is due expressly to the design of the incentive scheme, and the other due to changes in the recipient's psychology as the incentive scheme progresses.

Take the *simple case*:

> Simple Case: I pledge to give up alcohol for ten weeks, in return for £100. I could be paid the £100 in one lump sum at the end of ten weeks, providing I have not drunk any alcohol, or I could be paid £10 each week I do not drink any alcohol, or £1.43 each day I don't drink alcohol.

In order to exploit loss aversion, however, I could be given the £100 upfront, and then have money deducted each time I fail to stick to my pledge. So, for each week that I drink alcohol, £10 of the money I was given is taken back. By having the money in my possession initially, I incorporate it into what I believe to be 'my' property (even though it will not become fully mine until I have successfully achieved my target of ten weeks alcohol free). Call this the *custodian scheme*. In this scheme, loss aversion and the endowment effect mean that I will likely suffer a greater disutility due to deductions from the £100 than the utility I would gain from receiving it. Through the operation of loss aversion the same sum of money should have greater leverage over my preferences in the *custodian* case than in the *simple case*.

The second way the lack of a gain may be experienced as a loss relies upon the recipient's psychology altering as the incentive scheme progresses. Take the case where I have committed to not drinking alcohol for ten weeks.

Now, imagine that I have been extremely successful, and have not touched a drop for the first seven weeks. In addition, I have become accustomed to my weekly payments of £10, and I have been using the money to go to the cinema each week. In the eighth week, I fully expect to receive my £10 payment and have even picked out which film I will go to see with the money. But then disaster strikes: in a moment of weakness I indulge in a forbidden gin and tonic. This means I do not receive my payment of £10 for that week, and can no longer go to see the latest *James Bond* movie at the cinema.

Over the course of the preceding seven weeks I have become habituated to the weekly £10, and the absence of payment in the eighth week is very much experienced as a loss. In this, the *habituation* case, the lack of an incentive payment is experienced as a loss because I have grown accustomed to receiving the money and have incorporated it into my expectations and plans for the future.[21]

There will be other ways in which incentive schemes can be designed to exploit the fact that people are more sensitive to losses than gains. Loss aversion could, of course, be triggered by making the subject experience *actual losses*, for example, by requiring them to pay penalties. Some schemes could combine paying out a reward if participants succeed with issuing a penalty if they fail, whilst. Some penalty-based, or combined penalty-and-reward schemes, could be designed to exaggerate the influence of loss aversion, so that relatively small losses have a more significant impact. My focus here will remain on schemes providing rewards rather than penalties.

Back to Baselines

In the *simple case* the incentive scheme does not appear coercive according to either the normally expected, morally expected, of preferred preproposal baseline accounts. Key is that the offer of the *simple case* incentive does not appear to make the recipient worse off. We might generally think that incentives cannot coerce because they are offers: they can be turned down without the

[21] I do not claim here that any of the example schemes I propose would necessarily work in the ways I suggest – this will be an empirical question about how individuals experience the effects of payments under particular circumstances. I am only trying to provide examples in order to illustrate what I believe is a plausible proposition, namely, that people can experience the failure to gain as a loss, and that this could operate in the domain of health incentives.

recipient being made worse off or experiencing any harm. However, in the *habituation* and *custodian* cases it seems that incentives (or rather, the denial of incentives) are capable of making the recipient worse off.

On the agent-centric account the gain or loss experienced by an individual will be important to determining coercion. On the situation-centric account the absolute gain or loss will be more relevant. So, where the recipient of an incentive is made no worse off (financially) compared to their position before the incentive was offered, if we take a situation-centric view, there will be little to suggest that incentives could be coercive. Contrastingly, if we take an agent-centric view, the fact the recipient *experiences* a loss, even if they do not suffer any *absolute* loss, could be sufficient to indicate that coercion is possible.

In the *habituation* and *custodian* cases, I described scenarios where the recipient feels worse off for not receiving an incentive. The issue that must be addressed is whether this experienced 'worse off' can be explained in terms of a legitimate baseline against which coercion can be judged. In order for this to be so, there has to be a shift in the recipient's baseline during the course of the incentive scheme. When incentives are first offered, they act as in the *simple case*, and cannot be coercive according to a baseline account. It may be possible, however, that during the course of the incentive scheme the agent's expectations alter, and a change in agent-determined preferences and expectations could cause a shift in the baseline.[22]

In the *custodian* scheme, the recipient is given custody of the total amount that she would receive if she met all of her pre-specified targets. The example I used was one where I commit to not drinking alcohol for ten weeks, and receive £10 for each week I succeed in not drinking alcohol. On the *custodian* scheme, I would be given £100 in the first instance, and lose nothing for every week I do not drink alcohol, but lose £10 for every week that I do drink alcohol. Once the money has been given to me, I might plausibly incorporate it into 'my'

[22] It is worth noting here that the baseline accounts discussed in the previous chapter do make reference to recipient preferences when identifying baselines and coercion. On Nozick's account, recipient preferences can determine whether a normal or a moral baseline should be adopted. On Zimmerman's account, recipient preferences are used to determine if there is a relevant preferred pre-proposal situation, and what this is. On agent-centric accounts, recipient preferences and expectations are likely to be central to identifying coercion.

possessions; it becomes a part of my 'status quo,' and thus alters my reference point.

Likewise, in the *habituation* case, I become accustomed to my weekly payments of £10 and make plans about how I will spend the money and so forth: the weekly payments are incorporated into what I expect to happen each week. Thus, my reference point alters to include the weekly £10 payments, and the lack of a payment is experienced as a deduction from this reference point.

In these *custodian* and *habituation* cases, the framing of the situation alters during the course of the incentive scheme (that is, any time after the offer of the first incentive is made, until the payment / non-payment of the final incentive). So, payments that were initially framed as gains become neither gains nor losses, and non-payments become framed as losses.

It thus seems possible that an incentive could be non-coercive when offered at the beginning of an incentive scheme, but coercive when offered later in the scheme. This requires that incentive schemes be considered as a series of distinct proposals, rather than as one initial, continuing proposal. This is probably only plausible in the case of schemes that pay out incentives at intervals, rather than schemes that pay out a single lump sum at the end. For instance, in the case of abstaining from alcohol, at the end of each week I am either paid or not paid £10. The proposal of "you will receive £10 if you do not drink any alcohol this week" is then (implicitly or explicitly) renewed for the next week.

The reference point is fundamental to determining the baseline by which to judge coercion, as it identifies the point relative to which 'better off' and 'worse off' judgements are made. However, the baseline should not be considered as simply *equivalent* to the reference point, and this means that there is a potentially important distinction between the *custodian* and *habituation* schemes.

In the *custodian* scheme, the initial baseline can be taken as that where I drink alcohol and neither receive nor lose any money. This should be taken as the pre-proposal situation. Relative to this, as established, the proposal of an incentive does not make the recipient worse off and will not be coercive. If I accept the proposal I am immediately given £100. In fact, I am given 'custody' of the money – I cannot spend it until I meet the criteria of the scheme and it

becomes 'fully' mine. It may be initially paid into an account I will later be given access to, or I may be given 'Monopoly money' that can later be exchanged, or through some similar means. The key is that I am 'given' the money such that I incorporate it into what I consider to be my possessions.

After I have received the money the proposal alters. The proposal now presented to me switches from "each week you do not drink you will receive £10" to "each week you drink you will lose £10." The proposal now moves to a loss frame (lose £10 for every week you drink). The change in frame triggers a change in my reference point: after I have accepted the initial proposal my reference point alters to include the £100 I have custody of.

In the *habituation* case the initial baseline is the same (I drink alcohol, I receive no money). The proposal of the incentive before the scheme has begun, relative to this (pre-proposal) baseline is not coercive. When I accept the proposal my reference point initially remains the same. However, after a few weeks of successfully not drinking, I grow accustomed to the incentive payments: my reference point shifts to incorporate the weekly £10 payments.

The shifts here are similar, but not the same: in the *custodian* case the reference point changes without any associated changes in behaviour; in the *habituation* case the reference point only changes after a number of weeks of behaviour change. Whilst the reference point can be taken as the utility status quo (the marker against which gains and losses of utility are judged), the baseline must be taken as a broader, more holistic measure. In this case, the reference point is the financial situation from which I experience gains or losses, but we must incorporate this reference point *and my behaviour* (in this case, whether or not I drink) into the baseline.

In the *custodian* case, because the shift in reference point is not associated with a shift in behaviour, the new baseline is such that subsequent proposals can count as threats. In the *habituation* case, the reference point alters only with the associated behaviour change and thus the renewal of the proposal in the seventh week is not a threat.[23]

[23] I think the *habituation* case is in line with common sense judgements. Take the case of an employer and an employee: if the employee receives payment for performing a particular agreed upon job then they cannot expect to stop performing that job and still receive money from their employer.

To summarise, in the *custodian* case proposals could potentially operate as coercive threats, as the recipient may be better off in the pre-proposal situation than in the postproposal situation after she has received the first payment. I have argued that the same is not true for the *habituation* case, however, as neither the initial, nor any subsequent incentive proposals should qualify as coercive threats.

NON-BASELINE APPROACHES

The previous section sought to explore whether incentives targeted towards changing health behaviours could be coercive according to a baseline account of coercion. I have suggested that some schemes specifically designed so as to make the recipient feel as if they are being made 'worse off' when they do not receive a payment may coerce.

This does not show that health incentives actually *are* in practice coercive, merely that it is possible for them to meet the initial criteria. In fact, losing £10 may simply not be a big enough threat to warrant coercion. The trouble with the baseline account is, even when we have successfully identified an appropriate baseline, and determined whether a proposal makes the recipient better or worse off relative to this baseline, we still cannot say whether the proposal is coercive or not. Assuming that coercion is taken to be something over and above simply acquiescing to a threat, there will be many proposals that make the recipient worse off ('threats') but do not coerce.

Baseline accounts seem able only to take us part of the way: identifying instances that are definitely *not* coercive. I have shown that not all cases of incentivising health behaviour change can be ruled out in such a way. In order to identify those cases where the recipient actually experiences coercion, a non-baseline account may be useful. Such accounts, as adopted by Frankfurt (1988/2007) and Feinberg (1986), look at how the proposal affects the recipient's will and her decision-making ability.

Both Frankfurt and Feinberg emphasise the importance of the unique susceptibility of the recipient to a particular proposal, and make the psychology and preferences of the recipient central to judgements about coercion. Feinberg in particular argues that:

> For moral purposes we should always use the subjective "willingness" standard, and judge voluntariness by reference to the actual motivation of the person whose choices are at issue. That is the approach that is most respectful of autonomy and individuality, judging each individual as he is, not as the actuarial tables say he probably is.
>
> Feinberg (1986: 258)

Thus, the actual experience of the individual is fundamental to judging coercion, although Feinberg acknowledges that in practice this is extremely difficult to judge, and often we may have to use our objective knowledge of the situation as a proxy for this information. Frankfurt's stronger claim is that coercion will result in the complete absence of voluntary action (not simply a reduction): Coercive proposals must be "irresistible" (Frankfurt 1988/2007: 41):

> Coercing someone into performing a certain action cannot be, if it is to imply his freedom from moral responsibility, merely a matter of getting him to perform the action by means of a threat. A person who is coerced is *compelled* to do what he does. He has *no choice* but to do it.
>
> Frankfurt (1988/2007: 36)

On these non-baseline accounts it is possible for offers as well as threats to coerce. It still seems that schemes like the *custodian case* are likely to put more pressure on the subject than, for example, the *simple case*. Yet even the *custodian* scheme seems very unlikely to provide sufficient pressure to meet Frankfurt's criteria for coercion.

For Feinberg, the coerced act must be chosen, and thus involve a minimum degree of voluntariness. It must also meet the *polarity* and *proximity* requirements. For *polarity* the incentive may only coerce where its proposal makes the recipient 'worse off' in some way (for instance, the quasi-threatening *custodian scheme*) or where the recipient is in such a dreadful situation that she cannot countenance any thought of remaining in that situation (and the proposal represents her only chance of escape). In addition, the *proximity* condition requires sufficient pressure to be placed on the individual. It still seems unlikely that incentive schemes will meet both of these conditions. However, as the size of incentives gets larger, and the recipient more vulnerable and more susceptible to external pressure, it becomes feasible that incentives

could coerce. The relevant cases here are those where very vulnerable people are being offered (perhaps very large) incentives to change their behaviour.

I use the term 'vulnerable' with a deliberate vagueness. It could, for instance, identify very poor people, for whom even a small cash incentive could represent a significant sum; people addicted to drugs who may be desperate for money, and who may also have a reduced capacity for rational decision-making; or people suffering from mental illness, who may struggle to make consistent, considered choices. How external circumstances affect the vulnerability of an agent will depend on her individual character, although psychological research is likely to be able to help identify many situational indicators of susceptibility to coercive proposals.

Other Effects of Incentives

It is worth briefly mentioning how other aspects of health incentives might introduce psychological pressure that could contribute to coercion. There are at least two other features of incentive schemes that could vary such that incentives sometimes wield a large (potentially coercive) degree of influence over the will and decisional capacity of the recipient.

First, the wider social significance of the incentive: why the incentive is being offered. By this I am considering the implications of, say, offering cash incentives to pregnant women who smoke. The message behind such proposals could be interpreted along the lines of "you must accept this incentive and quit smoking for the good of your unborn child." Such statements employ 'emotional blackmail' by intensifying feelings of moral duty and guilt. Effects such as these could make the recipient 'worse off' if they prefer not to be reminded about the harm they do to others through their smoking. The stimulation of such feelings could make the recipient's current situation untenable (assuming the feelings were potent enough), or they might simply add to the power of the incentive.

Second, the identity of the provider of incentives may be relevant. If the offer of incentives comes from a person of authority one might be reticent to reject it for fear of causing offence, disappointment, or similar negative attitudes directed towards oneself. Alternatively, the identity of the offerer might produce a 'signalling' effect. For example, if an individual's GP offers her an incentive, she might assume he has her best interests at heart and that she ought to accept the incentive because it is his recommendation. Such an influence may simply

enhance the allure of the offer (so long as the recipient is still choosing between an acceptable situation and a better situation), though the signalling effect could also produce greater apprehension in the recipient about the consequences of refusing the incentive.

Summary Remarks

In Part Two, I have sought to apply the general theories of coercion discussed in Part One to consider if health incentives are likely to have a coercive influence. Rather than applying a single theory of coercion, I have included various elements from different theories, including more situation-centric and agent-centric accounts, those using baselines to judge coercion, and those focusing on the psychological experience of the agent. This is because it is not clear that any one of these theories of coercion is 'right': they all identify different aspects of proposals that seem coercive. I think it is worthwhile trying to capture any coercive elements of health incentives, rather than making a rather simplistic evaluation of their coercive potential on a single account.

That said, I have emphasised the importance of the agent's psychological experience and the quality of her will, rather than those situational factors which could indicate coercion. It seems that if we care about coercion insofar as it affects liberty and autonomy then we must be most concerned with whether the agent is able to act voluntarily, and this will depend upon the interplay of external pressures and psychological factors. To this end, information about the situation and the proposal will be helpful in predicting and judging coercion, though will not be definitive.

On the accounts I have sketched, however, often certain necessary situational conditions are imposed for coercion to occur: there is an extent to which an incentive is either able to coerce, or not. For example, one may stipulate that the proposal must represent a loss from the recipient's baseline; that it offers the only alternative to an extremely unpleasant outcome for the recipient; that it makes the recipient worse off in some way; or that the recipient is so vulnerable that she is unable to make a rational decision to refuse the proposal. Once these conditions have been met, the extent to which the recipient actually experiences the effects of coercion will differ in degree. Thus, incentives that could coerce (in that they meet the necessary conditions for coercion) may, in practice, place very little coercive pressure on the recipient.

In fact, I have suggested that generally health incentives will not coerce. There are, however, certain structures of incentive schemes which, when targeted at susceptible individuals, could meet the criteria for coercion on some accounts. Thus, if coercion is to be avoided, where some or all the necessary conditions for coercion are met (for instance, the proposal is a threat relative to the baseline, or it meets Feinberg's *polarity* condition) great care will be required to minimise the pressure placed on the subject's will. For instance, an individual who is deemed vulnerable (perhaps she is in debt, addicted to drugs, mentally unstable, has a number of dependents, and so on) may be coerced where the incentive offered is posed as a quasi-threat, or is very large.

There is only so much that a philosophical account can do here. Further work to inform when coercion is actually likely to occur will depend on how situational factors affect different agents' psychologies, and the extraction of such information requires empirical study. Thus, I am unable to conclusively answer the question of whether and when health incentives coerce, but merely say that, on the accounts I have described, it seems possible, though unlikely.

Moving on from the case of coercion, I will now discuss how incentives might alter behaviour in other ways that could be problematic. Persuasion, manipulation, and nudging are more likely to be justifiable than coercion as they are generally assumed to have less of an invasive effect on the agent's will, and thus, are less destructive of liberty and autonomy. However, manipulation and nudging can be quite contentious, and may be thought to be unacceptable ways of altering health behaviour. In Part Three, I will introduce briefly the key features of these different ways of changing behaviour, and will then move on to discuss the moral implications of how health incentives alter behaviour in the light of this discussion.

Part Three: Persuasion, Manipulation, and Nudging.

Introduction

Clearly, it is not always wrongful for one or more agents to act to intentionally change the behaviour of another agent or group of agents. In the previous chapter I discussed the possibility that the project of promoting healthy lifestyles through behaviour change strategies may be an illegitimate one for governments to undertake. The concern in that instance related to the liberty of

individuals to determine their own values and lifestyles, and the respect shown to citizens by different degrees of state neutrality towards conceptions of the good.

I argued that under a moderate perfectionist set up, encouraging healthy behaviour through interventions like health incentives should be considered justifiable. However, the issue deserves further consideration in light of the effect health incentives might have on individual psychology and free will. Most of the time, the state ought not to employ political policies which hinder psychological liberty, even where the intention is to bring about politically legitimate changes. Both the conceptual aims of the state and the methods used to bring about these aims will affect liberty and autonomy.

Here, I introduce persuasion, manipulation, and nudging as means for behaviour change. I will be briefer than in the discussion of coercion: the conceptual literature on these other forms of influence is smaller, in part because they are typically of less concern to political and moral philosophers interested in legitimate action. The discussion of coercion should also be useful for considering the relative disruption of these forms of behaviour change.

There are also other tools for influencing behaviour, such as facilitation, seduction, education, deception, and exploitation. It is not possible to discuss all of these, though some of them may play a role in the methods I do consider (for instance, education may be an important component of persuasion; deception might be used in manipulation). Many of these concepts do not have settled philosophical definitions, but carry with them particular philosophical baggage and colloquial usage.

My concern here is how incentives influence behaviour, and whether this affects liberty and autonomy in impermissible ways. Looking at these different methods of influence, and how they appear to be involved in the use of health incentives will be a helpful approach, but not in itself conclusive. It still remains to be considered how health incentives, by employing such methods of influence, affect agent liberty and autonomy, and whether this is objectionable. The final part of this section will seek to consider these questions.

Persuasion

Persuasion is generally presumed to be one of the least problematic methods of intentionally influencing another agent's behaviour. This applies far more to

rational persuasion than *non-rational* persuasion: whilst all persuasion involves appeals to reasons to influence an agent's behaviour, rational persuasion provides reasons that encourage the agent to deliberate rationally, whereas non-rational persuasion encourages faulty assumptions and failures in critical reflection (Rossi and Yudell [2012]; Beauchamp and Childress [2001]; Wertheimer [1987]; Faden and Faden [1978]).

In its ideal form, rational persuasion does not disrupt deliberative decision-making and operates entirely within this process of reasoning. Thus, rational persuasion enables the kind of 'free' deliberation which Scanlon (1986) argues is necessary for their being subject to moral appraisal. This requires "critically reflective, rational self-governance… less appropriately thought of as a kind of freedom than as a kind of intrapersonal responsiveness. What is required is that we be importantly dependent on our process of critical reflection, that that process itself be sensitive to reasons, and that later stages of the process be importantly dependent on conclusions reached at earlier stages." (Scanlon [1986: 176])

Non-rational persuasion appears to provide agents with a reason, and similarly engage with her deliberative capacities, but it may fail to give her a *good, rational*, or *valid* reason. Consider: I may be persuaded to buy a brand of face cream because it is advertised as having diamonds in it, and the impression is given using scientific language that these diamonds are key to making my skin look brighter. However, this form of persuasion may be encouraging me to make false assumptions about the effects of diamonds in face cream and the reasons I have to buy the product.

Non-rational persuasion begins to cross over into the realm of manipulation (see the next section), and is more contentious. Some theorists, however, propose that we should be wary of all persuasive influence (Faden and Faden [1978]). In part, this is because it is extremely rare (perhaps impossible) for persuaders to employ *only* methods of rational persuasion: they claim that other means of influence such as non-rational persuasion and manipulation will almost always operate in tandem.

This problem could go deeper in that persuasion might necessarily be manipulative insofar as it involves one agent trying to get another agent to change her attitude or behaviour (Mendelsohn [1968]). If such intentions create unavoidable manipulation then this issue will afflict any method of behaviour

change intervention, all of which are aimed at altering how people behave. These different methods of influencing behaviour will only differ in the way those intentions are translated into interventions. Education and facilitation may lie at the far end of the spectrum here, but if they are employed with the explicit intention of altering behaviour then, on some accounts, they must employ at least a modicum of manipulation.

MANIPULATION

Manipulation, then, tends to be identified when features of an agent other than her rational reasoning capacities are exploited in order to alter her attitudes or behaviour, although a number of different variations on this have been proposed (Rossi and Yudell [2012]; Blumenthal-Barby and Burroughs [2012]; Rudinow [1978]). As mentioned, manipulation may be used in conjunction with other influences such as persuasion and coercion, so the boundaries become easily blurred.

Warwick and Kelman (1973) distinguish between situational and psychic manipulation. The former operates by setting the social structure up in such a way so as to alter the decisions people are likely to make, whereas the latter works at the level of the individual psychology of the manipulated agent. Particular situations may be such that they have an inherent tendency to make subjects more susceptible to external influence, including manipulation, and (as discussed earlier) the psychological make up of an agent may also predispose her to being influenced by others. Manipulation may thus be rather hard to avoid, particularly in contexts where the emotional relevance of behaviours are highly salient, where people struggle to effectively reason about decisions and behaviour, or where there is a significant power imbalance between agents (O'Neill [1984]).

Sometimes manipulation can involve a considerable degree of sophistication on the manipulator's part, who may take the trouble to find out precisely what 'buttons to push' in order to exert maximal influence over her subject (Cave [2007]). This may be highly personalised, or it may derive from specialist knowledge about the workings of human psychology in general.

Often, manipulation is secretive and may deceive the subject such that she is unaware of being manipulated. Secrecy and deception is often thought a key indicator (and an objectionable feature) of manipulation. However, there

are instances where we may be at least partly aware of the manipulative efforts of others, and their likely effect on us (for instance, the use of cute labrador puppies to advertise loo roll), yet still be influenced by such tactics. Thus, manipulation may vary in the degree to which it deceives the subject or is explicitly enacted.

Due it its failure to engage the deliberative reasoning capacities of agents (often quite the reverse), and its frequent use of deception and exploitation of individual weaknesses in order to wield influence, manipulation is often thought to be undesirable. Yet, it should be borne in mind that manipulative tactics for altering behaviour may (often) be justifiable, either because the manipulation itself does not appear particularly disrespectful, controlling, or freedom limiting, or because the benefits derived as a consequence of the manipulation are sufficiently valuable. The permissibility or otherwise of manipulation must thus be judged on a case by case basis.

NUDGING

'Nudges' are another addition to the repertoire of behaviour change interventions available to health promotors. It is perhaps better to describe nudges as a category of interventions that employ other, more basic means of influence (in the same way that television adverts might employ persuasion). Yet it is worthwhile including a discussion of nudges here because they have become extremely prominent in debates about behaviour change and healthcare policy. The kinds of behaviour change that nudges might target overlap with those targeted by health incentives: everyday, lifestyle behaviours as well as 'one-shot' decisions where, quite often, individuals' choices seem to be inconsistent with many of their long-term preferences.

Thaler and Sunstein (2008) were the first to appropriate the term 'nudge' and use it to describe a range of interventions which alter behaviour through small changes to the decision-making environment (the 'choice architecture' of a situation). Thaler and Sunstein argue that the interventions they describe fall within a philosophy of 'libertarian paternalism': nudges do not reduce the choices available to an agent (hence libertarian), but they encourage the agent to make decisions that are in her own best interests (hence paternalistic) (Thaler and Sunstein [2008]; [2003a+b]).

Numerous criticisms have been directed at the conceptual framework proposed by Thaler and Sunstein. Often it seems that the interventions they group together as 'nudges' are highly variable and they do not provide a clear definition of 'nudges' which links these diverse interventions. Further, the philosophical underpinnings of 'libertarian paternalism' are rather weak, in part due to misapplication of the concepts of both 'libertarianism' and 'paternalism.' It has also been argued that the empirical evidence used to back up claims about the efficacy of nudges is poor (Hausman and Welch [2010]; Marteau et al. [2011]; Bovens [2008]; Mitchell [2005]).

Despite these and other criticisms, nudges have proved very popular, in particular with policy makers. The two authors have landed jobs working for the US (Sunstein) and UK (Thaler) governments, and nudge-like approaches to policy-making in healthcare and other domains have been influential over the last few years (Vlaev et al. [2010]).

The kind of interventions Thaler and Sunstein describe as nudges include presenting the food available in a cafeteria in such a way that people are encouraged to choose healthy food over unhealthy food; setting the default such that employees are automatically enrolled to make contributions to pension schemes (an opt-out rather than an opt-in system); or providing information about how one's energy usage compares to that of one's neighbours. What distinguishes 'nudges' from 'shoves' may be tricky to determine (Marteau et al. [2009b]), though it could relate to the costs associated with resisting the nudge (buying the chocolate cake, opting out of pension contributions, not taking any steps to reduce energy consumption); the degree of voluntariness the choosing agent is able to demonstrate; the benefits gained by the agent; the intentions of the nudger; and other things.

Those describing the potential applications of nudges tend to draw heavily upon the psychological and behavioural economic literature relating to the 'bounded rationality' of humans and to 'heuristics and biases' which affect much of our reasoning and behaviour, relating, for example, to risk, decision-making, preference formation, and recalling from memory (see for example Slovic [2000]; Tversky and Kahneman [1974, 1981, 1986, 1991]; Kahneman [2012]; Kahneman and Tversky [2000] Gigerenzer and Selten [2002]).

One example is the phenomenon of loss aversion, described in the section on coercion. Framing something as a loss rather than a gain can mean it

will make a bigger impression on the agent. Loss aversion relates to many other phenomena, such as the endowment effect (also mentioned earlier) and status-quo bias (which describes, as one would expect, our natural preference for what we already know and have) (Thaler and Sunstein [2008: 34]).

There are many other examples of heuristics and biases, and the literature on this subject is fascinating, but too extensive to properly discuss here. The main point is that the use of this privileged understanding of human psychological functioning is put to work by nudgers like Thaler and Sunstein in designing behavioural interventions. This may qualify as manipulation, at least some of the time, and so we might think that nudges are often manipulative. Furthermore, as much of the literature to do with heuristics and biases and nudges asserts, most of the time we are unaware of the sorts of factors influencing our behaviour. Thus, such nudges will often be invisible to the agent subject to them, and so also employ deception.

I will discuss shortly the extent to which influencing behaviour through these methods might have problematic effects on liberty and autonomy, and relate this to the use of health incentives. However, it is worth pointing out at this point that nudges are often thought to *correct* errors in reasoning; discrepancies between preferences at different times; lack of will power; and other factors that lead to what is sometimes described as 'boundedly rational' behaviour. Thaler and Sunstein argue that nudges help people to act in ways that are in their best interests and that, on reflection, they would choose to act anyway (Thaler and Sunstein [2008]). This may be correct in some instances, though it will not be true all of the time, and there may be no way of ensuring that nudges always and only operate in this way. Nonetheless, interventions which tend to correct for undesirable idiosyncrasies of psychology may be evaluated differently in terms of their impact on the autonomy of the individual they influence.

The Influence of Incentives

Parts One and Two explored the concept of coercion and developed an account of if and when health incentives might coerce the recipient. I now wish to look at how incentives employ other methods of influence: persuasion, manipulation, and nudging described above, to encourage healthy behaviours.

As became apparent in the discussion of coercion, incentive schemes can take many different forms, and so different schemes may well involve different degrees of persuasion, manipulation and nudging. Further, the use of these tools to change behaviour may be permissible in some cases, yet not in others.

First, incentives may operate by providing an agent with a reason to alter her behaviour. Consider someone who is on the cusp of getting tested for an STI, but who feels that it is not quite worth the effort. The offer of a £5 iTunes voucher might tip the balance of costs and benefits in favour of getting the test: the situation where she gets an STI test and receives a £5 voucher is preferable to the situation where she doesn't get tested and doesn't receive the voucher. If the incentive factors into the agent's deliberation in this manner it seems to influence via rational persuasion.

Incentives for 'one-shot' behaviours, such as STI testing, cancer screening, vaccinations, and so on, could work in this way: attaching additional benefit to one decision option to tip the balance in its favour. Such an effect sounds like the kind of incentives I have previously described as 'trade-type' as they provide an extra reason in favour of performing the incentivised behaviour, making that behaviour now the preferred action. Similarly, incentives for lifestyle changes could provide a reason in a similar way, persuading the agent to adopt a healthier lifestyle because the addition of the incentive to that behavioural option elevates it to being the most appealing.

Aid-type incentives could act as persuasive in a slightly different way: the agent in this case is presumed to already prefer the healthy behaviour, so she does not need to be given further reasons to think it is the right action for her. However, her current reasons are somehow incapable of motivating the behaviour she generally thinks is best. Incentives can provide a different kind of reason, the significance of which may lie in its propitious timing. Whilst the reasons an agent has to be healthy might govern a general attitude that adopting a healthy lifestyle will be wise, the incentive (when effective) provides a motivating reason: a reason to act in a particular way *on this occasion*. Whether it does so via rational persuasion or not will depend on the kind of reason the incentive provides and how it influences the agent's will.

Often, persuasive methods are thought of as those which draw the agent's attention to pre-existing reasons she has for acting in a particular way, for

instance, by pointing out to her the risks associated with not getting tested for an STI. Sometimes agents neglect the relevance such reasons hold for themselves: a kind of ego-centric, 'optimism bias' can cause people to assume that, although there are definite risks associated with a behaviour, *they themselves* won't suffer the consequences (Slovic [2000: 366]). It is plausible that incentives have this effect, if their introduction acts to remind the agent of pre-existing reasons she has, rather than by creating a new reason for her to act. Whether incentives create a new reason or remind the agent of pre-existing reasons (or perhaps both) will depend on the unique circumstances of a particular incentive.

Health incentives may thus influence via rational persuasion. It is equally plausible that they could influence via non-rational persuasion. In such cases, the incentives might appear to be a reason for the agent to act in a particular way, when in fact it isn't (in the sense that it doesn't provide a rational reason to act, or a 'good' reason to motivate her behaviour). Agents who are liable to overweight the value of incentives might be non-rationally persuaded, or those who are mistaken in what they understand the benefits of a particular action will be.

Some might argue that attaching money to a particular behavioural choice is not persuasive (does not engage with an agent's rational deliberative capacity), but rather, is manipulative. This could be the case if agents typically found cash incentives disrupted their deliberative capacities. In such cases, rather than being able to evaluate the costs and benefits associated with different courses of action, agents' reasoning processes might be confused by the introduction of a cash incentive. Money is thought to have peculiar effects on the way people make decisions, sometimes by mediating emotional affect. This, in turn, can have diverse effects ranging from encouraging self-sufficient behaviour to exaggerating autonomic responses (more impulsive, risky behaviour) (Vohs et al. [2006]; Bechara [2005]; Mantzari et al. [in submission, personal communication]).

Some of the ways money influences decision-making are not well understood and it would be rash to make general claims about the influence of incentives. A study by Mantzari et al. (in submission, personal communication) indicates mixed effects of monetary incentives on risk information processing, assimilation and understanding. The offer of money to take a pill as part of a

medical trial does not appear to undermine risk information processing or knowledge of side effects, but does affect the time participants spend viewing information about the risk of side effects.

Where health incentives do disrupt decision making due to the diverse effects money may have, and where these effects are harnessed in order to bring about desirable changes in behaviour, incentives may be considered manipulative. Some analogous areas where worries relating to the potentially manipulative influence of money on health-related decisions include payment for organ and tissue donation and participation in research trials (Becker and Elias [2007]; Olbrisch et al. [2001]; Dickert and Grady [1999]; Grant and Sugarman [2004]).

Once again, the experience of the individual is crucial here: the same incentive could act as a rational, reasons-compatible tool to persuade one individual to adopt a healthy behaviour, whilst disrupting the deliberative capacities of another and manipulating her. As discussed in the above section on coercion, identifying just how incentives actually work and the way they are likely to influence people will require psychological, rather than philosophical, insight. Even then, it seems that we will never be able to predict with certainty how a given individual will respond to a particular incentive.

Incentive schemes might also be specifically designed to exploit psychological features of agents in order to achieve maximum effect. Providing incentives at regular intervals structures efforts to change lifestyle behaviours by establishing short term goals and giving immediate, rewarding feedback. Some variations on this design may be tried, for instance, a scheme incentivising HPV vaccination uptake (where the individual must receive three separate injections of the vaccine on different occasions) offered different sized rewards of £20 for the first injection, £5 for the second, and £20 for the third and final injection (Mantzari et al. [in submission, personal communication]). Varying rewards in this way might be more successful at encouraging agents to receive all three injections than paying £15 for each injection or other configurations.

There is other information about human psychology that might be used to enhance the efficacy of incentives. Loss aversion and the endowment effect have been mentioned already (in the context of the *custodian* and *habituation* schemes I described in Part Two). These involve framing outcomes in terms of

losses rather than gains in order to enhance their 'force,' and encouraging agents to incorporate incentives not yet earned into what they consider to be 'their' possessions (or reference point).

Other examples could involve the selection of the sort of incentive used: cash might be more motivating than vouchers because of its fungibility and immediacy, though sometimes vouchers for luxury goods items are preferred (these can license the recipient to treat herself in ways she might not otherwise do) (Reilly et al. [2000]; Kivetz and Simonson [2002]). Behavioural economic research shows that people tend to overvalue lotteries with small probabilities of large rewards (Tversky and Kahneman [1974]; Gonzalez and Wu [1999]), so this might suggest offering lottery tickets as incentives will have greater motivational power than cash of equivalent expected gain.

Such design features might help influence agent behaviour, and they seem to do so by exploiting privileged knowledge that behavioural science experts have access to about certain foibles of human behaviour. Further, this knowledge often relates to non-deliberative features of decision-making behaviour, or at least, non-rational deliberation. Certain aspects of incentive schemes, then, appear to utilise the kind of sophisticated manipulation described earlier.

We might still distinguish between different aspects of the manipulative effects of incentive schemes. Providing rewards at regular intervals works on non-deliberative features in order to achieve an effect, but seems of a different sort to the manipulation involved in knowingly offering lottery tickets with a small chance of payout (that may nonetheless motivate misguided / optimistic agents). The latter might appear more cynical, exploiting as it does our tendency to erroneously evaluate lotteries. Further, some methods of manipulation may be more or less deceptive, or secretive about the basis for their efficacy.

An example of an incentive scheme that uses some of these manipulative techniques is 'Weight Wins,' which offers rewards for weight loss. On the front page of its website (on the 19/01/2013), Weight Wins features two individuals, one planning to lose 50lbs (23kg) in 12 months (who will earn £1185 if successful), and one who has successfully lost 56lbs (25kg) in 12 months (and earned £418 for doing so) (Weight Wins [2013]). Such examples may be misleading to potential participants with regards to the kind of success both in terms of weight loss and monetary rewards accumulated that they might expect

to achieve. On the 'Pounds for Pounds' scheme, an NHS pilot operating through the Weight Wins organisation, the mean weight loss for participants was 6.4kg. Clinically significant weight loss (defined as losing 5% or more of baseline weight) was achieved by 180 of the 402 participants (Relton et al. [2011]). Clearly (and unsurprisingly) the people behind Weight Wins select the most successful and ambitious participants to include in their promotional material.

Nudges are a form of intervention in themselves, and may employ a range of methods to influence behaviour. Nudges can provide informational feedback, emphasise the appeal of certain choice options, encourage the salience of particular social norms, exploit the 'stickiness' of defaults, and so on. The key interest of nudging lies in its link to the behavioural science literature and heuristics and biases. Rather than focusing on how a specific individual might be motivated to act in a particular way, nudgers look at trends of behaviour and general tendencies for 'choice architecture' to influence the decisions people make.

The main difference between incentives and nudges lies in their explicitness: nudges are typically secretive and hidden from the view of the person being nudged. It may be that, were the agent's attention drawn to the existence of the nudge, it would lose its effect. Incentives, in contrast, absolutely require the agent to attend to them: an invisible incentive will not motivate action.

Yet, in some sense, we can treat incentives as a kind of nudge. I have emphasised in this chapter the importance of the agent's experience and the effect on her will of some external intervention. Effective nudges, just like incentives, tweak the choice set available to someone in order to encourage a particular decision at the crucial time of action. Incentives make one behaviour (such as choosing to eat lettuce rather than cake) more appealing because of the prospect of a reward; nudges make that behaviour more appealing because of the position of lettuce in the canteen, or by informing the diner that 'most people choose lettuce,' or by providing lettuce rather than cake as a default addition to a meal. In the 'heat' of the moment, many considerations will have different pulls on the agent's will, and incentives and nudges are just two more inputs into her decision, and she may be more or less aware of the influence such inputs are having on her behaviour.

Incentives which are not overly large will surely not cross the 'libertarian' line of libertarian paternalistic nudges, insofar as they will not dramatically increase the costs of any of the choices available to the agent. The paternalist requirements of nudges may not be met by incentives, which could certainly fail to promote the best interests of the recipient. Yet, the same is the case for nudges: despite Thaler and Sunstein's insistence that nudges are paternalistic, it is not clear that there is any way of ensuring this.

Both trade-type and aid-type incentives could nudge: trade-type incentives by making one choice more appealing than it previously was; aid-type by compensating for the lack of tangible, immediate and rewarding feedback from healthy behaviour and providing motivation at the time and of the form needed to support agents in altering their lifestyle behaviours.

The way nudges work looks to be manipulative in that they do not tend to change behaviour by engaging the rational, deliberative capacities of the agent. However, an interesting feature of at least some nudges is that they seek to dampen non-rational influences on behaviour (or what are perceived to be non-rational influences), and thus allow agents to make decisions based more upon deliberation and stable preferences. Yet this activity takes place in the shadows: nudges, typically, are secretive and hidden from the conscious awareness of the one nudged. Further, nudges may not reduce the effect of bias, but simply add to it (albeit in ways that may make the agent better off) (Ashcroft [2011]).

Aid-type incentives may correct for 'irrationalities' in agent decision-making behaviour as well, supporting behaviour change in those who desire it. In this sense, incentives may also lay claim to not manipulating agents to exploit irrationalities in reasoning, but rather, counteracting those same irrationalities. In contrast to nudges, however, incentives are explicitly provided: the whole point being that agents must know about the prospect of the reward in order for it to have motivational power. Thus, one feature often thought to be involved in manipulation - secrecy - is absent at least in the fundamental action of incentives (as mentioned above, certain aspects of the design of incentive schemes may involve deception or secrecy).

Health Incentives, Liberty, and Autonomy

This chapter has sought to look at the means of influence health incentives use to effect behaviour change. If nothing else, it has argued that incentives, at least

in theory, can alter behaviour through a whole variety of different ways. In part, this is because the scope for designing different sorts of incentive schemes is so wide, and as such, different schemes can vary a great deal from one another. Additionally, different methods of influence over behaviour are poorly defined and frequently overlap even in the abstract. In real life cases, behaviour change interventions will invariably employ different combinations of coercion, persuasion, manipulation, and nudges, as well as other means of influence, in their operation. Incentives are no exception. Finally, individuals differ, and the way incentives affect different individuals will depend at least in part on psychological features of the agent.

I wish now to consider how such influences might impact upon liberty and autonomy. This will only permit, at most, indirect conclusions about incentives because (as discussed) generally incentives will employ multiple means of influence and the impact of different influences on agent liberty and autonomy will be moderated by contextual facts about the agent, the incentive, and other situational factors. However, this discussion should help to illuminate the potential for incentives to have harmful effects on liberty and autonomy.

Earlier, I discussed how autonomy has historically played an important role in bioethics. Too often, however, autonomy is presented as the most fundamental value of human existence, to be promoted above all others (Dawson [2010]). Yet there are many things we value (well-being and justice to name two), and it is not clear that we can sensibly incorporate so many values into a single concept, nor that we consider them to always be subordinate to autonomy.

Autonomy should not, therefore, be privileged above all other values at risk of destruction from disease or its treatment. Nor should it be the only value we seek to promote through healthcare (either in a proximal or distal sense). The other chapters in this thesis consider some of the other values of relevance to the ethical evaluation of a healthcare interventions. Despite devoting this chapter to the consideration of how incentives might impinge upon liberty and autonomy, I do not wish to suggest these two values are more or less important than other values both present and absent from this thesis. However, it does seem that these values are the basis for concern where arguments point to the influence incentives have on people's capacity to choose how to behave.

I emphasised earlier that when considering the negative impact of incentives on liberty and autonomy, it is important to have a concept of *valuable* autonomy in mind. That is, we should probably not be concerned about any kind of influence on an agent's behaviour, no matter how small or benign. Rather, we should worry when those influences significantly limit the agent's liberty to choose how to live her life, and disrupt her in a general pursuit of a self-determined life (and pursuit of the good as she conceives of it). Thus, it is a relatively modest form of autonomy I am concerned with, and one which seeks to be responsive to the limitations of actual human agents to direct their lives.

In this section, I will discuss how those coercive, persuasive, manipulative, and nudging influences might bear on autonomous action, and what this might mean for the moral status of incentives. I will then also provide some discussion informed by research from health psychology, relating to the mechanisms by which we control our behaviour. My intention here is to consider how incentives interact with our mechanisms of behavioural control and in particular how trade- and aid-type incentives can be thought to have different influences here.

I argued that although it is conceptually possible, incentives will rarely, if ever, prove coercive. Key to this conclusion was the assertion that incentives are unlikely to place sufficient psychological pressure on the agent to restrict her ability to make voluntary decisions in such a way as to be considered coercive. Were incentives to coerce, they would also *prima facie* restrict liberty.

Liberty restriction is not always unjustifiable: it is at least likely to be permitted where it prevents one individual from further restricting the liberty of (or harming) another (Mill [1859/2006]). However, generally, liberty restriction is difficult to justify, particularly where it is enforced without some form of consent from the individual concerned. Regulations to prohibit smoking under a certain age and requiring seat-belts to be worn in cars are supported by explicit paternalistic reasons, and are generally well-accepted. Yet the introduction of liberty-restricting legislation is often met with hostility and state governments are often (sometimes inappropriately) reticent to use such regulation (House of Lords [2011]).

The potential for incentives to restrict liberty looks weak. As discussed, incentives are unlikely to place sufficient pressure on the recipient's will through coercion, or indeed persuasion, manipulation, nudging, or other means, to threaten the voluntariness of the agent's choices and behaviour. The extent to which incentives actually do pressure recipients' wills is a matter for empirical assessment, yet common-sense tells us that relatively small rewards (£10 or £20 a week) would be unlikely to place a significant (liberty-restricting) pressure on the majority of individuals.

Liberty might be threatened were the agent's deliberative processing to be disrupted significantly, to the point where she is sufficiently confused or distracted that she is no longer able to make a 'free' decision. However, we ought to take liberty to be a reasonably robust notion, present in the majority of agents making decisions under ordinary circumstances.[24] This liberty ought not to be vulnerable to disruption from offers such as a relatively small cash sum. If this were the case, there would be many greater threats to liberty than that posed by incentives. Through pressure or disruption, then, it looks unlikely that incentives will be liberty-restricting.

Where agents lack the capacity to understand the options open to them and / or are unable to make a well-informed decision about how to act then they may be said to lack *positive* liberty. This form of liberty is closely connected to autonomy, the general capacity for self-determination. Autonomy will be disrupted where an agent's choice set is significantly altered such that choices that are, in principle, available to her appear unavailable, or she is unable to choose them for some reason.

Generally, influences that are liberty restricting will also be autonomy restricting.[25] Coercion, then, and any other influence undermining liberty through pressure or disruption to the agent's will, will also undermine autonomy. But autonomy will also be vulnerable to influences which 'hide'

---

[24] Clearly, such an *ad hoc* specification of liberty is not entirely satisfactory, and were my intention to fully justify the value of pursuing liberty then a more comprehensive discussion would be necessary, and a better explanation of how to specify the lower limits for 'liberty' would be needed. However, I do not intend to provide such a justification here. I assume that practices which accord with our everyday notions of 'liberty' will be acceptable, and take this to be the standard against which incentives (and the means by which they influence the agent) should be judged.

[25] There can be exceptions, which I will discuss later.

choice options, or manipulate the agent's behaviour in more subtle ways so as to control her decision-making. Where the offer of cash has an especially large (though not coercive) influence on the agent, her autonomy may be compromised by making alternative (non-incentivised) decision options significantly less 'available' to her.

Autonomy, further, requires that agents are able to direct their own lives; that they 'authentically author' their behaviour. Kantian autonomy places particular emphasis on rationality as a fundamental component of autonomy: the agent who acts inconsistently and irrationally cannot be autonomous in this sense. To be an autonomous agent will also require one to be treated with a degree of respect by others. The Kantian tradition characterises this as being treated always as an end in oneself, and never merely as a means by which others may achieve their own ends (Kant [1785/1993]; Korsgaard [1996]).

Autonomy in the sense of being respected as a self-determining, rational agent, is a richer notion than positive liberty. Clearly, it will be valuable to have power over one's own life course and not merely to be used as an instrument by others. However, it is not always obvious how agent autonomy can be protected and promoted, for a number of reasons. For instance, autonomy can have a rather complex interaction with freedom (in the sense of both positive and negative liberty); the time scale autonomy may be considered over can vary (less autonomy now may be necessary to guarantee future autonomy); it will always depend on agent psychology and so is highly sensitive to individual idiosyncrasies; it can interact with other values in variable ways (some of which may contribute to agent autonomy, or may require its sacrifice).

To begin, there are two important candidates for arguing that providing health incentives is not compatible with a concern for agent autonomy (aside from the ways in which incentives could hinder liberty through coercion or extreme pressure, although as argued, this seems very unlikely to arise). Both of these involve incentives operating in ways that would be described as manipulative on my account.

First, incentives may cause people to act for inappropriate reasons. This can occur at the quasi-conscious level: the agent who is so distracted by the offer of cash that she fails to evaluate her other options properly. It may also be unconscious (or preconscious): the agent who is unwittingly more moved to be

motivated by the offer of a lottery ticket because she overvalues the small probability of winning the prize. We would not wish to say that such an agent is unfree to choose properly between different options, however, her capacity to do so is either compromised by the offer of money, or influenced by inappropriate features of the scheme that she may be unaware of.

This kind of manipulation (as well as non-rational persuasive influences) could undermine autonomy by limiting the extent to which the agent acts as a rational, reasonable (and reasons responsive) individual. Whilst she might view herself as acting according to reasons which she recognises as 'good' (and which she considers move her appropriately), in fact she is acting according to poor reasons, or things which are not reasons at all.

Second, the provider of the incentive may manipulate the agent's behaviour in order to further the provider's own ends. Some public health interventions provide benefits at the societal level, and sometimes the individual herself will be unlikely to benefit, and may even experience harm. For example, a given individual may do better not to receive a vaccination, and to rely on herd immunity instead. Yet public health interventions which aim to create herd immunity must target (almost) everyone in the community. Incentives may thus be offered where the intention is only partially, or perhaps not at all, to benefit the recipient, but rather to get her to alter her behaviour in order to benefit others (or society at large). It is also important to note that even where the intention *is* (entirely or mostly) to benefit the incentive recipient, the provider of the incentive may judge wrongly what is in the recipient's best interests. Thus, encouraging someone who derives extreme pleasure from eating fatty foods to diet may not be a good way of promoting her interests.

Agents who are manipulated in this way are not given the opportunity to act as self-directing individuals, with their own values and life plans. As such, they are not treated with the respect owed to autonomous moral agents. Further, where their interests are not promoted by the manipulator, then agents subject to such manipulation may have their future interests harmed, suffering not only a loss of autonomy but potentially well-being as well.

Manipulative influences which encourage action that is based upon inappropriate reasons, or which is designed to make instrumental use of the agent, may hinder the agent's capacity to act in ways consistent with being an autonomous, self-determining agent. Further, such actions fail to treat agents

with the respect due to autonomous persons. However, influence which does not manipulate in such ways, but instead engages the agent's deliberative capacities to make decisions about how she wishes to behave and to make (so far as can reasonably be expected) autonomous choices, will not have such a disruptive effect on autonomy.

Reflective and Impulsive Control Over Behaviour

As mentioned, when discussing the effect healthcare interventions such as incentives may have on liberty and in particular autonomy, it is important to begin with a realistic conception of human agency. When discussing decision-making, and particularly the interference with the decision-making behaviour of other agents, there is a tendency to exaggerate the value of 'independence.' In part, the assumption that independent (uninfluenced) decision making is preferable seems to rest on a false assumption that much human behaviour is determined in this way. This is supported by the 'fundamental error of attribution'; the tendency to overestimate how much one's character is responsible for one's behaviour (and the underestimation of the importance of situational factors in determining behaviour) (Nisbett and Ross [1980]; Ross [1977]).

It is, I think, worth saying a little about current social psychological theory on behavioural control, which has particular relevance to health behaviour and the evaluation of behaviour change interventions. This is the 'two systems' model, which describes how behaviour control operates through both reflective and impulsive systems. The following discussion draws heavily upon Strack and Deutsch's (2004) influential article which builds on previous dual-process accounts and focuses particularly on the interaction of the reflective and impulsive systems, and how these systems may operate synergistically or antagonistically.

Furthermore, Strack and Deutsch consider the effect of the two systems on behavioural outcomes, rather than on judgements and information processing (often thought of as the antecedents to behaviour, but in fact not always present). Strack and Deutsch's account therefore seeks to make sense of phenomena such as *akrasia* (weakness of the will) and instances where agents do not act to maximise utility, or in ways that appear to thwart their preferences. This is of particular relevance to health behaviour where agents

often act to achieve short term benefits and in ways inconsistent with their long term goals and values.

Reflective system (also know as 'system 1') behaviour is based upon reasons, values, and an understanding of the future and what desirable or undesirable consequences will be brought about by present action. Thus, the reflective system directs behaviour as a result of conscious reasoning processes. In contrast, the impulsive system works through associative links and often controls habitual (routine) and unconscious (or pre-conscious) behaviours. The impulsive system therefore requires far less cognitive effort than the reflective system, can operate more quickly, and is less vulnerable to disruption from suboptimal conditions.

It is thought that around 45% of behaviour is habitual, and mostly controlled by the context-cued impulsive system (Neal et al. [2006]; Marteau [2010]). This will include many everyday behaviours which make up one's 'lifestyle' as a whole, such as dietary choices and exercise. It takes a large cognitive effort to intervene in habitual behaviours under impulsive control and to alter this behaviour to bring it into line with reflective processes: this is often what is required for agents to resist the chocolate bar and choose fruit instead, or to make similar healthy substitutes for unhealthy behaviours.

Only the reflective system is capable of explicitly generating a time perspective (and thus adapting present behaviour in order to achieve future goals). Often, then, unhealthy behaviours are maintained by the operation of the impulsive system, responding to immediate cues and desires. Where the impulsive and reflective systems act antagonistically, agents may struggle to modify their behaviour in line with their reflective wishes, and impulsive behaviours may be enacted in spite of efforts to inhibit them.

The consideration here, then, is how incentives interact with these systems of control, and what relevance this holds for agent autonomy. The distinction I wish to draw is between trade- and aid-type incentives, and the different impact they have on the orientation of the agent's behaviour. In the trade-type case, the agent (before the offer of an incentive) is happy engaging in her unhealthy behaviour (be that smoking, eating a poor diet, failing to get vaccinated, and so on). It is generally the case that behaviours which produce immediate pleasure and which tend to be cued by our environments are those which will typically have negative effects on health in the long run. Thus, agents

subject to trade-type incentives can be assumed to have their reflective and impulsive systems oriented in the same direction (towards unhealthy behaviour[s]).

In contrast, aid-type incentives are those offered to agents who already hold some desire to change their behaviour. In these agents, the impulsive system leads them to display unhealthy behaviours, whilst the reflective system struggles (and generally fails) to bring about healthier behaviours. The impulsive and reflective systems of the agent receiving aid-type incentives are therefore oriented in different directions.

When the incentive is introduced, let us assume that this could either fail to alter the orientation of either system; it could succeed in altering the orientation of one system but not the other; or it could succeed in altering the orientation of both systems. It is not clear that we can state which system incentives operate through: potentially, an immediate reward could figure in the operation of the impulsive system, although it seems more likely that it would be involved in the reasoning of the reflective system.

In the aid-type case, then, the incentive orients towards healthy behaviour and away from unhealthy behaviour. This is consistent with the agent's reflective system, but contrary to her impulsive system. This agent already has antagonistic systems, and an incentive may be effective in changing behaviour in one of two ways. First, the incentive may simply boost the power of the reflective system to determine eventual behaviour (by adding additional reasons or motivation for behaviour, or structuring motivation in some helpful way, or similar). In this case, the agent adopts the new (healthy) behaviour, which is consistent with her reflective preferences, though not with her impulsive desires.

Second, the incentive may bring these systems into line, resolving the conflict by re-orienting the impulsive system towards healthy behaviour. We may accept that the incentive is more or less successful so, although it may bring the two systems into alignment some of the time, it may fail to do so all of the time. Moreover, this second effect may take time to become established, and may result from the first: the agent is initially able to act in line with her reflective system by over-powering her impulsive system but, as the healthy behaviour is repeated more and more, it becomes habitual and comes under the control of the impulsive system.

In the aid-type case, the effective incentive will either cause the agent to act in line with her reflective system (though a conflict between reflective and impulsive systems will remain), or it will resolve the conflict between both her systems, bringing her behaviour into line with both reflective and impulsive systems.

The trade-type case differs, because here the agent's systems are initially oriented in the same direction and she experiences no conflict. The successful (behaviour-changing) incentive will either have the effect of altering the orientation of one or other system, or it may alter the orientation of both together. If the incentive has this latter effect, the agent will alter her behaviour, will not experience conflict between her two systems, and will act in line with her reflective system. If the (successful) incentive re-orients only the reflective but not the impulsive system, the agent will alter her behaviour and this will be consistent with her reflective preferences, though not with her impulsive desires. Alternatively, the agent's impulsive system may be re-oriented, but not her reflective system, leading to her acting consistently with the former but not the latter.

Sometimes, then, the incentive may change behaviour but create a conflict that was previously absent between the agent's systems of behavioural control. There is another concern with the trade-type case, and that is where the incentive re-orients one of the systems but fails to change behaviour. Here the incentive creates a conflict, as before, but does not produce any health benefits that changing behaviour would produce.

Seeking to alter behaviour with trade-type incentives would seem to create potential problems for autonomy. Creating a conflict that was previously absent between the agent's reflective and impulsive systems may be problematic. Such antagonism can create feelings of temptation and frustration in the agent, and undermine her experience of being able to fluently control her behaviour (Strack and Deutsch [2004: 230]). Further, the incentive may cause the agent's behaviour to diverge from her reflective preferences. This will occur either where the impulsive system is re-oriented and the behaviour is changed, but the reflective system remains oriented towards the original behaviour, or when the reflective system is re-oriented towards the new behaviour but the agent does not alter her behaviour (the impulsive system still controls the final behaviour).

Although the incentive could result in either the reflective or the impulsive system being frustrated, it seems worse from the point of view of autonomy that the reflective system ends up opposed to the action performed. This is because it is the reflective system which directs reason-based action, consistent with the agent's values, goals, and self-perception. Where the agent's behaviour is not in line with such things, she becomes alienated from her behaviour.

Trade-type incentives, then, risk creating conflict within an agent in relation to her two systems for behaviour control, and between an agent's reflective system and her behaviour. Whilst the trade-type incentive may bestow benefits in terms of well-being and improved health, it may sacrifice a degree of agent autonomy in bringing this about. Aid-type incentives, on the other hand, would seem to promote autonomy: in this case, the incentive either brings behaviour into line with the agent's reflective system, or it brings both the agent's systems into agreement.

This has been a necessarily simplified discussion of how behaviour might be altered using incentives. The real picture will be far more complicated, particularly as preferences may change depending on time and situational factors, and behaviour may be inconsistent to begin with. It is a further complication, still, to consider how the methods of influence described earlier - coercion, persuasion, manipulation, nudging - feed into the way trade- and aid-type incentives influence behaviour through the impulsive and reflective systems. This, I suspect, would be a step too far into speculation in a discussion which is already vulnerable to such a criticism. I shall leave it here, then, but I hope to have provided at least a plausible account of how incentives might influence agent behaviour, and the impact this might have on liberty and autonomy.

Concluding Remarks

Influences which pressure agents to act in particular ways; exploit peculiarities of (non-rational, non-deliberative) decision-making; disregard the agent's own interests and rather promote the interests of others or society, may undermine liberty and autonomy. Where incentives are likely to do this, they may be less defensible than where they avoid these effects. My aim in this chapter has been

to describe what kind of features of proposals will result in different forms of influence, to assess the extent to which health incentives possess such features, and finally to consider the impact of health incentives on agent liberty and autonomy.

It is worth reiterating that trade-offs may be required. Examples of permissible coercion have been described, but there will also be cases where concerns for welfare, justice, and other values (including the autonomy of other agents) justify interventions that limit liberty and autonomy. It thus matters how much harm an intervention does to agent autonomy, and what is gained from the intervention, and what harms might arise from non-intervention. However, without knowing what the benefit-side effects of an intervention will be, those interventions with the least potential to undermine liberty and autonomy will look more likely to be justifiable.

Incentives likely to result in coercion will probably be difficult to justify. As discussed, in reality, this seems unlikely to occur given the size of incentives generally offered and design of incentive schemes. However, where incentives are large, structured to play on loss aversion and other psychological biases, offered to vulnerable individuals by powerful others, and so on, then they will have optimum potential to coerce, or otherwise have an illegitimately powerful influence over the recipient.

Incentives that encourage agents to act from something other than rational, deliberative decision-making processes, may also be difficult to justify, due to an undermining effect on autonomy. This is a more demanding requirement than simply avoiding coercion, and may rule out more instances of incentive use. For instance, it may be impermissible to target agents likely to struggle to deliberate once offered incentives. If psychological or social risk factors were known such that agents like this could be identified (for instance, if this were descriptive of a particular group, such as drug addicts or people suffering from particular mental illnesses) then it might be inappropriate to use health incentives as an intervention for behaviour change in those groups.

Another area of concern will arise where incentives fail to promote the interests of the agent. This may occur because the agent is not the intended (main) beneficiary of the intervention, or where she is intended to benefit but the intervention fails to promote her interests effectively. The latter will be harder to avoid than the former, although all schemes must be considered

carefully to ensure that their likely outcomes have been evaluated as accurately as possible. The more widely targeted the incentive, the more difficult it will be to assess whether the population influenced by it will be likely to benefit (or be harmed) by the intervention.

Incentives that exploit cognitive or volitional deficiencies may also be problematic. This means that schemes which use lotteries may be less justifiable than those which use certain rewards. This is a difficult line to draw, as all well-designed behavioural interventions (incentive schemes included) ought to consider how best to achieve cost-efficiency. Nudges seem more prone to this potential for undermining autonomy than incentives, as they frequently employ secretive, non-cognitive means of behaviour alteration, such as setting defaults or exploiting social norms so as to influence behaviour. Sometimes such interventions will, in reality, seem fairly benign and their impact on agent autonomy is, I believe, limited. It has been suggested by Bovens (2008) that the test for legitimacy here should lie in whether agents can, if they so desire, identify the means by which they are being influenced.

It may also be argued that, insofar as nudges work to *counteract* the effects of biases ever-present in human psychology, they may help to promote rational agent action, and thus, autonomy. Similarly, incentives can support agents to act in accordance with their own long term, considered, rationally deliberated life plans. In the section on the reflective and impulsive control of behaviour, I suggested that aid-type incentives seem to act (where they are successful) to bring an agent's behaviour into line with her considered preferences, and perhaps to resolve conflict between her two systems of behavioural control. Such synergy would appear to be promoting, rather than undermining, of agent autonomy, facilitating more consistency between the agent's preferences, goals, intentions and her behaviour.

The two systems theory also suggests the potential for trade-type incentives, targeted at individuals with no prior intention of altering their behaviour, to create conflict. Antagonism between an agent's two systems, or alienation of her behaviour from her reflective system is likely to be autonomy undermining. A desire to preserve autonomy, then, may lead us to avoid offering trade-type health incentives to individuals.

Rather than indicating that incentives are, on the whole, permissible or impermissible, this discussion has shown that it is very much dependent on

features of the incentive scheme, the agent, and the surrounding situational factors. In short, the effects of health incentives on agent liberty and autonomy will be highly *context dependent*. However, I have indicated certain features of schemes that will make them more likely to have pernicious effects on liberty and autonomy. Although each individual will experience incentives differently, it seems likely that, where a number of these features cluster together (which could occur when targeting incentives at particular groups or at certain diseases) incentives may have more potential to undermine agent liberty and, in particular, autonomy.

The justification of the use of health incentives in a given situation cannot be determined by consideration of liberty and autonomy alone. Rather, the other effects of such interventions on things we value - their ability to promote goods or create harms - will need to be taken into consideration for an overall judgement to be made.

Five

The Difficulty with Desert: Fairness, Responsibility, and Freedom

Introduction

So far, I have considered a number of themes of criticism relating to health incentives. Chapter Three looked at some political features of healthcare interventions: what kinds of values the state ought to promote (through healthcare); where the limitations of state intervention should be set; how the state ought to aim to distribute resources amongst its citizens. In Chapter Four, I looked at a more personal aspect of healthcare interventions: the effect on the individual's psychology, and the extent to which different means of influence have the capacity to constrain one's freedom and autonomous action.

There are, of course, always overlaps between the political and the personal. Discussions of the political permissibility of an intervention are in part concerned with how that intervention will affect the individual. Similarly, questions about individual interactions can be 'scaled up' to concerns about patterns of treatment by the state and population wide effects. The topic considered in this chapter particularly encapsulates the dual relevance of bioethical concerns to political and moral philosophy. Here I will discuss the issue of desert and how we should conceive of people as responsible agents and treat them fairly. Desert and responsibility are both notions typically operating at the personal level, where agents interact with one another. Yet it is also common to extend this notion of desert to the political level, and to demand that

states treat citizens fairly and as they deserve to be treated. Though highly intuitive, there are difficulties with this extension of desert into the political sphere which will become apparent.

The concern about desert is often spelled out in the press in terms of incentives 'rewarding unhealthy behaviour'. This criticism could also indicate a concern that rather than discouraging unhealthy behaviour, incentives actually encourage it. Such a perverse effect of incentives would mean they entirely fail in their intended task, becoming worse than useless. Once again, this is a criticism that is dependent upon there being evidence which shows such an effect of incentives, and is not something I make any attempt to support or challenge here.

The aspect of 'rewards unhealthy behaviour' that I do intend to consider here is that which relates to fairness. Specifically, that incentives do not treat people according to their *just deserts*. This can either mean that those receiving incentives are getting *more* than they deserve, or that those not receiving incentives are getting *less* (or perhaps that both these are true).

In this chapter I will flesh out the claim that some individuals are undeserving of incentives, and thus that incentives are unfair. I will look at the notion of desert, as well as the related concepts of responsibility and blame. I consider how the distinction between incentives as a psychological aid, and incentives as an economic trade may be of relevance to the assessment of the fairness of incentives. I outline a number of arguments that criticise incentives on the basis that they provide rewards for people who are undeserving of them (or less deserving than others). In seeking to respond to some of these criticisms, I consider the link between moral responsibility and health behaviour. For this, I draw upon Philip Pettit's work which conceptualises freedom as 'fitness to be held responsible,' and consider how circumstances of deprivation, and the psychological mechanisms controlling health behaviour fit into this account of freedom. I argue that there is reason to think that, at least some of the time, agents will not be entirely free, nor fully morally responsible, nor deserving of blame for their unhealthy behaviour.

WHY CONSIDER DESERT?

The criticism that incentives fail to treat people in a just and fair manner arises in the media under the guise of a number of arguments: Misuse, Rewards Unhealthy, Responsibility, Nanny State, Giveaway, and Universal can all be construed as involving claims that health incentives fail to meet some standards of justice (Parke et al [2011]).

The aspect of justice that is striking in the 'rewards unhealthy behaviour' claim is that of *desert*. The criticism, as I consider it here, is that some (or all) of those who are offered incentives to adopt healthier lifestyles are not deserving of this form of healthcare intervention.

Yet most recent work on justice in political philosophy has focused on the distribution of resources in society. If desert were included as a principle of distributive justice, it would look something like: 'a just distribution will be that which treats everybody exactly as they deserve to be treated.' This would mean that the patterning of resources across society (or, perhaps, the patterning of *access* to resources) would reflect the deservingness of individuals.

Although intuitive and commonplace in our ordinary thinking about fairness, desert is relatively unpopular as a principle of justice. Samuel Scheffler describes how "none of the most prominent contemporary versions of philosophical liberalism assigns a significant role to desert at the level of fundamental principle." (1992: 301) Scheffler points to a whole range of theorists whom he identifies as neglecting, or explicitly excluding any principle of desert from their conception of distributive justice. These include the contractualists John Rawls and T. M. Scanlon; utilitarians such as J. C. C. Smart; 'Lockean libertarian' Robert Nozick; and communitarian Michael Sandel.

This consistent exclusion of desert from such a variety of theories (albeit, theories still falling within the broad 'liberal' tradition) occurs because of the difficulty inherent in identifying agents as morally responsible, and thus deserving, of any features which leads to their having a greater or lesser share of the distribution of resources in society. For instance, imagine that intelligence is crucial to determining lifetime earnings, and is itself determined by biological and environmental factors over which the agent has little or no control. It is difficult to see how an agent can deserve her higher / lower earnings given that

she had no control over (was not responsible for) the intelligence she was born with.

These sorts of claims conflict with ordinary notions of moral responsibility, desert, and justice, where it is neither unusual nor problematic to talk about an agent 'deserving' benefits and burdens such as income. In part, this may be because the picture is always more complicated than the example above allows: we may be taking account of the fact that, even with natural attributes that make high earnings possible, the agent must still exert herself in order to achieve those earnings and that such exertion entails desert. Thus, one problem with thought experiments such as this is the difficulty inherent in imagining what our intuitions would be *were it the case that* high income had no relation to factors within the agent's control.

At the deepest level, we reach the problem of free will and the question of compatibilism. I do not intend to deal with this issue here, but our ordinary notions suggest that there is room for free will in human action, which would entail that some form of compatibilism is true, or determinism false. If we assume that free will, and thus moral responsibility, is possible, the difficulty then becomes where we should make the cut. It still seems that there will be many morally relevant features of an agent and her actions which she cannot plausibly be said to control. There is a gap between our ordinary notions of fairness which suggest a principle of desert is appropriate, and our philosophical theories of justice which tend to exclude such a principle.

Despite the philosophical scepticism, I will consider the fairness of incentives in terms of desert here. In the main, this is because it is so central to our ordinary notions of fairness. Numerous criticisms of incentives, particularly in the lay media, stem from desert-based concerns, and it is necessary to make a reasonably full exploration of the basis for these criticisms.

Further, desert might not be so thoroughly absent from contemporary theories of justice: Nozick criticises Rawls's assumption that "the foundations underlying desert [must be] themselves deserved, *all the way down*" (Nozick [1974: 225]). Even Rawls allows that desert may play an important role in *retributive* justice, thus acknowledging the appeal of the idea that people should 'get what is deserved' in some cases (1971/1999: 314-315). Theories described as 'luck egalitarian' (see, for example Cohen [1989]; Dworkin [1981a, 1981b,

2000]), seek to compensate for unequal distributions resulting from 'brute bad luck', but not those resulting from 'option luck'. Luck egalitarianism may derive some of its appeal from its apparent consistency with notions of personal responsibility and desert.

Personal responsibility in healthcare is also discussed in the bioethics literature (Buyx [2008]; Schmidt [2009]; Thornton [2009]). This may not always be framed in terms of 'desert,' but occasionally the notion that some individuals are more deserving than others seems to motivate these arguments. Rhetoric surrounding desert and responsibility is also frequent in media articles debating the prudence of health incentives, often demonising the potential recipients.[26]

In considering this notion of desert I will also consider the related concepts of moral responsibility and blame. As I have argued throughout, incentives may be conceived of as a psychological aid to behaviour change, as well as a form of economic trade. The claim: 'individual A does not deserve the *healthcare intervention* X', may be quite different from the claim: 'individual A does not deserve the *payment* X'. The basis for asserting that some individual is not deserving of healthcare supplied by the state will look quite different from the basis precluding the state from trading with that individual.

Preliminary Remarks: Requirements for Healthy Living.

Some of the arguments that people do not deserve incentives assume that those receiving them act *wrongfully*. I do not consider this assumption in full here, but it is worth indicating briefly the sort of argument that might underlie it. I

---

[26] See, for example, Harcombe (2009); Chapple et al. (2004) for discussion of the stigma attached to diseases associated with obesity and smoking. Titles arising during the search for articles to include in Parke et al.'s (2011) media analysis included '*Pay nurses.. not these mollycoddled tubbies*' The Mirror, 2009; *'Lose weight and get £425 from the NHS; ...meanwhile patients are denied life-saving drugs'* Daily Mail, 2009; *'Dying patients are denied drugs but the obese can get £425 from the NHS to diet'* Daily Mail, 2009; *'Take up two bus seats and they'll pay your fare'* The Times, 2008; *'Finger on the pulse; should patients be paid for doing little more than taking their medication, asks Max Pemberton' The Daily Telegraph, 2009; 'Do addicts deserve freebies?'* Yorkshire Evening Post 2008; *'Why reward the least deserving?'* Aberdeen Evening Express, 2007; *'NICE wants incentives for addicts days after refusing cancer drug'* Aberdeen Evening Express, 2007. All these titles were retrieved from either online sources or the Lexus Nexus database and were provided via personal communication with Hannah Parke.

suspect justification may rest upon the idea that one has duties, either to oneself or to others, to take care of one's own health (for instance, to avoid becoming a burden on the healthcare system). This would mean that failure to adopt a reasonably healthy lifestyle would be grounds for criticism. There is a considerable literature on duties that follows a broadly Kantian line of reasoning (Kant [1785/1993]; Herman [1981]; O'Neill [2002]).

Alternatively, we might think that something like a principle of solidarity requires citizens to contribute to, and avoid taking too much from, the common resources, including healthcare. Solidarity compels agents to recognise that we are 'all in it together', and must take care of the interests of others as well as of ourselves. I discuss the notion of solidarity further in the next chapter.

Finally, some minimal degree of healthy living might be required in order to avoid harming others. An agent who lives unhealthily and suffers poor health as a result may directly or indirectly harm the interests of others. For example, she may subject others to the harms associated with passive smoking, additional burdens on the healthcare system which everyone must bear, or having to cover the work of an absent colleague. This 'harm' based requirement for healthy living is slightly different from the positive duties one has to be healthy, or those arising from solidaristic obligations.

Any factors demanding that agents adopt healthy lifestyles will be limited, and must be traded off against considerations of the well-being of the individual and other values. Some formulation of the basic libertarian principle of respect for the rights of the individual to act as she wishes, and not face unacceptable limits on her freedom to choose how to live her life, will also be important in explaining why any 'healthy lifestyle' requirement on citizens, if it exists, must be in a restricted form.

For the purposes of this discussion, I assume that it is possible to support some version of the claim that agents have at least a minimal duty to live reasonably healthy lives. This still leaves open the question of 'how healthy is enough?' That is, just how healthy must an agent's lifestyle be in order for her to be considered as having fulfilled her obligations. This will depend on factors such as those described above: how strong the imperative to be healthy is, and what other considerations place limits on it.

1. TRADE-TYPE INCENTIVES

The distinction between trade-type and aid-type incentives seems particularly relevant to the discussion of fairness and desert. In part, this is because we have different norms governing the use of money in buying / selling transactions around consumer goods and labour, and the use of monetary resources in healthcare interventions. Hence, I want to devote this first section to considering trade- and aid-type incentives in turn, and exploring the different kinds of objections that they attract. After introducing the objections to trade- and aid-type incentives in turn, I will consider all the objections together in more detail.

Where incentives are viewed as a trade - a straightforward 'payment' - there are a number of ways they might be viewed as objectionable. We can first distinguish between those arguments which claim that the trade, in principle, is fair, and those which claim that such a trade is never fair. Within the first group, we can pick out at at least two distinct arguments. First, it might be accepted that incentives can be deserved as a form of compensation, reward for effort, contribution to society, or similar. However, one might still argue that incentives, in fact, are unfair because they tend to provide people with more than they deserve. Second, one might similarly accept incentives in principle, but argue that it is not fair that they are only provided to a small sub-group of people.

Both these arguments make a claim about the relevant *desert base* when considering the fairness of incentives. It is generally accepted that, in order for an agent to deserve something, she must posses a desert base for that thing. For instance, Usain may deserve to win the olympic gold medal because he is the fastest sprinter in the world. What will count as relevant desert bases in different situations is a central, much debated aspect of desert theory.[27]

The claim that incentives are greater than recipients deserve is a claim that those receiving incentives do not posses the relevant desert base in the required amount to deserve the size of incentives they receive. The claim that all people who show the healthy behaviour should receive rewards is a claim that

[27] Lamont (1994: 49) argues that we cannot extract the correct desert bases merely by analysing the concept of desert itself. Rather, he makes the point that we must use external goals and values from the practice we are considering in identifying the relevant desert bases for that practice.

all those people also posses the relevant desert base and ought not to be excluded.

Of course, one might argue that those receiving incentives do not posses the relevant desert base at all: they are not deserving of any rewards, because adopting healthy behaviours is not a desert base for receiving incentives. This may be the strategy that many desert-related criticisms of incentives assume. It raises an interesting issue about the relevance of desert when considering trade-type incentives but can, I think, be set aside. For consider: it is clearly not the case that simply lacking a desert base for some benefit means that it is unfair that one receives that benefit. Take, for example, my finding £10 on the street: there is no desert base that makes it obvious that I deserve that £10. However, neither is it wrongful or unfair that I keep the money.[28]

Olsaretti (2003: 197) distinguishes between *positive* and *negative* desert claims. The former, stronger claims, assert that an agent actively deserves something. So, I may deserve a reward because I have the relevant desert base for that reward. The latter are weaker claims, and assert that an agent lacks the relevant desert base for something, and so does not (actively) deserve it: if I have done nothing to merit the reward, I do not deserve it. Similarly, I have done nothing to deserve the £10 I find on the street, so the negative desert claim that I do not deserve it will hold.

Where we are considering trades, negative desert does not seem sufficient to make the trade unacceptable. The mere lack of a desert base for an incentive is not sufficient to make it wrongful or unfair that an agent receives that incentive. Some may deny that the effort made by agents to adopt healthier lifestyles, or the sacrifices they must make in doing so, are relevant desert bases. However, this amounts only to a negative desert claim: recipients of incentives do not possess the relevant desert base to deserve rewards.

Those who wish to argue that the incentive trade is always, in principle, unfair, must make a positive desert claim, or show why the negative desert claim is sufficient. This positive claim is that those receiving incentives have a desert base for *not* receiving them: incentives should be denied to these people

---

[28] I appreciate the two cases here are not analogous, and there may be further reasons for thinking the incentives to be unfair. I will consider some of these later in the chapter.

because they have acted in a particular way (for instance, they have committed wrongful acts) that render them *actively* deserving the denial of incentives. The positive claim thus further implies that people *should not get incentives*, and that it would be *unfair* for them to receive incentives.

These implications do not follow from the negative claim, unless additional conditions are brought to bear. It seems likely that such conditions would be relational claims, such as that the state must not provide (negatively) undeserved rewards because it ought to conserve resources, or similar. However, in the case of incentives, the efficacy of this intervention (which may save the state money in the long term) would look likely to override the negative desert claim. This notion of relational desert, and negative desert claims will be more active in the the aid-type case than here, where free-market norms are fairly permissible with respect to the size of a payment.

I now want to focus on the potential for positive desert claims to object to the provision of trade-type incentives. I will discuss two strategies for this kind of claim. The first is that incentives reward *cynical* bad behaviour, and the second, that they reward *non-cynical* bad behaviour.

I have already outlined how we might conceive of unhealthy behaviour as wrongful action on the agent's part. For the purposes of this discussion, I assume there is at least a minimal requirement for agents to avoid unhealthy lifestyles, and that failure to do so qualifies as wrongful action.[29] The *non-cynical* bad behaviour argument proposes that the agent's wrongful action (in failing to adopt a healthy lifestyle) is sufficient grounds for denying her the opportunity of the incentive trade. In terms of desert, the claim is that unhealthy behaviour renders the agent actively undeserving of incentives.

The *cynical* bad behaviour case adds in an extra layer of wrongful behaviour. Here, the agent intentionally adopts an unhealthy lifestyle so as to become a candidate for incentives, and thus profit from receiving incentives to revert to the healthy behaviour. This is a form of 'gaming', where the agent cynically manipulates the practice of incentive-giving so as to financially benefit from it.

[29] This assumption is highly contentious and I do not mean to endorse it as true.

To summarise, the different desert-based arguments against trade-type incentives that I have outlined are:

Trade in principle acceptable
Amount is too great
Exclusion of other individuals is unfair
Trade in principle is unacceptable
Cynical bad behaviour
Non-cynical bad behaviour

2. Aid-Type Incentives

Before discussing these lines of criticism in more detail, I will first introduce those criticisms deriving from the conception of incentives as psychological aids. This alters the status of incentives from trade to healthcare intervention. The role of notions like desert and fairness (and justice more generally) tends to be different within the domains of trading and of healthcare provision.

Often, *entitlement*, rather than desert, is used to ground reasons for providing an agent with healthcare. Entitlement requires that, for justice to be served, individuals' rights be respected. Entitlements are formalised through the structure of institutions which provide for agents' entitlements, and obviously depend upon what rights agents are thought to have, and what corresponding duties and obligations are owed to them.

Desert often tracks entitlement, but the two are distinct and may sometimes conflict. For instance, had Usain tripped in the last metres of the 100m sprint, and Yohan passed him and won the gold medal, we might say that, although Yohan is *entitled* to his medal because he won the race, it is really Usain who *deserves* to be the champion because he is the faster sprinter.

In healthcare it is normally necessary and sufficient that individuals are entitled to certain sorts of treatment: we need not consider further whether individuals deserve treatment in order to think that they should be treated. The critic of aid-type incentives, however, argues that this form of intervention is not deserved, and that because of this, incentives should not be included in the basic healthcare provision citizens are normally entitled to.

Within a socialised welfare system, such as the NHS, the extent of healthcare provided will normally be determined by things like cost and likely

improvements in well-being. Facts about the agent are normally relevant only insofar as the agent must be capable of benefitting from the treatment, and that she be a citizen (or holder of health insurance) which allows her access to healthcare in general.

Desert requires the additional assessment of responsibility in order to be incorporated into decisions about what healthcare people should receive. If people are to be considered blameworthy for their illness they must also be responsible for the causes of that illness. The practical work done by desert here is to suggest that agents who are responsible for their illness should not receive treatment for it, but rather deserve to suffer the consequences of their behaviour.

It is not clear why critics of incentives often prefer to adopt the language of desert, rather than of entitlement, here. The parallel claim would be to say that those agents who are responsible for their illness are no longer entitled to treatment for it (or something along those lines). It is perhaps because desert also carries with it the assumption of moral blame. To claim that agents are not entitled to incentives could be a claim that those agents do not qualify as citizens; that incentives are not effective; that incentives are too expensive to be provided; and so on. These can be claims about the intervention, the social institution, or about the agent herself. Desert, however, makes a claim specifically about the status of the agent: there is some salient moral feature of the agent which means she does not deserve this intervention. Thus, the absence of desert is a form of condemnation upon some aspect of the agent's character or actions.

Once again, it seems the critic must make a positive desert statement in order to claim that agents should not receive incentives. Given that there is a presumptive entitlement to treatment (assuming the other necessary features for entitlement are in place: incentives are cost-effective, recipients are citizens, and so on), an active desert claim that agents do not deserve incentives because of some desert base they posses must be made.

As with trade-type incentives, such a positive desert claim will rest on the assertion that agents act wrongfully in failing to adopt healthy lifestyles. If it can be shown, along the lines I gestured at earlier, that agents who adopt unhealthy lifestyles fail in their duties to themselves or society, or harm others, or do not

adhere to a principle of solidarity, then these may count as relevant bases for denying such agents health incentives.

This strategy may be modified somewhat by conceiving of desert as a comparative feature. This takes central account of the fact that healthcare provision involves divvying up resources between different interventions, illnesses, groups, and individuals. Resources not spent on one intervention can be spent on another instead. The comparative desert claim is that those who are responsible for their illness are *less* deserving of treatment than those who are not responsible for their illness. Further, healthcare provision should be prioritised in favour of those who are more deserving.

The sharpest example of using a comparative desert-based criterion for allocating healthcare resources is in the case of organ transplants, as illustrated by Thornton (2009). There are often insufficient organs to treat everybody who is in need of them. Where all other relevant features of two candidate recipients for a transplant are the same, Thornton argues that moral responsibility should be used to break the tie. Different justifications can be given for using moral responsibility in this way, but one possible motivator is the connection between moral responsibility and desert.[30] In this case, it need not be preferable that one who is responsible for her illness suffers for it, but it must be preferable that she suffer from her illness, rather than that some other agent (who is not responsible for his illness) suffers.

Although incentives do not constitute a clear 'tie-break' situation, they do involve diverting resources in one direction, and benefitting a particular set of people, rather than directing those resources differently to benefit a different set. Desert (as deriving from responsibility) could be used as a criterion to factor into decisions about where those resources are directed.

This argument assumes that those targeted by incentives tend to be responsible for the diseases they suffer. Incentives are generally offered to individuals whose lifestyles contribute to their poor health prospects, either because they eat poor diets, take little exercise, do not vaccinate against disease or attend disease screening programs, and so on. It is also assumed that there are plenty of people requiring healthcare who are not responsible (or

[30] It should be noted that Thornton does not use desert as the motivator in this case.

significantly less responsible) for the diseases they suffer. Those who receive incentives may thus be considered less deserving of treatment than those who receive other forms of intervention.

The implications of this argument may go further than the critic intends: we must remember not only to treat incentives as a healthcare intervention, but also to treat other healthcare interventions *as we treat incentives*. The implication here is that if it is unfair to provide incentives to those with unhealthy lifestyles (in order to reduce chronic disease) because they are deemed undeserving, it may also be unfair to provide some other forms of (costly) intervention, such as support groups, drug therapies, and so on, to those same individuals. The consequences of introducing a principle of desert into healthcare provision decisions will affect more interventions than just incentives, and more conditions in need of treatment than merely those resulting from the kind of lifestyle behaviours mentioned. For instance, skiing and getting pregnant are activities that one generally engages in voluntarily and which increase one's risk of needing medical assistance.

To summarise, I have described how the desert-based argument regarding aid-type incentives can be presented in a non-comparative and a comparative form:

Non-comparative case

Individuals should, in general, bear the consequences of actions they are morally responsible for

Comparative case

Individuals who are responsible for their illness are less deserving of treatment than those who are not responsible for their illness

Analysis

I will now consider in more depth the various arguments outlined above that suggest the use of incentives is unfair. I have highlighted different arguments in relation to trade-type incentives and aid-type incentives. I believe not only are the arguments presented in support of the desert-based account different in these two cases, but the threshold for the arguments to succeed (for them to make us think that incentives are inappropriate) is different.

In relation to this second, threshold claim, T. M. Scanlon's (1986; 1998; 2012) work is instructive.[31] Scanlon insists that, when thinking about moral responsibility and blame, it is important to keep one eye on the consequences of such evaluations. For example, if the evaluation of an agent as blameworthy means that we are justified in punishing that individual, the requirements for establishing blame must be set quite high. We may alternatively wish to identify blame in situations where no justification of punishment follows, and in this case the requirements for blameworthiness may be more easily met.

It need not follow that finding incentives to be incompatible with notions of desert, and thus unjust, means that, all things considered, they shouldn't be used. However, desert-based criticisms might constitute serious reasons for thinking incentives should not be used (and the most forthright proponents of these criticisms will want to claim that incentives should not be used).

To deny trade-type incentives is different to denying aid-type incentives. The former is to stop one individual (or rather the state or similar institutional provider of incentives) from freely trading with another individual. In this case, we prevent these two agents from benefitting from the advantages of the trade. The latter involves denying individuals access to healthcare that could help them, and that they would otherwise be entitled to.

I propose that the bar be set higher for denying access to aid-type incentives than to trade-type incentives. That is, there must be excellent reasons for supposing that it is unfair to provide aid-type incentives to people in order for this to constitute a good reason to prohibit the use of incentives. We may not need to set the bar so high for trade-type incentives, as the consequences of denying access to these incentives are not so serious (though I do not wish to imply that such a denial of access should be considered trivial).

In the discussion that follows, I first consider three of the arguments made against trade-type incentives: that the amount represented by incentives is too great; that incentives ought to be offered to everybody; and that those receiving incentives display cynical bad behaviour. I do not offer a full defence of

[31] It should be noted that Scanlon falls into the group of theorists who deny that desert should be included in a theory of distributive justice.

incentives in the light of these arguments, but suggest that each has only limited strength.

Next, I consider the arguments that base their claim on the idea that agents receiving incentives act wrongfully in some way. These come under the non-cynical bad behaviour argument; the argument that individuals ought, in general, to bear the consequences of behaviour they are responsible for; and that individuals who are responsible for their illness are less deserving of treatment than those who are not responsible for their illness. In considering these arguments, I consider the key notion of responsibility, and discuss the extent to which agents can be considered morally responsible for their unhealthy lifestyles. In doing so, I present some evidence from research into the social determinants of health, and the mechanisms of health behaviour.

Economic Desert

Trade-type incentives are payments to individuals in return for something the incentiviser wants. Such payments may be thought 'deserved' along the same lines as one might deserve a wage for performing one's job, or deserve compensation for participating in medical research. Feinberg (1970) insists earned income can only be deserved where it is compensation for hazardous or costly work. Sadurski (1985) in contrast proposes that effort should be the main basis for deserved income. Other desert bases could include contribution to society, exposure to risk, need, current wealth, opportunity costs, and so forth.

It will often be very tricky to determine the exact payment - not too great nor too little - that any individual deserves. Attempting this involves assessing the 'absolute' desert of an agent. Different metrics could be used to tell us how much each desert base should contribute to the economic deserts of an individual. It seems likely that some plausible metrics we could use would tell us that many of the payments made to real-world citizens are not fully deserved.

Take some real-life example schemes (see also Appendix A):

- 'Give it up for Baby' offers £12.50 in supermarket vouchers per week to women who smoke no cigarettes that week;
- FIAT (Financial Incentives for Adherence Trial) offers £15 cash per anti-psychotic depot medication the agent receives;

- A trial incentivising HPV (Human Papilloma Virus) vaccination offers £20 for the first injection, £5 for the second, and £20 for the third and final injection.

I do not know how much effort, sacrifice, risk exposure, or productive contribution was required for any agent to meet the demands of one of these schemes. It is not clear that knowing this for any individual, or knowing the average for all participants, would help us to judge absolute desert, since there is no clear guidance as to how much value we should place on effort, sacrifice, risks, and so on.

Alternatively, we may consider the *comparative*, rather than *absolute*, aspects of desert. In this case, we can only determine what one individual deserves by taking into account what other individuals will receive. In order to determine comparative desert, we may compare the payments made in incentive schemes to payments made in comparable situations, and make some approximate judgement of their appropriateness on that basis.[32]

However, it is not clear what should act as a precedent for this judgement. Should we use the effort : pay ratio involved in working as a teacher and apply it to the case of incentives for healthy behaviour? Or should we match incentive payments to the compensation per unit of sacrifice involved in medical trials? The comparative approach does not seem to move us on much further.

All that I propose to say here, I fear rather unhelpfully, is that it is extremely hard to judge one way or the other whether the size of incentives matches the deserts of recipients. We might make intuitive responses about £10 weekly payments, or £650 cumulative payments, but such assertions would not seem to be robustly justified. It might be possible to provide some sort of judgement on the matter, through sophisticated sociological analysis, combined with a thoroughgoing philosophical apparatus instructing us as to what the appropriate desert bases are and how they should be translated into payment. Such ambitions are far beyond this thesis.

I put to one side, then, the claim that incentives are larger than recipients deserve. I have not ruled out this argument, but it has also not shown itself to be particularly compelling. It seems that, whilst incentives remain relatively

---

[32] For more on the distinctions between absolute, proportional, and comparative desert, see Scheffler (2003: 83-84)

modest, the gap between what people receive and what they deserve is unlikely to be too great.

Unfair Exclusion

The related claim is that *everybody* who shows the incentivised (healthy) behaviour should receive payments, not just those who newly adopt the healthy behaviour. This assumes that the appropriate desert base is the healthy behaviour itself, thus, anybody showing the healthy behaviour is deserving of incentives.

An obvious reason for providing incentives only to those who currently exhibit unhealthy behaviours is cost: it will be much cheaper to provide incentives only to those whose behaviour it will be advantageous to alter. Targeting incentives in this way pays for the *change* in behaviour: smoking to not smoking; unhealthy diet to healthy diet; failing to get vaccinated to getting vaccinated; non-adherence to adherence to treatment. Recipients of incentives only receive them for the duration of the scheme, in order to bring about the transition from unhealthy to healthy behaviour.[33]

Identifying the appropriate desert base can be difficult. Different approaches to economic desert have been mentioned: economic payments may be deserved as compensation; reward for effort; for productive contribution to society, and so on. All those who exhibit healthy lifestyles contribute to society by doing so, by reducing the likelihood of illness and the negative effects this will have on society. This contribution alone may constitute a sufficient desert base for incentives. Further, it may be a struggle for these individuals to maintain their healthy lifestyles, and they may have had to work very hard and make significant sacrifices to achieve them. Considering these features as well, we may have more reasons for thinking that lots of individuals who have already adopted healthy lifestyles are no less deserving of payment than those who do so in exchange for incentives.

There are some plausible desert bases which will tend to recommend incentivising (according to desert) those who currently show unhealthy behaviours in preference to those who already show healthy behaviours. For

---

[33] Incentive schemes are not intended to continue after behaviour change has been established. However, empirical evidence is lacking and it is not clear how long incentives may need to continue in order to ensure that behaviour change is sustained.

instance, features like need, social deprivation, and health prospects could play a role in determining someone's desert. These features will arise more frequently in individuals who currently have unhealthy lifestyles, and who are also more likely to be subject to greater deprivation (discussed later).

Nonetheless, strict adherence to desert may require that at least some who already show healthy behaviours are deserving of incentives as well. It is unlikely that it will be practical to adopt incentives if they must be given to people who already have a healthy lifestyle. Thus, excluding such individuals from eligibility to incentives would be a necessary structural feature of practices of providing incentives.

There is a danger here of over-stating the imperative to adhere strictly to what is required by desert. Most of the time, monetary transactions occur free from scrutiny as to their particular desert bases. If I buy a bicycle from Chris rather than from Bradley, that is not made wrongful if we come to understand that Bradley has put in many hours of hard work to repair his bicycle, and really needs the money, whereas Chris is quite wealthy and is simply selling a bicycle he was given as a gift but no longer wants. There may be facts about Bradley and Chris that make us think that Bradley deserves the sale more. However, we accept that, where a market operates with relative freedom, transactions will often be based on criteria other than strict desert.

Giving incentives to people who already live healthily is paying for nothing (or perhaps, back-paying them for something they have already done). Rather than being obligated to do this by the requirements of desert, this seems like a case of excessive generosity, and is not demanded by principles of fairness.

Once again, there is some basis to the criticism that providing trade-type health incentives only to those whose behaviour we wish to alter fails the demands of desert. However, I think the power of this criticism is limited. In the domain of transactions, desert may have some relevance, but it seems that other factors (ownership rights, fulfillment of contractual obligations, absence of exploitation, and so on) will constitute the main criteria for determining the fairness of a trade. In the case of trade-type incentives, then, concerns about desert will be less worrisome. As with the argument relating to the appropriate

size of incentives, whilst incentives remain relatively modest, desert concerns are unlikely to constitute a significant objection.

'Bad Behaviour'

The second set of arguments that I described relating to trade-type incentives were those which posit a positive desert claim that individuals receiving incentives actively do not deserve them. I proposed that the 'do not deserve' here comes from the claim that those offered incentives have acted wrongfully in failing to adopt healthy behaviours.

The criticisms of aid-type incentives rest on similar claims: that those who receive incentives act wrongfully and thus do not deserve the treatment, or at least, are less deserving of treatment than agents who live healthily (and who therefore do not act wrongfully on this front).

In what follows, I will consider the link between agent behaviour and desert. I argue that, even if we accept the claim that agents ought to adopt healthy lifestyles, their failure to do so does not mean that they are morally blameworthy, nor that they are undeserving of incentives. I propose that sociological evidence concerning the social determinants of health, and psychological research into the mechanisms of health behaviour, undermine the kind of responsibility link between an agent and her actions that will be required to render her undeserving.[34]

Gaming

First, I will briefly consider the *cynical* bad behaviour claim, outlined earlier. This was the argument leveled at trade-type incentives which proposed that those being offered incentives might actually adopt unhealthy behaviours with the very intention of making some money from the incentives they would subsequently be offered to quit those behaviours. Effectively, the criticism is that incentives are vulnerable to 'gaming,' and that those who are willing to exploit the system (through wrongful behaviour) will reap the rewards.

---

[34] Once again, 'underserving' here refers to both undeserving of trade-type incentives (payments), and undeserving of aid-type incentives (healthcare). To be rendered undeserving of these two things is quite different. However, I do not think the discussion shows that 'bad behaviour' renders the agent undeserving of either type of incentive.

This criticism can probably only be resolved through empirical evidence that can tell us, one way or the other, whether people really do game the system in this way. There are other ways that people may 'cheat' in order to appear as if they are adopting healthier behaviours, when really they are not. Perhaps one could fake being a smoker, or intentionally become non-adherent with one's treatment regimen in order to be paid to quit or to become adherent. It may also be possible to 'fake' the tests of behaviour change, such as the carbon monoxide breath test used to assess whether someone has been smoking or not.

Any individuals who engage in these sorts of cynical bad behaviours are likely to be deemed undeserving. However, it is not clear how frequently people are likely to behave like this in order to get at incentives. Consider the effort required for someone to intentionally gain weight, then gradually lose it so as to earn rewards from an incentive program. In some cases, the cost of the incentive is unlikely to cover the direct financial cost of the 'bad' behaviour (for instance, incentives to quit smoking are often less than the cost of a packet of cigarettes).

Without research and evidence about the prevalence of gaming behaviours it is hard to say how powerful this argument is. Given the effort required and minimal rewards involved in these forms of gaming I would hazard a guess that rates of successful gaming would be minimal, but such speculation is no substitute for empirical evidence on the matter which is, thus far, lacking.

Further, as with the other ethical considerations of health incentives, the need to avoid rewarding undeserving 'gamers' should be balanced against the potential good incentives can bring about. If only a very small proportion of those receiving incentives do so through cynical methods, this may be an acceptable trade if other (genuine) recipients are helped.

BLAMEWORTHINESS

There are three arguments I have not yet considered. First, the argument that agents do not deserve incentives because they display non-cynical bad behaviour. This should be distinguished from the cynical bad behaviour claim. In the cynical case, the objection is to the (potential for) gaming behaviour. In the non-cynical case, it is proposed that the behaviour of failing to live healthily is sufficient to render the recipient undeserving of trade-type health incentives.

The other two arguments relate to aid-type incentives and have similar motivations. One of these asserts that recipients are *absolutely* undeserving of incentives (this is a judgement about the recipient in isolation, independent of facts about other people). The other asserts that recipients are *comparatively* undeserving (a claim that, in comparison to what others receive, incentive recipients get more than they deserve).

Though these arguments have some different features, there is significant overlap. In particular, they all require that the agent in question be considered *blameworthy* for her wrongful behaviour, so as to show that she is undeserving of trade-type or aid-type incentives. In this section, I will consider this accusation of blameworthiness, and the implication that some who are unhealthy might be undeserving (or less deserving) of healthcare.

Reactive Attitudes

One influential approach to analysing notions such as blameworthiness is based upon the 'reactive attitudes.' In his seminal article 'Freedom and Resentment,' Peter Strawson (1963) proposes that our practices of holding people to be blameworthy and praiseworthy are expressions of 'reactive attitudes.' Reactive attitudes are how we respond to the attitudes held by other agents, usually as revealed by their actions. For instance, if you spill hot coffee on me, and make no effort to apologise, I might come to resent you for your lack of remorse. Resentment such as this is a reactive attitude. Other reactive attitudes include indignation (where I criticise you for wronging another agent), and guilt (where I feel I have wronged someone). Holding reactive attitudes such as these and expressing them through the practices of blaming and praising are, according to Strawson, central to interpersonal relationships.

That attitudes such as resentment and indignation lie behind desert claims in the media is apparent. The language used in criticising incentives, and their recipients, is highly evocative, and often demonizing of those with unhealthy lifestyles.[35] Yet it is clearly not the case that every instance where an agent experiences resentment will be a case where the subject of that

---

[35] A few sample headlines: *'Pay nurses, not these mollycoddled tubbies'* The Mirror, 2009; *'Will scheme turn fatties into fitties?; Should overweight parents be paid to walk their kids to school'* Evening Times, [Glasgow], 2008; *'Why reward the least deserving?'* Aberdeen Evening Express, 2007.

resentment is rightly to be considered blameworthy. First, the attitude assigned to the agent, based on her behaviour, may not be the right one. Second, the reactive attitude we hold in response to the agent's behaviour may not be correct. Third, the agent may not be the appropriate subject of reactive attitudes (Scanlon [1986: 161]).

If any of these cases holds, it will not be appropriate to express the attitude of resentment towards the agent, nor to blame them. In the discussion that follows, I consider the question of how agents come to adopt unhealthy behaviours, and whether or not we should consider them morally responsible for their behaviours.

Some theories explicitly include responsibility in their criteria for desert: agents must be responsible for the desert bases which convey praise or blame upon them. Those positing that agents who are responsible for their illness are not deserving of treatment place responsibility central to their criticism. The desert base here is the unhealthy behaviour, which conveys blame upon the agent responsible for it.

On the reactive attitudes account, the responsibility status of the agent will be relevant to judging the appropriateness of resentment (and thus blame). The three factors that can render reactive attitudes inappropriate - attitude assigned to the agent is wrong, reactive attitude is wrong, agent is not appropriate subject of reactive attitudes - could be affected if the agent is not responsible for the relevant behaviour. If the agent was not responsible for the action to which the offending attitude has been assigned, the basis for thinking that the agent holds that attitude will be undermined. If the agent does hold the attitude for which we resent her, yet lacks responsibility for holding this attitude then it might yet be inappropriate to resent her. Finally, if the agent is incapable of controlling any of her actions in a way compatible with responsibility, we may think that she will not be an appropriate subject for resentment.

Responsibility, thus, has relevance for desert-based approaches to criticising the provision of health incentives. It seems important, therefore, to consider to what extent agents are responsible for their unhealthy lifestyles. In doing this I shall first look at some of the evidence from research into the social determinants of health, which explores the links between social, environmental, and economic factors and health outcomes. I will then summarise some of the

literature from psychological research into the mechanisms underlying health behaviour, and the barriers to behaviour change.

In considering the conclusions about health incentives that we can draw from this evidence I will consider how the freedom an agent has in choosing how to act relates to her responsibility for her actions. I draw upon Pettit's work on freedom as 'fitness to be held responsible' for this discussion. I then seek to connect this discussion of freedom and responsibility to the discussion of those factors influencing health and health behaviour, and thus the responsibility of agents for their unhealthy lifestyles. Finally, I seek to connect this back to the discussion of blame and desert, and propose that where agents are not fully responsible for their unhealthy behaviour, they will not be blameworthy, and will not deserve the denial of incentives.

Socio-Economic Factors and Health

Research has identified clear links between health and deprivation. Deprivation is most commonly given in terms of the 'Index of Multiple Deprivation' (IMD). The domains along which IMD scores are calculated are income, employment, health, education, crime, access to services, and living environment (Department for Communities and Local Government [2011]). IMD scores apply to areas, not people, though they are calculated by assessing the circumstances of those living within that area. Not all of those living in a deprived area will be deprived, and not all deprived people will live in deprived areas.

Socio-economic factors are correlated with health: levels of health and life expectancy are lower in more deprived areas than in less deprived areas. There is a gradient such that, for every move an individual makes up the socio-economic 'ladder,' the better her health prospects will be. A wide range of factors, which stem from inequalities in power, money, and resources, contribute to this distribution of health outcomes reflecting differences in deprivation. These factors include:

> [M]aterial circumstances, the social environment, psychosocial factors, behaviours, and biological factors. In turn, these factors are influenced by social position, itself shaped by education, occupation, income, gender, ethnicity and race. All these influences are affected by the socio-political and cultural and social context in which they sit.

'Unhealthy behaviour' is effectively a broad term to describe behaviour which exposes the agent to risk factors for various diseases. For instance, tobacco smoke is a risk factor for lung disease, so cigarette smoking counts as an 'unhealthy behaviour.' Many of the sorts of unhealthy behaviour that incentives target are incorporated into everyday behaviours and general lifestyles, including diet, physical activity, and smoking. Risk factors in these domains tend to contribute to the likelihood of suffering from non-communicable diseases (NCDs). NCDs include diabetes, cardiovascular disease, cancer, chronic respiratory disease, and mental disorders. Unhealthy behaviour, in the context of those behaviours targeted by health incentives, can also include failure to adhere to treatment advised by physicians, and failure to utilise healthcare resources (cancer screening, STI testing, vaccination uptake, and so on).

Deprivation, exposure to risk factors, and health outcomes, interact in a complex fashion. I cannot discuss these intricacies in too much detail here, however, a few examples can gesture towards the wider trends linking socio-economic and environmental factors to health. Take weight and smoking for example: women in managerial and professional groups have a 19% rate of overweight and obesity, compared to 29% in routine and semi-routine groups (Nuffield Council on Bioethics [2007]); 17% of men in professional occupations smoke compared to 31% of those in manual occupations (Marteau [2010]).

Utilisation of healthcare services also tends to be lower in deprived areas. Human Papilloma Virus (HPV) vaccination uptake has been shown to be significantly lower in deprived areas and amongst ethnic minorities (Roberts et al. [2011]). Further, girls who did not receive the HPV vaccine were less likely to have received childhood immunizations such as those for measles, mumps, and rubella (MMR).[36] Vaccination against influenza was also shown to have slightly lower rates of uptake in more deprived areas (Coupland et al. [2007]). Cancer screening tends to be less well utilised by the socially deprived, as shown by McCaffery et al. (2002) in the case of colorectal cancer screening, and Maheswaran et al. (2006) in the case of breast cancer screening.

[36] The HPV vaccine is given to girls aged 12-13 as part of the national vaccination programme. A catch-up programme also provides the vaccine to some older girls.

Such differences in exposure to lifestyle risk factors and use of healthcare services tend to manifest themselves in disease prevalence and life expectancy: so, in one of the wealthiest areas of London male life expectancy is 88 years, whilst in one of the poorest it is 71 years. Throughout England, those living in the poorest neighbourhoods can expect to die seven years earlier than those in the wealthiest; this gap remains at six years even when the poorest and richest five percent are excluded (Marmot et al [2010: 37]).

These data establish a correlation between various indicators of deprivation, exposure to health risk factors (including lifestyle and uptake of health services), and health outcomes. Whilst the patterns are well established in the social determinants of health literature, they are only general, population-level trends, and thus will not apply to every individual living in a particular post code, or doing a particular job.

Furthermore, although compelling, the evidence presented here cannot establish causation between deprivation and health (i.e. that those factors marking out deprivation, either directly or indirectly, *cause* unhealthy behaviour and poorer health outcomes). It is possible that there is a common cause for these factors which would explain why poor health is often seen in areas of high deprivation. Indeed, the explanation of these trends still leaves room for free choice and autonomous agential control over health behaviour. Thus, the evidence from research into the social determinants of health might lead us to be suspicious of assumptions that people are in control of (and morally responsible for) their health behaviour, but it ought not to persuade us that this is the case.

The additional information needed in order to make judgements about responsibility relates to *why* the agent acts as she does; what control she has over her behaviour; and could she act otherwise. Psychological research into the mechanisms of health behaviour and behaviour change can help illuminate these kind of issues. Before presenting some evidence from that field of research, I will give a brief sketch of Phillip Pettit's account of freedom, which conceptualises freedom as 'fitness to be held responsible' (Pettit [2001]). Pettit's account is helpful in directing our attention towards those features of free human action that will be necessary if we are to judge an agent as responsible for her behaviour.

PETTIT: FREEDOM AS FITNESS TO BE HELD RESPONSIBLE

On Pettit's account, to be free and to be responsible are separate but overlapping states. He identifies three aspects of the free agent:

> Pettit's conditions of freedom:
>
> 1. *Free Action:* "the freedom of an action performed by an agent on this or that occasion;"
> 2. *Self Identification:* "the freedom of the self implicit in the agent's ability to identify with the things thereby done, rather than having to look on them as a bystander;"
> 3. *Social Status:* "the freedom of the person involved in enjoying a social status that makes the action truly theirs, not an action produced under pressure from others." (Pettit [2001: 4])

These conditions must be met for the agent to be considered free. Pettit continues: "To be free, in the most general sense, is to be fully fit to be held responsible; it is to be fully deserving of the sort of reactions, say those involving resentment or gratitude, that characterize face-to-face relations." So freedom relates to desert and responsibility, and the reactive attitudes, in this consistent way. Agents who are rendered unfree will not be fully fit to be held responsible, may not be the appropriate subjects of the reactive attitudes with which other agents tend to respond to them, and may not deserve praise or blame for their actions. It will also be possible for an agent's freedom to be partially undermined, where some of the conditions for freedom are not met, or are only partly met. This would still mean the agent is not *fully* fit to be held responsible, though she may bear some responsibility for her actions (and some desert as a consequence).

The first condition of freedom that Pettit describes is that of *free action*. Actions will be free so long as they are under the control of the agent. The agent must not be coerced through extreme pressure (such as death threats) or physical force.

The second condition is that of *self identification*. In this case, what is needed for freedom is that "[t]he agent cannot be detached from the action, or

from the process leading to the action, in the way they may be detached from a reflex or a pathology or even an obsession or compulsion" (Pettit [2001: 10]). The agent must be able to take ownership of the action she performs; she must be able to talk about it in the first-person, and fully accept that it is *she* who performed it. Agents who act under the influence of drugs, hypnosis, or whilst asleep are unlikely to meet the criteria for this aspect of freedom.

The final condition of free action relates to *social status* and the absence of significant external pressure. Examples of people rendered unfree in this way include those who are pressured into prostitution or crime by the desperate circumstances they find themselves in.

These different aspects of freedom will be relevant to the discussion of desert that follows because of their connection with responsibility. I will now introduce some of the psychological literature that has been directed at understanding the connection between social deprivation and poorer health outcomes, by elucidating the mechanisms of health behaviour.

THE PSYCHOLOGY OF HEALTH BEHAVIOUR

Intuitively, it may seem that it is within one's power to avoid many of the health risk-factors mentioned. One cannot control who one's parents are, whether they smoke, where one goes to school, or the foodstuffs on discount in the supermarket. Ultimately, however, each agent can still choose how she lives. After all, there are plenty of individuals who are subject to disadvantage who nevertheless manage to lead healthy lifestyles without the need for interventions from the state.

Yet, human reasoning has been shown to incorporate the 'fundamental attribution error' which produces a "tendency for attributers to underestimate the impact of situational factors and to overestimate the role of dispositional factors in controlling behavior." (Ross [1977: 183]) Thus, although it is intuitively appealing to assume that we are 'free' and 'responsible' for the actions we take in our everyday lives, we should be wary of assuming this is generally the case.

An extensive literature seeks to explain the psychological mechanisms underpinning the social patterning of health. Some of this evidence leads us to question assumptions about people's ability to 'choose' to be healthy. One of the key insights of this research is that much human nature is 'automatic.' The 'two

systems' psychological model, as discussed in Chapter Four, proposes that, while some behaviour is influenced by values and conscious reasoning, another pathway of behavioural control operates with little or no cognitive engagement (Strack and Deutsch [2004]). Where an agent is not normally required to consciously control their behaviour, it takes a large and sustained psychological effort to then intervene to alter these everyday behaviours and habits.

This habitual nature of behaviour makes it highly susceptible to environmental influences: factors like a high density of fast-food outlets, and being surrounded by others who are engaging in unhealthy behaviours. These features can have a considerable negative impact on lifestyle and are typically found in deprived areas (Marteau [2010]; Cummins et al. [2005]; Macdonald et al. [2007]).

Agents who live in deprived areas also tend to be exposed to greater levels of stress. The physiological effects on the brain of stressful living environments can cause agents to act more impulsively. The sorts of health risk factors that contribute to chronic, lifestyle-related disease often satisfy these kind of impulsive desires. The capacity for resisting such impulsive behaviours - self-regulation - is also influenced by features of the social environment. In particular, immediate family and parenting style can help foster the development of robust self-regulatory skills. However, the kind of environment needed to develop such skills is undermined by poverty, deprivation, and low levels of parental education (Marteau [2010]; Lexmond and Reeves [2009]).

The greater tendency for environmental circumstances to cue unhealthy behaviour means that those living in deprived areas must exert greater effort in diverting their 'automatic' behavioural system away from unhealthy behaviours. Moreover, such self-regulation will often be harder for agents who are subject to numerous indicators of deprivation because they are less likely to possess the skills required to curb impulsive behaviour. Kotz and West (2009) present evidence that seems to support this, showing that smokers from the most deprived socioeconomic groups are equally likely to attempt to quit smoking (and to seek help in doing so), but only half as likely to succeed as those in the highest socioeconomic groups.

I have sought to present some of the evidence from research into the social determinants of health, and the mechanisms of health behaviour, in order to help consider questions of whether those who adopt unhealthy behaviours bear responsibility for doing so; are blameworthy for some of the harmful consequences that result; are deserving of the ill health they suffer; and are undeserving of health incentives.

Whilst the sociological research identifies general trends that link aspects of deprivation with poor health outcomes, the psychological research goes further, suggesting a complex causal relationship connects these factors. The issue of relevance for questions about the desert of agents with unhealthy lifestyles is whether this relationship undermines the moral responsibility of those agents for their behaviour.

Let us return to Pettit's analysis of freedom as fitness to be held responsible. Recall, the three aspects of freedom that Pettit highlights relate to *free action*, *self-identification*, and *social status*. I believe that our best information about the mechanisms underlying health behaviour leaves open the possibility that agents are not fully free in adopting unhealthy lifestyles, and that they are thus not fully fit to be held responsible.

First, consider the requirements for *free action*. The agent who eats an unhealthy diet is not rendered unfree by the employees of McDonalds holding a gun to her head and forcing her to consume endless boxes of chicken nuggets. Freedom-limiting influences in the health behaviour context are less dramatic and, invariably, far weaker than this. However, as I have argued in the context of what should constitute coercion, the important factor is the psychological pressure placed on the agent's will. I think the same holds true here: we are interested in whether the action is an expression of the agent's own will, or of factors overriding her will in some way.

Some unhealthy behaviours incorporate pathologies of addiction, dependence, or compulsion. In these cases, the pressure to act in a particular way (even though that pressure comes from 'within' the agent) can be extreme. In other instances, an agent may wish to act healthily, yet fail to muster the willpower to do so (sometimes described as *akrasia*). The pressure exerted on the agent's will in the first set of cases is more extensive, and will compromise

the freedom of any given action to a greater extent than in the second case. This needs to be the case in order to distinguish cases of addiction and compulsion from cases of weakness of the will.

It is, of course, not possible to enter somebody's mind and experience what she experiences, or to objectively measure the degree of pressure her will comes under to behave in a particular way. We cannot make an objective assessment of the effort required for agents to overcome such pressure, to shun unhealthy behaviours and adopt new, healthier behaviours. The agent herself is the only one with access to this information, which makes for a rather dissatisfying conclusion for those wishing to assess the freedom of actions (and the responsibility agents should bear for their actions). Such an emphasis on psychology also leaves open the possibility that those who are particularly susceptible to pressure on their will, and who are likely to succumb easily to temptation (for instance, the dieter who gives in at the merest whiff of a freshly baked croissant), will be rendered 'unfree' to some extent.

I would stand by this assessment: what is significant is the agent's will, the pressure she comes under, and her capacity to resist that pressure.[37] Further, the strategy proposed on Pettit's account provides three different forms of freedom. So far, I have only made mention of the first: *free action*. An agent whose freedom of action is compromised slightly on a given occasion may nevertheless be judged largely 'free' in a more general sense where she suffers no freedom limitation along the other two aspects of freedom. For an agent to be *fully* fit to be held responsible, she should also be *fully* free. However, an agent may still be considered largely responsible where she meets most of the conditions for freedom.

The second aspect of freedom is that of *self-identification*. Agents are free in this sense when they can take full 'ownership' of their actions: see them as fully 'theirs' and accept them as an extension of a part of their own character. This requires more than the mere fact that the agent physically performed an action. Consider the difference between accidentally tipping a drink down somebody you are having an engaging conversation with, and throwing your drink over

[37] I do not suggest that this is *all* that matters in making a moral appraisal of the agent. If she has cultivated a sensitivity to external pressures then we might feel less inclined to consider her 'unfree' in a fuller sense. The picture will be more complicated than I express it here.

somebody who has just made a series of insulting remarks to you. In the first instance you are unlikely to see that action as 'yours' in any significant way.[38] In the second case, you may identify with the action to a greater extent as it was intended and an accurate expression of your feelings towards the individual.[39]

When thinking about an agent's self-perception, we may take her long term goals to be important: how she sees herself and how she wants to be in the future (the discussion of autonomy in Chapter Four is relevant here). Rational, reasonably well-informed agents are capable of forming plans about how to achieve their short, medium, and long term desires, and about what it will be feasible for them to achieve. Agents who act in the short term in ways that will obviously disrupt their long term plans and harm their interests may not strongly identify with those actions, particularly where those actions are made in 'the heat of the moment,' or in response to fleeting physical or psychological stimuli.

Freedom of action and *self-identification* seem vulnerable to the effects of deprivation that contribute to unhealthy lifestyles. First, the fact that much lifestyle behaviour is habitual, and not generally under conscious control means that factors external to the agent are likely to have a lot of influence over certain aspects of her behaviour. This could contribute to the undermining of the freedom of any given action (where agents struggle to consciously intervene and redirect behaviour), as well as damaging the extent to which the agent self-identifies with her behaviour.

The abundance of cues for unhealthy behaviour is one feature of deprived areas that makes it more likely that individuals in these areas will have their freedom undermined. Further than this, recall that other features typical of these environments make it likely that agents will have reduced capacity for self-regulation and are more likely to behave impulsively. Eating foods high in sugar, fat and salt, smoking cigarettes, failing to conscientiously follow a treatment regimen, or to take other precautionary steps to avoid ill health are all

---

[38] It might even be inappropriate to call such an occurrence an 'action': often the term preferred is a 'happening' or an 'event' as there is nothing active in the agent's will.

[39] Though, this may be undermined if you acted in 'the heat of the moment', perhaps if you were overcome with rage at the person's ignorant comments. You might say that such an action was 'out of character.'

more likely to occur where an agent acts impulsively. Instances of health-harming impulsive behaviours could well also be instances where an agent's freedom of action is undermined, as well as the extent to which she is able to self-identify with that action.

The final aspect of freedom is that relating to the *social status* of the agent, specifically, 'the freedom of the person involved in enjoying a social status that makes the action truly theirs, not an action produced under pressure from others.' Pettit is partly concerned with political liberty. Hence, the sorts of freedom-limiting factors that Pettit has in mind here typically relate to imbalances in power relations: ensuring that citizens have the capacity to express their preferences without fear of persecution, and to uphold the ideals of democracy.

The concerns motivating Pettit's account still resonate in the context of health behaviour. The pervasive connection between socioeconomic factors like wealth, education, and occupation, and health, mean that some individuals are placed in a position of disadvantage from the instant they enter the world. A lack of resources and power leaves some people vulnerable to health risk factors and the poorer health prospects that attend them. There may well be instances where the powerful and resource rich actors in society harm the interests of the more deprived through self-interested behaviour. Consider, for instance, the placement of fast food outlets in areas of high deprivation, so that companies can maximise profits by exploiting the fact that residents in these areas are more likely to be susceptible to cravings for fast food.

The thought here is that there is something particularly unpleasant about freedom-limiting forces coming from exploitation by other agents (not merely the mindless effects of the environment, or our own weakness of will). Thus, the exercise of power by one agent or group of agents, over another agent or agents, subjugates the interests of one under the other. Where we value freedom and equality, we will generally strive particularly hard to avoid the entrenchment of such vastly unequal power relations. Thus, we might feel that the sorts of factors which contribute to inequalities in health are particularly offensive to our ideals of democratic freedom.

From Freedom to Responsibility, Blame, and Desert

I have proposed here that a better understanding of the contributory factors to unhealthy lifestyle adoption gives us cause to question the extent to which agents who behave in this way act with complete freedom. I have discussed how each of Pettit's three aspects of freedom may be compromised when we consider the actions of agents who adopt unhealthy lifestyles, particularly where agents are subject to deprivation.

Limitations on freedom have implications for attributions of responsibility. The conception of freedom adopted here is that which conveys 'fitness to be held responsible.' Where freedom is undermined, the agent should not be considered fully responsible for her actions. Responsibility here is 'merit-based,' and is the kind of responsibility needed in order to make moral appraisals of agents. This can be contrasted with 'as-if,' 'consequentialist,' or 'forward-looking' responsibility. These forms of responsibility do not depend upon identifying any crucial, morally relevant feature of an agent. Rather, they rest upon the balance of positive and negative consequences likely to be brought about as a result of treating an agent as responsible. So, if allowing people to suffer the consequences of their unhealthy behaviour *as if* they are responsible for them causes more people to adopt healthy behaviours, saving the NHS money, then this would constitute a reason for attributing responsibility to those agents. However, insofar as I am concerned with responsibility relevant to desert, I am not considering 'as-if' responsibility.

Agents who are free in the required way, and who can thus be considered responsible, will also be 'reaction-worthy' (they are apt candidates for reactive attitudes as described by Strawson). On the desert-based view, these agents will meet the key initial criteria for being deserving agents. In contrast, those agents who are not fully free, nor fully responsible, nor reaction-worthy in some respect, will also not be appropriate subjects for desert. It is worth emphasising here that these attributions of freedom, responsibility, reaction-worthiness, and desert can apply to isolated instances of an agent's behaviour, and need not be judgements passed about an agent in general. This means that an agent may ordinarily act with responsibility and be an appropriate subject for desert attributions, yet on a given occasion not be.

The arguments criticising health behaviour that this discussion is intended to respond to are those which suppose that agents receiving health incentives act wrongfully in doing so, and are rendered undeserving of health incentives on this basis. The key criteria here is that agents be morally responsible for the desert base which conveys blameworthiness in this context: the unhealthy behaviour. This analysis of the social determinants of health and mechanisms underlying health behaviour has shown that agents may well not be fully morally responsible for the unhealthy behaviours they adopt. Thus, agents may not be blameworthy for their unhealthy lifestyles. Agents who are not blameworthy in this way can no longer be accused of being undeserving of health incentives on this basis.

As mentioned earlier, it is important to remain aware of the consequences of attributions of moral responsibility and blame. In this case, blameworthiness (and desert) is used as a justification for either a) denying the opportunity for agents to be paid to adopt healthy behaviours (in the case of trade-type incentives), or b) denying agents access to a healthcare intervention which could potentially improve their quality of life (aid-type incentives).

Although there are some arguments which are made in support of responsibility (and perhaps desert) as a criterion for allocating healthcare resources, it is unusual for the desert of agents to be taken into account either when determining general healthcare policy, or the treatment of any individual.[40] I do not claim that agents are not *at all* responsible for their unhealthy lifestyles: freedom (and responsibility) may only be subject to some limitation, not abolition. Nor do I claim that this will be the case for all agents, nor even all agents subject to social deprivation. However, I think enough doubt can be cast on the assumption of responsibility and blameworthiness to conclude that it would not be right to deny (otherwise advisable) preventative healthcare interventions to agents who are at risk of developing disease. As discussed, the 'bad behaviour' argument, in the case of aid-type incentives, would have to be sufficiently strong to overcome the general assumption of entitlement to treatment. I do not think it succeeds.

---

[40] It may be that desert is not explicitly taken into account at all. It is, however, plausible that desert-based concerns could implicitly drive policy in a particular direction (for example, if curative, rather than preventative interventions are preferentially funded), or in the treatment of individuals.

An alternative version of the bad behaviour argument directed at aid-type incentives drew upon the comparative nature of desert. In this case, the claim was that those receiving incentives were not completely undeserving of treatment, but that they were less deserving then some. Once again, the consequence of this argument being accepted would be the denial of aid-type health incentives.

Does the comparative case give us more reason for thinking that incentives should not be provided? I do not think so. The analysis presented here has sought to cast doubt on the claim that agents freely adopt unhealthy behaviours, and are responsible for these behaviours. The responsibility requirement would still need to be upheld where the desert assessment is made in comparison to other agents (whose behaviour, we assume, is not subject to criticism). It is still the case that we cannot establish with enough certainty that agents are blameworthy in the way that is required to deny them access to healthcare interventions. It bears repeating that, in the case of aid-type incentives, we should treat incentives as we would any other healthcare intervention. Thus, if agents are undeserving (or less deserving) of this intervention, they must also be undeserving of other forms of health intervention directed at reducing unhealthy behaviours. After all, these forms of intervention also cost money; money that could otherwise be spent on alternative healthcare interventions.

Finally, the desert-based criticism can be directed towards the use of trade-type incentives. The assumption of entitlement is absent here: we do not think that agents are entitled to be offered the incentive trade in the way we might think they are entitled to healthcare. There is a sense in which agents have an entitlement to trade *in general*, which ought to be protected. If we accept that there is no prior entitlement to the incentive trade, specifically, blameworthiness will have less work to do here (its consequences do not involve denying a prior entitlement, and thus seem less severe). So far as desert may play a role in determining whether or not it is appropriate to trade with a given individual, the threshold for the responsibility requirement may be easier to attain. To put it another way, even where agents seem not fully responsible for their behaviour, we may still think they are sufficiently responsible to be judged as blameworthy. If this is so, then the desert-based criticism may be more successful when directed at trade-type incentives.

RESPONSIBILISATION

In this chapter, I have emphasised *moral* responsibility, and the role this plays in desert (praising and blaming agents for their actions). My concern has been whether the appropriate form of moral responsibility is present when agents adopt unhealthy behaviours to justify denying trade-type or aid-type incentives. I suggested that the influences that make it more likely agents will adopt unhealthy behaviours appear sometimes to undermine their freedom, and to some extent, to undermine their fitness to be held responsible.

There are, however, other forms of responsibility, and there may be other values to be gained from incorporating both moral and other forms of responsibility into healthcare. The requirements for an agent to be morally responsible for her unhealthy behaviour must, on my analysis, be set quite high if the potential consequence of attributing responsibility is blame, and even the denial of healthcare.[41] But we may wish to identify responsibility in a more basic sense, for instance, I may be 'responsible' for breaking the vase because it was my arm that knocked into it. This leaves open the question of whether I am also blameworthy (I may have been pushed by someone else) or even praiseworthy (I used the vase to knock out a burglar by smashing it over her head). Thus, responsibility may be used in this less interesting form.

Further, I have already mentioned *consequential* responsibility. Recall, this is where responsibility is used in order to bring about desirable consequences. So, we may want people to bear the burdens (or reap the rewards) of actions they perform (whether or not they are morally responsible, in some sense, for those actions) because this will alter how they behave in ways that produce more utility.

There is a growing trend for 'responsibilisation' in healthcare, and policies that claim to promote opportunities for agents to 'take responsibility' for their health are popular across political parties (Brown [2009]). An NHS 'Constitution' has been produced, the purpose of which is described thusly:

> **This Constitution** establishes the **principles** and **values** of the NHS in England. It sets out **rights** to which patients, public and staff are entitled,

[41] These two need not follow from moral responsibility, but moral responsibility will be a necessary condition if they are to follow.

> and **pledges** which the NHS is committed to achieve, together with **responsibilities** which the public, patients and staff owe to one another to ensure that the NHS operates fairly and effectively. All NHS bodies and private and third sector providers supplying NHS services are required by law to take account of this Constitution in their decisions and actions.
>
> NHS (2012: 2, emphasis in original)

The impression is given of a cooperative service where different actors have duties and responsibilities, as well as rights and expectations, regarding the provision of healthcare. Fairness is mentioned here, and seems to relate to the idea of solidarity - that we all ought to realise that our outcomes are dependent upon others, and theirs upon our behaviour, and that this should encourage pro-social behaviour.

Responsibility may also be a means to efficiency. First, if agents can successfully be encouraged to do what is within their own capacity in order to promote their own health, without the need for costly interventions and medications from the health service, the NHS will save money. The extent to which agents will actually be capable of 'taking responsibility' for their behaviour may be limited, but encouraging the belief that they have control over their behaviour (encouraging self-efficacy) may improve the extent to which agents are able to control their behaviour.

Second, if the outcomes of agents' actions are felt most acutely by the agent (so the burdens or benefits her behaviour creates are not distributed equally across all individuals served by the health service) then this will limit the action of 'moral hazard.' Moral hazard arises when the burdens of costly behaviour are spread amongst a large group. Consider, for example, when a lot of people go out for dinner: invariably, some of the party will eat three courses (generally including steak) and drink as much champagne as possible, knowing that their indulgence will be subsidised by the group. Linking behaviour and outcome by identifying responsibility can help to avoid costly effects like this.

There may be yet further benefits to responsibilisation strategies in healthcare, as Brown asserts:

> Personal responsibility can matter for reasons of fairness, but it can also matter for reasons of utility, self-respect, autonomy, human flourishing, natural duty and special obligation.

It is easy to see why the rhetoric of responsibilisation is popular amongst healthcare policy makers, given the prospect of such benefits. Yet, my earlier analysis suggested that it may often be mistaken to hold agents to be fully morally responsible for unhealthy behaviours, given the influence of social, environmental and economic factors on those behaviours, as well as the importance of the impulsive system of behavioural control (not readily altered by the agent's values and considered preferences).

I do not wish to enter into a discussion about the requirements for other forms of responsibility, and what kind of conditions must hold for particular consequences of identifying responsibility to be justified. However, we might be suspicious of claims that policies claiming to enhance responsibilisation will result in greater 'self-respect,' 'autonomy,' and 'human flourishing,' given the potential for moral responsibility in relation to some health-related behaviours.

Regardless of the outcomes of an analysis directed at the potential for responsibilisation strategies to produce these benefits, it seems unlikely to do much harm. At least, it looks more benign than strategies to incorporate moral responsibility into healthcare with a view to rewarding or punishing behaviour as deemed *deserving*. Responsibilisation may involve additional efforts at education and information provision, in an attempt to create citizens who are well informed about the health-related consequences of their behaviour, and what they can do to be healthy.

Further, the move towards increasing 'personalisation' of healthcare, whereby agents are (supposedly) given more control over what healthcare they receive, where, and from whom, can be seen as an additional push towards responsibilisation. It is not clear the ways and extent to which such additional input into healthcare will always be beneficial to the agent, and it may be that there are also harms associated with greater personalisation, as well as benefits (Cribb and Owens [2010]; Owens [2012]).

Incentives, too, might be seen as a form of responsibilising healthcare, in that they encourage behaviour change in recipients. This can be seen both as an encouragement to avoid ill health: 'responsible' in the sense that one does not burden others but acts solidaristically; and as a way of providing for one's own

healthcare needs; 'responsible' in the sense of independence and self-sufficiency.

It is not too much of a stretch to think that incentives could be accused of undermining responsibility: rather than allowing (and encouraging) individuals to choose their own lifestyles, incentives prod them into making what some other agent deems to be the 'right' decisions. Incentives also require healthcare resources that would be spared if agents were able to change their behaviour independently.

The problem here is that the term 'responsibilisation' and the interpretation of the effect of incentives are sufficiently plastic as to make it easy to argue in either direction and difficult to settle on a conclusion. I do not have a particular interest in arguing either way. I do, however, think that the extent to which (aid-type) incentives support agents to act in line with their considered preferences regarding health behaviour; engage with their own capacity to control their behaviour (and acknowledge other factors likely to influence their behaviour); understand the risks and benefits associated with lifestyle choices; and assist people in adopting healthier behaviours (and reduce disease prevalence); incentives will offer many of the benefits available through 'responsibilisation' strategies.

In contrast, incentives which offer little support other than payment of a reward, and which target individuals who have no prior interest in altering behaviour (trade-type incentives) are likely to lack such benefits. Moreover, these incentives might have the kind of negative effects on the potential for agents to 'take responsibility' for their actions that critics are concerned about. In particular, those incentives that looked likely to be impermissible according to the account of different means of influence in Chapter Four (coercive, highly pressuring, deceptive or otherwise objectionably manipulative incentives) are likely to disrupt the potential for an agent to be responsible for her actions in an important sense.

Concluding Remarks

This chapter has considered the criticism that those receiving health incentives are not deserving of them. Desert plays a role in considerations of fairness (and justice). Although many contemporary political philosophers tend to leave desert out of their schemes of distributive justice, the idea that people should be

treated *according to their just deserts* is important to our everyday conceptions of fairness. The intuitive importance of desert, and the prominence of responsibility-based rhetoric in both responses to health incentives and healthcare policy development more generally, motivated the discussion in this chapter.

In discussing whether or not recipients of incentives might be thought to deserve them, I have drawn out a few different claims. These related to both trade-type and aid-type incentives, and located the problem with incentives in various aspects of the size of the incentive, the desert base, and the moral worth of the recipient. It is this last aspect that motivated the discussion of freedom and responsibility, and the information we might glean from research on the social patterning of disease and the mechanisms underlying health behaviour.

This evidence on the social determinants of health and mechanisms of health behaviour suggests that assumptions that individuals are largely able to control their health behaviour might be misguided. The pervasive effects of deprivation make it far more likely that those subject to indicators of deprivation will suffer poorer health outcomes.

I considered how the freedom of an agent's health-related behaviour (her freedom to 'choose' to live a healthy life) might be compromised along the three aspects of freedom described by Pettit. I argued that, given the likely impact of social and environmental factors on people's ability to choose to adopt healthy lifestyles, it is quite possible that their freedom, and thus fitness to be held responsible, will be undermined. This will not be true for every agent subject to the pernicious effects of deprivation, and nor will those of higher socio-economic statuses be immune from such influences. Indeed, it might be that we should consider those with healthy lifestyles (largely as a result of environmental influence) are perhaps not very 'free' in terms of the choices they have made. However, I was concerned to look at the extent to which those with unhealthy lifestyles might be blamed, and it looks like, at least some of the time, agents who adopt unhealthy lifestyles should not be considered fully free in those actions.

Considering the potential consequences of blaming agents for their poor health (i.e. at the most extreme, reducing their access to healthcare), I argued that it will not be appropriate to blame agents for unhealthy lifestyles and resultant poor health. The case is probably clearer where we are considering

aid-type incentives, and we assume there is a prior entitlement to healthcare. In this case, denying people access to incentives on the basis that they do not deserve such an aid seems to be based on a false assumption about moral blame and desert, and would itself be an injustice.

Finally, I considered the relevance of health policy discussions relating to 'responsibilisation' in healthcare. This implies that further benefits can be gleaned from developing policies which link agents to their actions. The actual benefits possible from such styles of policy are unclear, and the sorts of factors that undermine moral responsibility that I have discussed could hinder the prospects for other efforts at incorporating responsibility into healthcare provision. I argued that some of those benefits thought to arise from responsibilisation may plausibly arise from incentives, particularly where these are of the aid-type variety and tend towards supporting agents to alter unhealthy habits.

There may be further value to efforts at responsibilisation where interventions aim at encouraging considered choice-making by agents. Scanlon (1986) in particular, thinks there are a number of ways in which choices can be valued, and this may be of relevance to healthcare provision and incentives. I will discuss the value of choice, along with other values in the next chapter, which considers whether the introduction of money into healthcare can have corrupting effects.

Six

Does Money Corrupt? Values, Reasons, and Motivation

Introduction

Thus far, I have considered three general themes of argument criticising the use of health incentives. These themes were derived from media criticism specifically directed at incentive schemes, combined with philosophical arguments relating to the ethics (including political and moral philosophy) of healthcare interventions more generally.

Chapters Three to Five covered concerns relating to efficiency, the proper ends of state action and limitations of the state in pursuing those ends; the influence incentives have on the individual and the impact this might have on agents' autonomy and liberty; and the extent to which agents engaging in unhealthy behaviours are responsible, in some sense, for their poorer health outcomes, and whether they are any less deserving of healthcare because of their actions. In many instances, the criticisms of these chapters have been applicable to healthcare interventions quite widely, rather than applying narrowly to health incentives.

In this penultimate chapter I will consider a complaint more specific to health incentives: that the involvement of money in healthcare interventions corrupts something that we value. It is not entirely clear what, precisely, is vulnerable to corruption, nor how this happens when money is used in healthcare, and often part of the appeal of arguments referencing 'corruption' of some sort is their vagueness, hinting toward a 'funny feeling' that the use of money just isn't quite right in these circumstances. I will seek to clarify what the

criticism is, and how it applies to health incentives; it will be interesting to see whether this criticism bears up under scrutiny, or whether its attraction fades along with the rhetoric.

In the press, health incentives are variously referred to as incentives, vouchers, tokens, payments, bribes, rewards, gifts, sweeteners, carrots, nudges, benefits, aids, grants, bonuses, boosts, freebies, prizes, and more descriptions besides. Clearly, there is some uncertainty, or at least variation, of opinion surrounding how we ought to conceive of health incentives. This is perhaps fair, as the use of incentives elides two (usually distinct) worlds of monetary transactions and healthcare. This unfamiliar fusion may lead to hesitation, apprehension, suspicion and hostility. Of course, the opposite may be true, and were the new intervention considered innovative and novel in a positive way then it might be responded to more charitably.

First, I will introduce the corruption criticism as presented by Sandel, and seek to identify where in the process of monetary transactions Sandel recognises the problem of corruption as occurring. I identify two processes through which corruption seems to be thought to arise: *incommensurability* and *diminishment*, and discuss the extent to which these processes seem likely to have the harmful effects proposed.

Next, I discuss what it might mean to act from the 'right reasons' as opposed to acting from the wrong ones, and I briefly introduce some of the theory on practical reason. Next, I consider what values might be candidates for forming the basis of 'right reasons' (those values the promotion of which *ought* to motivate people to adopt healthy behaviours). I consider the plausibility of the claim that reasons for action based on these values are preferable to reasons based on incentives. Further, I discuss the criticism that to act from certain reasons is degrading to the individual, and should be prohibited.

Finally, I assess the claim that motivation for adopting healthy behaviours can be distinguished according to its degree of 'intrinsic' or 'extrinsic'-ness, and that the former kind of motivation is preferable. This section draws upon evidence (and lack of it) from research in behavioural economics into motivation crowding theory. I discuss how trade- and aid-type incentives might differently act as reasons, and motivate agents to alter their behaviour, and the significance of this for an ethical analysis of incentives.

THE CASE AGAINST: SANDEL'S DISTASTE FOR MONEY

Attitudes towards money differ between individuals, contexts, and cultures. What is appropriate in one context (paying for food in a restaurant) is not appropriate in another (paying for food at a friend's house). Money may be used to purchase (or at least make more accessible) objects, experiences, statuses, and relationships, previously available only via other avenues. Michael Sandel (1998, 2012), a philosopher who has achieved prominence beyond academia over the last few years,[42] opposes what he views as a continuing infiltration of money into all aspects of life:

> [T]he extension of markets and of market-oriented thinking to spheres of life once thought to lie beyond their reach… is by and large a bad thing, a development that should be resisted.
>
> Sandel (1998: 93-94)

Sandel has two reasons for suggesting the reach of money markets should be circumscribed. First, Sandel is concerned that money often results in coercion, and second, that monetary exchange can corrupt those values it encounters.

I wish to focus on this second, corruption, argument here. It is, however, worth noting that the coercion argument Sandel presents is slightly different from that discussed in Chapter Four. My focus was on the potential for a monetary exchange to place sufficient pressure on the recipient's will such that freedom becomes compromised. Sandel's concern is focused at the societal, rather than individual, level. The worry here is that the creation of markets where some agents are inevitably worse off than others creates large disparities in power. Those with more resources thus wield more influence over those with fewer resources. Sandel is concerned that the imbalance in wealth and power that results from background economic conditions can lead to pervasive coercion and exploitation of the weakest, across many different contexts of human life. If incentives were to contribute to this general imbalance of wealth and power, or perhaps more modestly to contribute indirectly through the normalisation of market forces, then they may result in these socially realised

[42] Including presenting the Reith lectures in 2009, being the BBC's 'public philosopher' in 2012, and authoring a number of popular books.

coercive effects. I do not consider the severity of this concern here, though there is scope for further investigation.

Much of Sandel's corruption argument centres around social norms, and the detrimental effect that introducing money into certain areas of life can have on these norms. Such norms control the ways we ordinarily interact with people in social situations and determine what will be an appropriate way to behave. Sandel argues that when money is introduced as a way of influencing behaviour, it can 'crowd out' (effectively diminish or abolish) these norms. In some cases, the destruction of social norms may have a detrimental effect on a particular behaviour, such that incentivisation results in less of that behaviour being shown, rather than more. If this happens, then incentivising a behaviour may not be a good way of promoting it.

The example often provided by Sandel here is the case of an Israeli day-care centre where a fine was introduced for parents who were late collecting their children (Sandel [2012: 119]; Gneezy and Rustichini [2000]). The expectation was that the fine would discourage parents from being late, and reduce the number of late pick-ups. However, the opposite occurred and children were collected late more frequently. Sandel's explanation of this effect is that, whilst previously parents felt guilty for collecting children late (and made a concerted effort to be on time), the introduction of the penalty seemed to legitimate being late. Parents now felt it was ok to arrive late as they were paying the staff extra money for looking after the children overtime.

We might question the way Sandel interprets the evidence here, in terms of 'crowding out' a useful social norm. There are other explanations for the observed increase in late pick ups from the day-care centres. For instance, Frey and Jegen seem to agree with Sandel that the day-care centre case involves an unfortunate monetisation, arguing that "The introduction of a monetary fine transforms the relationship between parents and teachers from a non-monetary into a monetary one." Yet in the original paper, Gneezy and Rustichini describe how the day-care centres where the trial was conducted were all private ones, meaning that parents directly paid the centres for the services they provided (Gneezy and Rustichini [2000: 3]).[43] Moreover, when fines for late pick-ups

[43] The authors also describe that day-care centres in Israel, where the trial took place, are composed of a mixture of public and private centres.

were paid this was added onto the regular monthly bill and paid to the owner. It therefore doesn't look quite so obvious that there was a clear change in the relationship between parents and day-care centre staff.

It is further worth pointing out Gneezy and Rustichini's own explanation for the observed effects. They suggest that, prior to the introduction of a fine, the relationship between parents and day-care centres represented an incomplete contract (Gneezy and Rustichini [2000: 10]). That is, parents were unsure what would happen if they arrived late because this was never specified. After the introduction of the fine, however, parents knew exactly what the consequences of arriving late were, and perhaps decided that these were not so bad (at least, less bad than having their child expelled, for instance). The completion of the contracts may have had the effect of reassuring parents, causing them to become more relaxed about arriving late.

There are alternative explanations that could be provided for the day-care centre study, and we need not assume that Sandel is correct in his assertion that here some 'desirable' social norm is destroyed and replaced by some less desirable one. However, it could yet be the case that, in some instances, incentives or disincentives may have peculiar effects, encouraging the penalised behaviour or discouraging the rewarded behaviour. This need not require explanations in terms of social norms in the way Sandel presents it. I will provide some more discussion later in this chapter in the section on intrinsic and extrinsic motivation.

Further to this efficiency-diminishing effect that money may have, Sandel also argues that when we corrupt these social norms through commodification we lose something else of value. Sandel states that whether or not an instance of commodification will be morally objectionable will depend on the nature of the thing commodified:

> In the cases of surrogacy, baby-selling, and sperm-selling, the ideals at stake are bound up with the meaning of motherhood, fatherhood, and the nurturing of children. Once we characterize the good at stake, it is always a further question whether, or in what respect, market valuation and exchange diminishes or corrupts the character of that good.
>
> Sandel (1998: 104)

There are many examples provided by Sandel (2012) of instances where the use of money makes us feel a bit squeamish. A variety of activities are identified, including paying to be able to queue jump at airports and amusement parks; upgrading to a nicer jail cell; paying drug addicts to get sterilized; monetary rewards for children who get good exam grades; trading rights to pollute the environment; hiring people to act as friends; and paying to be able to kill an endangered rhino. Amongst these examples are "bribes to lose weight", also a case where Sandel objects to the use of money (Sandel [2012: 55-60).

There look to be a lot of different things going on with these examples. For instance, killing an endangered rhino looks problematic, quite aside from the fact somebody might pay to do it. It seems that the examples Sandel provides could prove wrongful according to various different moral arguments, some of which have been discussed in earlier chapters (treating people fairly and ensuring distributive justice, and avoiding harm to others in particular). It is a simple point, though one which bears repeating, that beneficial outcomes will not always outweigh wrongful actions needed to create those outcomes. The complexity of Sandel's examples means that it is not clear why *money*, and its corrupting effect on social norms, is a particular problem. I will now discuss further how commodification and the introduction of money into certain areas of life is supposed to bring about an undesirable corruption of values.

COMMODIFICATION : *INCOMMENSURABILITY* AND *DIMINISHMENT*

Sandel argues that when we introduce monetary transactions into areas previously governed by social norms we engage in commodification of those norms and values involved. Commodification means that something that was previously not tradable becomes something that can now be bought and sold and traded with other things; it becomes a commodity. Sandel's objection rests on the claim that turning certain things into a commodity corrupts their value (Sandel [1998: 95]).

Clearly, not all instances of commodification are objectionable, and this practice is commonplace and plenty of commodities are considered extremely precious (and not thought of as inherently 'corrupted' due to their commodity status). Radin (1993) distinguishes between 'fungible' and 'personal' property which are valued quite differently. Using Radin's taxonomy, Duxbury describes

the distinction thus: “Whereas fungible property has a purely economic or instrumental value, personal property is that with which the owner becomes bound up to such a degree that its loss would cause him or her pain that could not be relieved simply by replacing the object with other goods of equal market value.” (Duxbury [1996: 333])

Clearly, the market value of something (when treated as a fungible property) may fail to capture its value in the fuller sense (when we take into account its role as personal property as well). Consider a Van Gogh painting: this would seem to cross many domains of monetary, personal, sentimental, aesthetic, and symbolic value. Putting a £-value on a Van Gogh may only capture one aspect of its value (particularly where this £-value is determined simply by market forces). When commodified, then, objects or activities may have their value ‘downgraded’ because their ‘full’ value is incommensurate with a single, monetised metric. This is the *incommensurability* claim.

Incommensurability is part of the problem where metrics like QALYs are used, as it seems very difficult to capture the full value of treating some disease in these terms. The worry is that when systems of measurement like QALYs are commonplace, their imperfections may be overlooked: we start to assume that the QALY value of an intervention *just is* its full value. Even if we avoid this trap, the work QALYs are put to in developing policy will mean that certain interventions will not succeed in having the right amount of influence because we have failed to value them properly.

Slightly different to the *incommensurability* claim is the *diminishment* claim. In this case, the mere act of placing a monetary value on an object through commodification diminishes the ‘full’ value of that object. In terms of the Van Gogh, one might suppose that by placing a monetary worth on it and thus valuing it as a tradable commodity we diminish some (or all) the value it holds as an art work, in terms of its aesthetic, sentimental, or symbolic value. Much discussion of this *diminishment* problem arise in debates around commercial surrogacy. For instance, Radin (1987) and Anderson (1990) argue that arrangements where a woman is paid to carry a child to term treat the surrogate as a ‘baby maker’ and reduce the parent-child relations of love and protection to ones of contracts, ownership rights, and exchange. They argue that to do this is to diminish the value of such agents and interpersonal relationships.

Commodification results in the improper valuation in both the *incommensurability* and *diminishment* cases. In the first, the assignment of a monetary value to the object is simply to fail to value it properly: its worth cannot be captured in monetary terms. In the second case, to involve the object in a monetary transaction is not to treat it properly, and diminishes the (non-fungible) value that object has. The effects of *incommensurability* and *diminishment* may be to permanently corrupt the valued object, although it seems that sometimes the full value may be reinstated.

I think these stories about *incommensurability* and *diminishment* sound plausible, though not very convincing. It is not clear how valuing something in terms of its monetary, tradable value will necessarily (or ordinarily) be problematic where we also value that thing for other reasons. As described, we often value things *both* in terms of their worth as fungible *and* as personal property. In the vast number of cases it seems we have little problem discriminating between the value of something on the open market as a commodity, and its value in other ways. The work to be done by shared intuition here is too great: Sandel, Radin, Anderson et al. rely on us having shared intuitions that to pay for friendship, queue jumping, surrogacy, gametes, and so on will be to engage those things in monetary transactions in an inappropriate way; will fail to value those things properly; and will corrupt the good in question such that its non-fungible value is diminished.

It may, however, be possible to supplement (and strengthen) this argument with another claim: that to act for the *wrong reasons* may be to fail to value something properly, and may result in corruption. In the next section I will introduce briefly what it is to act for a reason, and indicate how it might matter what reasons motivate our action. In the following sections I consider a number of candidate values which, it might be argued, are vulnerable to corruption by the use of incentives to encourage healthy behaviour.

Right Reasons for Action

One might worry that, when agents act to adopt healthier behaviours and discard unhealthier ones they do so for the *wrong reasons*.[44] That is, it is

[44] See 'The Wrong Message' argument from Parke et al. (2011)

inappropriate that incentives motivate people to change their behaviours, whilst other things do not. I think formulating the concern in terms of what agents take to be reasons for acting can give us more of a grip on the corruption claim.

Without the involvement of reasons, the corruption criticism looks weak: as I have argued, simply valuing something in terms of cash, or as a commodity does not generally destroy the value of that thing separate from this valuation. We are quite adept at distinguishing the market value and personal value of things. Further the process of valuation is highly social: much of the time, values are created through the processes of exchange with other agents, and the way we are seen to treat objects, relationships, and states as valuable. Our experience of things as valuable or disvaluable in particular ways affects how we value them. In turn, the way in which we value these things is expressed through our treatment of them and the attitudes which we hold with regard to them (Anderson [1993]).

Attitudes and valuations can be expressed through acting upon reasons. If I choose to avoid eating meat for reasons of concern about the environmental impacts of the meat industry then this provides an insight into the values I hold about the environment and food. Similarly, if my vegetarianism is motivated by a dislike of the flavour of meat, or the expense of meat, or a desire to ensure as many vegetables are killed as possible, then quite different information about my values may be inferred.

In terms of health incentives, then, choosing to act in the presence or absence of monetary rewards may indicate the value an agent places on goods relating to health. The criticism from the point of view of corruption is that adopting healthy behaviours, when this is motivated by the prospect of monetary gain, fails to express the appropriate attitude (valuation) of health and health-related goods; in expressing inappropriate attitudes towards these goods, the agent contributes to the corruption of the values of those things. This complaint will be amplified when the offensive attitudes are expressed publicly, by those in positions of authority (including the state and healthcare professionals).

It will first be worth giving a little background as an introduction to the idea of acting from a reason, in order to consider what it is to act from a 'wrong' reason (or to fail to act from a 'right' reason). I will then consider the sorts of things

that might count as 'good' reasons for adopting healthy behaviours, and whether failing to act for these reasons (and acting instead because an incentive has been offered) corrupts the value of those things in some way.

There is a considerable philosophical literature on the topic of practical reason (reasons which guide action), and I can only give the barest of summaries here. Hopefully, this will be sufficient to shed some light on the status of health incentives in this context. One of the most influential philosophers of practical reason is Donald Davidson, who proposes that:

> Whenever someone does something for a reason... he can be characterized as (*a*) having some sort of pro attitude toward actions of a certain kind, and (*b*) believing (or knowing, perceiving, noticing, remembering) that his action is of that kind
>
> Davidson (2001: 3-4)

According to Davidson, when agents act on reasons that meet the criteria described above those reasons *rationalise* their actions. That is to say, a reason provides both an *explanation* for why the agent acted as she did, and a *justification* for her actions. We can distinguish between these two functions of a reason, which may be more or less conceptually connected (more, on Davidson's account). So, we can identify reasons that are motivating (or explanatory), and reasons which are normative (or justifying).

Normative reasons - count in favour of or against some prospective action; justify why the agent did what she did

Motivating reasons - those things which motivated the agent to act; provide an explanation for why the agent did what she did

One problem with accounts like Davidson's where reasons are based upon pro-attitudes (such as desires) is that, in fact, it seems one can have reasons one is unaware of, or that one does not consider to be reasons; thus, one may have reasons for things that one lacks a pro-attitude about. For example, Lord Grantham loves pavlova, particularly made by Mrs Patmore. However, on one occasion, Mrs Patmore's poor eyesight means that she accidentally sprinkles salt on the pavlova rather than sugar. Eating salty pavlova is something generally to

be avoided in Lord Grantham's opinion, but unless he believes the pavlova in front of him to be a specimen of salty pavlova, he will not have a pro-attitude towards not eating it (quite the reverse). Yet it seems Lord Grantham *does* have a reason for not eating this helping of pavlova.

To avoid this problem, some theorists introduce further conditions for an agent to have a reason. Often, these involve idealised versions of the agent or the situation. For example, an agent might have a reason when she has a pro-attitude plus a belief and *she is in possession of all the relevant information*, or she may need to be *perfectly rational*, or *perfectly virtuous*.

Another worthwhile consideration is that of a 'primary reason.' Consider: an agent could have a reason to perform an action, be aware of this reason, recognise it as a reason for her to act, be sufficiently rational to act on such reasons generally, and so on, and yet still not be motivated to act by this reason. The agent might have a conflicting reason that pulls her in another direction; or she might be 'weak-willed' and fail to act; or, she might even perform the action, but on the basis of another (better) reason. Thus:

> *Primary* reason - this is *the* reason why an agent acts (not just *a* reason counting in favour of the action). Primary reasons are thus necessarily connected with motivation.

Actions can be subjected to philosophical scrutiny as well. In the context of practical reason, it is important to distinguish between 'actions' and mere 'happenings' or 'goings on.' Korsgaard (1999) insists on the importance of ownership: an 'action' must be authored by some agent. This requires that it be attributable to an agent as a whole, and not merely to some sub-personal part of that agent. For Davidson (2001) the involvement of the agent in the action is also significant, this time requiring intentionality on the agent's behalf to bring about the action.

There are grey areas in determining what will properly count as an action and what will not. For instance, my spilling soup on you because somebody bumped into me will not be an action; intentionally upending my soup onto your shoes will be an action. If I throw my soup on your shoes because I mistakenly think that they are on fire (they aren't), and my intention is to rescue your feet, then I *have* acted intentionally, but just not with the results I hoped

for. If we think intentionality is crucial to action, then throwing the soup over you will count as an action because of the involvement of intentionality. In cases where there is an 'action', the agent 'does something' in a significant way, rather than just being involved in its coming about.

VALUES

I now want to consider some of those values which, in the context of health incentives, might be thought vulnerable to corruption when agents act for the 'wrong reasons.' I will discuss: the intrinsic value of heath itself; the importance of avoiding harming others; solidarity; and the value of choice. I will introduce each in turn and briefly consider how the use of incentives to encourage healthy behaviour might be considered to corrupt that value in some way. My aim initially is just to characterise the values at issue; I will give a more generalised account of incentives and the corruption objection (in relation to reasons) after discussing these values.

The first claim is that health is valuable in and of itself, and that it ought to be a sufficient reason to motivate behaviour change, independently of instrumental benefits it may bring like incentives. Second is the idea that we ought to adopt healthy behaviours because failing to do so may directly or indirectly harm the interests of others. Third is a claim deriving from the value of solidarity: we ought to aim to be healthy so as not to place unnecessary burdens on the healthcare system which must support us all. Finally, I will discuss the value of choosing itself, and how this process might be corrupted by the introduction of an incentive.

Health

Health is clearly of huge importance and value in making an agent's life go well. There are criticisms that to value health in terms of metrics such as QALYs is to fail to value it properly: I have discussed some of the controversy relating to QALYs in Chapter Three. The account I described initially, that identified corruption in *incommensurability* and *diminishment* suggests that processes like QALY calculation or exchanging health or healthcare for money could result in the corruption of the 'full' value of health. As mentioned, however, I do not think this version of the complaint has much bite: to value health in terms of QALYs may well be to do a disservice to its true value, and may further be an

undesirable way of incorporating the value of different healthcare interventions into comparative policy decisions. Yet, I see no reason why this should be seen as unavoidably corrupting the value of health in some important sense.

Money *must* play a role in healthcare provision - the labour of physicians and nurses and other staff must be paid for, equipment must be manufactured and installed, drugs must be developed and produced, and so on. But the mere buying and selling of the paraphernalia surrounding health and healthcare does not change how important health is to *me*, nor to any one: the value of health, in this sense, remains in tact.

I also think there is an ambiguity when supposing that 'health' is commodified through the use of incentives. It is not clear at what this criticism is targeted: is it the state of being in good health itself that is commodified; or is it the behaviour that leads to good health (eating a healthy diet, exercising, getting vaccinated, and so on); or is it the delivery of healthcare in terms of the relationship one engages in with the healthcare provider? In the former case, incentives may miss the mark, as they are targeted at quite specific changes in behaviour. It is difficult to see how rewarding an agent for, say, receiving a vaccination can be characterised as commodifying 'the state of being in good health.' Incentives will not guarantee good health, but may increase the likelihood that one will experience better health in the future.

If we take it that the behaviour is subject to commodification (the quitting of smoking or the consuming of a healthy diet), then I think we struggle to see the problem. What is the deep value inherent in the action of attending a clinic to be screened for cancer or an STI? It is not clear that these behaviours, in and of themselves, hold value that we should be concerned about commodifying. That is not to say that values related to these behaviours, such as health itself or a concern for the welfare of others, have value (which could plausibly be corrupted), but that seeking to locate the value in the behaviour itself seems misguided.

Finally, the kinds of interpersonal relationships involved in healthcare encounters may be something of value that could be corrupted through commodification. I think this criticism has the most purchase: the way physicians and patients interact matters greatly to the comfort of both individuals, the capacity for the physician to provide good healthcare and the extent to which the patient is able to benefit from those efforts. Care (or

benevolence) is a central feature here, as is trust, autonomy, professionalism, and so on. All of these values are (or ought to be) promoted through the way physicians and patients interact. One might worry that if physicians began offering patients money to adopt healthier behaviours then the fostering of these values through physician-patient encounters could be disrupted.

I think this is a real concern, though I do not think expressing it in terms of 'commodification' quite captures the problem. We are not worried that the physician-patient relationship itself becomes 'commodified,' but rather, that the role played by a physician who offers incentives to her patient is not compatible with interactions where care and trust are of primary importance.[45] The most obvious problem looks to be one of inferred paternalism. That is, the patient perceives the physician's offer of incentives as paternalistic because it appears as if the physician is simply trying to 'push the patient's buttons' to get her to behave how the physician thinks she ought to behave. Such inferred paternalism might be experienced as disrespectful, overbearing, infantalising, and manipulative in a way that offering the patient a more conventional therapy such as drugs or a treatment programme or support group, is less likely to be (Goold and Lipkin [1999]; Goodyear-Smith and Buetow [2001]; Goold [1998]).

Clearly, it will matter very much how incentives are presented to agents. Trade-type incentives, targeted at any agent displaying 'unhealthy' behaviour, seem likely to engage these feelings of being subject to undesirable paternalism. This is the case where the incentive is used as a 'payment' to get people to change their behaviour, and would seem likely to disrupt the kinds of values involved in the physician-patient relationship described. Yet in the case of aid-type incentives, the incentive can be seen as one amongst a number of supportive interventions to help people alter their behaviour. If presented carefully, with an explanation as to why incentives may assist with behaviour change (by providing short term goals and rewarding feedback), then they may be perceived as falling within the scope of 'conventional' medical care.

Further, incentive schemes may be presented as *part* of a supportive behaviour change scheme. So, a smoking cessation or weight loss programme

---

[45] There are related worries that an explicit involvement of money in healthcare encounters could corrupt the physician-patient relationship. For instance where patients understand that their General Practitioners are responsible for commissioning particular healthcare treatments they may feel pressure to 'make their case' for why they should receive a particular treatment.

may involve incentives, as well as frequent meetings where advice and support can be provided. It may be that providing incentives in forms other than cash dilutes the negative effects of raising the offer of incentives. Or it might be better that such interactions are distanced from the physician-patient relationship, and that incentive schemes should be provided though means other than an individual's GP or other physician.

It is very difficult to predict how health incentives could affect the values central to physician patient relationships. This becomes more complicated still when we consider the variety of forms incentives can take and the behaviours they can target. It may, for instance, be a very different thing for an agent to be offered cash to take her anti-psychotic medication and for her to be offered an Amazon voucher for completing an STI test. The way individuals personally experience these interactions, as well as the way physicians experience them and behave, will be of central importance, and will be very difficult to predict.

We might supplement the mere involvement of money in health-related domains with the claim that to act for a reason other than the promotion of the good of health itself is not to value it properly. Alternatively, we might frame this as there being something corrupting about taking money as a reason for action. The suggestion is that to fail to take good health as a reason for adopting healthy behaviours is to fail to value it properly, and to express an indifference to this good which eventually results in its corruption.

The problem as I perceive it is that health is enormously important for a whole variety of reasons. First, health may have an intrinsic value of its own. This can be a rather tricky notion, but I do not think we need be too concerned with it here because the overwhelming instrumental value of health is surely more pertinent. That is, a certain amount of health is fundamental to almost anything anybody wants to achieve in life. The value of being healthy is just so pluralistic and so significant that it doesn't seem vulnerable to the kind of corruption we are worried about. In part, this is because health will clearly be of value even when we fail to consciously value it. Our obliviousness to its great value need not, however, diminish its value.

Avoiding Harm to Others

Unhealthy behaviour may result in harm being caused to others. This may be relatively direct, through passive smoking, or failure to receive testing and treatment for STIs, for example. Harm brought to others may be indirectly caused by unhealthy behaviour, for instance, if one's own poor health requires resources from the health service which are then unavailable to others. In all of these instances the behaviour of the individual is likely to impact on other agents in negative ways.

The imperative to avoid harming others will often require agents to behave in particular ways. Some of these are of sufficient importance to be incorporated into law, such as requirements not to assault others or to drive under the influence of drugs or alcohol. Other times, one is morally, though not legally, obliged to constrain one's actions in order to avoid harming others, such as where I ought not to insult you for no reason.

There must be limitations on the obligations to avoid harm, otherwise we would be obliged to forego most of our everyday activities. Feinberg (1984) describes harm in terms of a setback to the interests of an agent. When we talk of obligations not to harm others, this does not usually refer to a requirement to avoid setting back others' interests, but rather, to avoid the *wrongful* setting back of their interests. If we do not include the wrongfulness stipulation, then driving a car down the street, choosing to read the newspaper online rather than buy a hard copy, and many other activities would become impermissible.

The kind of health behaviours where the harm claim becomes relevant, then, may be a much smaller subset of those behaviours that may be targeted for health incentives. One of the most likely candidates is surely smoking whilst pregnant. In this case, there is a clear risk of harm to an identifiable individual, and the sacrifice required of the harmer (the mother) is, I would contend, acceptable given the benefits it would be likely to bring about.[46]

The commodification-based criticism applies here where the obligation to act in order to avoid harming others is replaced with a payment: rather than valuing the rights of others not to be harmed, the agent values actions only

---

[46] The evidence of harm resulting from smoking during pregnancy is clear, indicating a wide variety of problems are more likely to arise in children whose mothers were exposed to tobacco smoke whilst pregnant. See, for example, Olds et al (1994); Ward et al. (2007); Montgomery and Ekbom (2002).

insofar as she may materially benefit from the outcome. On this account, the act of paying people to avoid harming others destroys the value inherent in that action.

Once again, I am not sure this will always be the case. The opposite could plausibly be true, for consider, if I am willing to pay you a great deal to stop you from harming another, does not that indicate my concern for that other individual's well-being? Placing a high instrumental value on certain goods may highlight their importance.

If we consider the importance of reasons here, we might take the concern to be partly that agents *fail* to be motivated by a need to avoid harm, and partly that they *succeed* in being motivated by money. The reasons behind actions can express agents' attitudes towards others and valuations of others' interests. A pregnant woman who takes as a reason to quit smoking the fact that she will be given a cash reward if she does so, but who does not take the risk of harm to her child as a reason shows a callous attitude toward her child, incompatible with the values of motherhood, compassion, and love.

This seems undesirable: it is surely preferable that mothers act out of love for their children rather than being motivated by avarice. Yet, I see a problem with the proposal that incentives are the wrong-making property in this situation. For consider, in the absence of the incentive offer, the mother is not sufficiently motivated by a concern to avoid harming her unborn child to quit smoking. This behaviour might be taken to express the attitude 'avoiding harm to my child is not a good reason for me to give up smoking.' Already, we seem to have a problematic expression of disvaluing the avoidance of harm to the child. The corresponding expression that the incentive is a good reason to give up smoking seems superfluous here.

It is possible that the incentive does make a difference. First, it may be more corrupting to the value of avoidance of harm that agents both fail to act from this reason and succeed in acting from the incentive-reason; there may be an additional expression of comparative attitudes which undermines the value of avoiding harm further. Second, there is also the possibility that offering incentives makes it less likely that the value of avoiding harm will motivate future agents to act. This second point is related to discussions of intrinsic versus extrinsic motivation which I will discuss later in the chapter.

As with the value of health, we must consider both the value the avoidance of harm might have in itself, and the value it has as an instrumental contributor. So, there may well be value to be had from agents acting from the motivation of wishing to avoid harming others, but there is also a huge amount of instrumental value to be derived from people *actually avoiding harming* others. The benefits brought about by a pregnant woman who quits smoking are likely to be significant, as are the benefits of successfully getting more people to be tested for STIs. We should not overlook these benefits (and the costs of foregoing them) unless we are convinced of the negative implications that the actions required to bring these benefits about will have.

Solidarity

Solidarity, as a good, appears slightly different from the values of health and avoiding harm. Both these first two goods should be valued, I have argued, primarily for the instrumental benefits they bring. Solidarity, however, is a good where much of its value derives from the expressive significance it plays in interpersonal relations. One aspect of solidarity is the notion that one ought to avoid harming others, insofar as one recognises and respects the interest of other agents. But further than this, solidarity involves a more communal view of obligations between agents in a shared society. Solidaristic feelings are ones of 'we are all in it together,' and bind people's interests to those of society as a whole. Hans-Martin Sass characterises it as:

> A richer notion than either social or legal reciprocity among freely contracting individuals, it is both pre-supposed by the sphere of self-interested social interaction, because it is a personal virtue, and complementary of that sphere, insofar as it is a principle of social morality, justifying institutions of social justice and welfare.
>
> Sass (1992: 367-368)

Understanding that others experience suffering in the same way that we do, and acknowledging values such as empathy, loyalty, and fraternity leads us to solidaristic systems, such as a socialised healthcare system, which invoke a strategy of risk-sharing (Weale [1990]; Segall [2007]). The continued importance of solidarity, as a means of creating obligations to care for the sick;

promoting social cohesion and 'togetherness'; to avoid draining resources from the healthcare system; and as a means by which to criticise selfish behaviour, is gestured at by a number of theorists writing about medical ethics and healthcare policy (Buyx [2008]; Segall [2007])

The reciprocal nature of solidarity means that members of a solidaristic system are expected to avoid taking too much out of it. That is, if we can easily avoid placing extra pressure on the healthcare system then we ought to do so. Once again, it is not clear what the extent of our obligations are in this case. It would be unreasonable to expect people to refrain from a lot of the activities they enjoy just so as to avoid even a small risk that they might become unwell and require medical attention. Yet people who frequently fail to attend GP appointments, and unnecessarily waste NHS resources clearly do not meet their solidaristic obligations.

Some small changes to lifestyle could be thought appropriate under solidarity, so it is possible that this value is instantiated where agents make such changes voluntarily. Thus, the offer of incentives could act to disrupt the way solidarity is valued in these circumstances.

Indeed, solidarity looks quite incompatible with payment: paying an agent to act with solidarity may very well be to miss the point of what solidarity is about. For to act on the basis of incentives is to act in one's own interests, and not to act with the intention of promoting the interests of society in general. In this sense, it is plausible that incentive payments could diminish the extent to which solidarity operates as a value worthy of motivating action.

Once again, an agent who fails to act solidaristically also fails to do so in the absence of incentives. Thus, the criticism starts to bite only if the introduction of incentives adds a further layer of corruption, which it is not clear that it does. If, however, as mentioned above, the offer of incentives gradually diminishes the tendency for solidarity to be valued as a motive for action, and reduces the tendency for solidarity to be expressed through action, then incentives could play a more active role in this corruption. Such a story sounds like it could make sense, but I cannot make a judgement about the likelihood and extent to which this would happen.

Choice

The final value I wish to discuss is the value of choice, and to consider whether there is value inherent to the act of choosing that is corrupted by being offered health incentives. It is often thought desirable that agents are able to choose certain things about how they live their lives and what healthcare treatment they receive. Choice is often characterised as being a necessary feature where we are interested in linking an agent to her actions via responsibility, and where we want to maintain the freedom and autonomy of an agent (to ensure that her actions are not controlled by others in objectionable ways).

Scanlon (1986, 1998) proposes three distinct values of choice: predictive, demonstrative, and symbolic. The first, predictive value of choice, depends upon the agent being a good predictor of what will make her life go well. There is thus value in being able to chose how to act in that the agent will be more likely to get what she prefers than if someone else tries to judge what is best for her (Scanlon [1986: 178]).

The demonstrative value of choice rests upon the desire to have a particular outcome be attached to aspects of one's character and to the action of choosing. The example Scanlon provides is the demonstrative value of choosing an anniversary gift for one's partner: is it important that one chooses the gift oneself, even if some other agent might be capable of making an equally good or better choice in terms of what one's partner will enjoy (Scanlon [1986: 179]).

Finally, the symbolic value of choice lies in the importance of being seen as an agent who is capable of choosing. In contrast, being denied choice may indicate that one "fell below the expected standard of competence." Thus, even where one is unable to instrumentally improve one's lot effectively through choice, or where one is unable to make a choice that effectively represents some feature of one's character in the choosing, one may still appreciate being treated as an agent who is capable of choosing (Scanlon [1986: 180]).

The expressive values of action might be understood to be instantiated in both the demonstrative and symbolic values of choice described here: to be able to demonstrate something about one's character through the choices one makes is to express one's attitudes about certain things to other agents. To treat another agent as capable of making a choice (and so to allow her to enjoy the symbolic value of choice) is to express an attitude towards her precisely of a moral agent capable of making a choice.

In Chapter Four I discussed whether the influence of incentives was compatible with the values of autonomy and freedom. The concern about choice is related: is the influence of incentives on behaviour compatible with the value of choosing? This could be the case if choice is considered 'commodified' in some sense through the use of incentives, with the result that the value of choice is corrupted. Or, alternatively, if acting for reasons based on incentives expresses an indifference towards making independent choices and undermines the value of choosing.

The predictive, instrumental value of choice is not, I think, likely to be much threatened by incentives. This would occur in the cases discussed where sufficient pressure is placed on the agent to choose in a particular way that she is no longer able to make choices that will track her interests.

The demonstrative value of choice may be disrupted by casting a doubt over the agent's reasons for action. If an agent chooses to adopt a healthy behaviour, this might ordinarily express her valuation of the goods and values associated with this behaviour, perhaps those of solidarity, avoidance of harm, or health itself discussed above. Once an incentive becomes involved, the agent may be perceived as expressing a favourable attitude towards money and wealth, rather than towards those other values. We might well respond differently to an agent we perceive as acting 'for the money,' than to an agent acting on the basis of other values.

It will of course depend on how the incentive influences the agent's choice as to whether we are right to infer what we do into her choices. An agent receiving an aid-type incentive, for instance, may be motivated by a variety of reasons (including those values discussed), and not merely (or even mostly) by the incentive. To assume such an agent's actions in accepting the incentive and altering her behaviour express an indifference to those values and a love of money would be mistaken. Yet another agent might hold exactly those attitudes, and consider the incentive a good reason for acting but the alternative values not good reasons. It seems that our response to these agents - our interpretation of the attitudes expressed through their actions - will not cleanly track the attitudes they *actually hold*.

The symbolic value of choice is also reliant on agents being *perceived* as fit choice-makers (and not simply *being* fit choice-makers). Incentives look plausibly to obstruct this value of choice as well, because recipients of incentives

may be viewed as incapable of making a choice in the absence of incentives, or as not being firm in their choices (and thus being swayed by the incentives), or for a similar reason that suggests recipients of incentives are inferior choice-makers in some way.

The problem for the symbolic value of choice here lies in how we interpret the actions of the incentive-provider, what we believe her reasons for acting to be, and what we interpret her as expressing through those actions (what her attitude towards the incentive recipient are). We might think something like this: the incentive provider does not believe the recipient is capable of making the right choice about her health behaviour; the incentive provider thus decides to seek to control the recipient's behaviour by guiding her choice; the offer of an incentive expresses the provider's attitude that the recipient is not a moral agent capable of making good choices independently.

This may be experienced in the first person, as well as the third. So, if someone offers me an incentive to alter my behaviour I might infer from this that she thinks I am not an agent capable of making good decisions in the absence of incentives and guidance from others. In both the third and first person cases the recipient may experience a loss in the value of symbolic choice. Moreover, the attitudes perceived as being expressed in relation to incentives and a particular form of behaviour may be generalised, and have broader, knock-on effects. Being perceived as an inadequate choice-maker in one situation (with regards, say, to receiving a vaccination), may detract from one's overall status as a moral agent and competent decision-maker.

Degradation

Criticisms are made by those such as Sandel, Radin, and Anderson that the use of money to influence behaviour in certain situations can have a corrupting effect. Often these criticisms, although superficially appealing, seem to lack a thoroughgoing theoretical underpinning that can tell us what is corrupted, in what circumstances it is corrupted, how this corruption comes about, and what significance this corruption holds. I have sought to make this a little clearer in the preceding sections. Sometimes, it seems to be claimed that the mere involvement of money can result in the objectionable commodification of some good, through commodification leading to incompatible valuations (corruption due to *incommensurability* and *diminishment*).

I supplemented this discussion with some theory from work on practical reason. I argued that the way we take certain things to be proper reasons for action, and the attitudes we express by acting on reasons (and not acting on other reasons) may be important to our lives as social beings.

Finally I discussed a few examples of values that might be at stake in the case of health incentives: some of the things we might worry about getting corrupted when incentives are offered to people in exchange for the adoption of healthier behaviours. Inevitably, this discussion will be very limited, dealing with only a few values when there will clearly be many more candidates.

There is a final sense in which corruption may have negative effects that I wish to discuss. The focus thus far has been on the values, which in some cases may be discussed relatively independently from human agents. In some instances - in particular when thinking about the value inherent in choosing, for example - part of the significance of these values relates to the impact they have on the moral status of the agent. There is, then, a sense in which the corrupting effect is imposed on the agent herself. This can be described as *degradation*, and can occur either because some aspect of the agent is 'commodified,' or because the agent acts for particular reasons.

The 'commodification' of the agent is sometimes thought to happen in cases of commercial surrogacy or where there is a market in organs, tissues, or gametes. This idea was touched upon in the section on the value of health, above, where I questioned what exactly about 'health' might be commodified. Possibly this could refer to the act of the agent in altering her behaviour. This sounds quite similar to commonplace arrangements where employers pay employees a salary in order to direct their behaviour in particular ways. When labour is commodified in these cases we generally find it unproblematic, although there are exceptions. Where the employee is underpaid and exploited, or where the work she is asked to perform is very risky or demeaning, we may find the employment contract unacceptable.

The behaviours involved in health incentives do not, themselves, appear demeaning (eating broccoli or going for a run are fairly acceptable everyday activities). Nor is it clear we can describe the payments as exploitative, given that there is no clear market to determine the worth of these behaviours and they would also ordinarily take place without any payment being made. Also, it

is just not obvious that people, or their behaviour, or their health or some other aspect of them is actually commodified in the way proposed.

The other opportunity for degradation arises from the expressive importance of acting for reasons, and the proposal that to adopt healthy behaviours 'for the money' is a degrading thing to do. Sandel (2012) frequently uses words like 'distasteful' and 'unsavoury' in describing the exchanges taking place for money in certain cases. The problem seems to be that, in some cases, and perhaps where health incentives are offered, the monetary exchange is viewed as being beneath human dignity. It might be thought similarly 'distasteful' to have sex in exchange for money, or to drive around in a souped-up sports car.

However, rarely does this kind of disapproval alone justify forbidding agents from engaging in a practice. The point here is that there is just too much variation between different people as to what is considered acceptable behaviour, and what is to be deemed unacceptable, demeaning, and to be met with disapprobation. There are some cases where the freedom of agents is restricted in order to avoid causing offence to others. I am thinking of occasions where entertainment media is censored if it includes explicit material. Further, people may find jobs and friendships at stake if they use racist language or otherwise offensive terms. Thus, more or less informal mechanisms may be in place to define the parameters of acceptable and unacceptable behaviour.

These restrictions of freedom, where they are knowingly and intentionally imposed, may be done both in order to protect others from harm, and for paternalistic reasons (where the agent's freedom is restricted for her own good). Punitive responses to offensive behaviour are an example of the former, whereas outlawing the solicitation of clients by prostitutes is an example of the latter, paternalistic reasons for restricting freedom.

Formal restrictions (that can be met with institutionalised punishments) based on harm to others and paternalism will be relatively rare, as it must be shown that the suffering in each case warrants the restriction of freedom. It will be particularly hard to restrict freedom on the basis of disapproval where the agent herself is deemed to be at risk of harm through degrading herself: if the individual experiences no disgust towards her actions then how can another judge her to be degraded, and harmed by that degradation? I might think it

degrading for people to get drunk and act like morons, or to enjoy watching Jeremy Clarkson, but this alone is clearly no reason to ban these activities.

It is, of course, possible that agents will later come to be embarrassed by their debauched behaviour, but unless this regret is likely to be in the extreme, and unless third parties are capable of reliably judging when this will be the case, there is little chance interference to restrict an agent's liberty will be justified. There may be a very few instances where individual liberty should be restricted on this basis. Arneson (1992) provides one example:

> I would agree that some possible contracts involve a degrading subordination and that the human dignity of the contractor who would be degraded demands that such contracts be condemned and prohibited. Without arguing the specifics of the case, I suppose that slavery contracts should be banned on this ground."
>
> Arneson (1992: 161)

There may actually be rather a lot wrong with slavery contracts, and the degrading effect on the subject may not be the best reason for prohibiting them. However, this may still be an instance where paternalism, on the grounds of degradation, is sufficient reason to limit another's freedom.

It is not clear, however, that incentives provide us with sufficient grounds for worrying about the recipient's dignity so as to restrict their use. It might be argued that agents who accept monetary payments as reasons to alter their health-related behaviours are degraded by doing so. But, as a criticism of an intervention with potential benefits to contribute to agents in terms of well-being and health, this seems weak and unlikely to provide sufficient reason to prohibit the use of health incentives.

Trade-Type and Aid-Type Incentives and Reasons

As argued, different sorts of incentives will interact with money, reasons, social norms, and other features that will be important to determining whether corruption is likely to take place. Context is thus very important, including the nature of the incentive, the recipient and provider, the behaviour incentivised, the situation in which the incentive is offered, and so on. All of these factors could influence how the incentive is perceived and experienced, and so the

extent to which it has the potential to corrupt the agent herself or some other thing that we value.

I want now to look briefly at how the distinction between how trade-type incentives and supportive, aid-type incentives fit into this account. I think the key difference here is the distinct ways these incentives provide reasons for agents to alter their behaviour.

In the trade-type case, the agents have little or no prior intention of altering their behaviour and it is only the offer of the incentive that brings about any change. In contrast, aid-type incentives involve agents who already have some wish to change their behaviour, but have failed (so far) to successfully do so. Here, the incentive increases the likelihood that agents will succeed in changing their behaviour.

In both the trade- and aid-type cases the agent may have reasons she is unaware of, or that she does not recognise as reasons, to change her behaviour. For instance, smokers could be unaware of the health risks of smoking, or underestimate the personal risks to themselves. Clearly, unknown reasons such as these cannot motivate an agent to act. It is also the case that agents do have reasons which they recognise as such, but which still fail to motivate changes in behaviour. This will be the case where those normative reasons are just some amongst many, and overall, the agent is not motivated to alter her behaviour. Or, the agent may think she should change her behaviour, but is unable to do so (through weakness of the will or other barriers to behaviour change). In all of these cases where the agent does not alter her behaviour, she has normative reasons but lacks motivating ones.

Trade-type incentives act both as a normative reason and as a motivating reason for agents to alter their behaviour. Recall that normative reasons are those which justify the agent's actions, whereas motivating reasons are those which explain why she acted as she did. Aid-type incentives may also act as both normative and motivating reasons, but in this case, the agent has plenty of other normative reasons she does take to be reasons. We might further say that this agent has a preference to act upon these reasons: she wishes her normative reasons to also be motivating reasons, whereby this results in her successfully changing her behaviour. It seems sensible to say that the normative reasons this agent had prior to the incentive are sufficient to justify her behaviour change.

However, in order to explain it we have to include the incentive, for without the incentive the change would not have come about.

The difference between trade- and aid-type incentives seems to be that the trade-type incentive provides the primary reason for an agent to alter her behaviour, whereas the aid-type incentive provides one amongst a number of reasons for behaviour change. Clearly, behaviour control is influenced by a whole variety of interacting reasons: agents will rarely be straightforwardly motivated by a single reason, with no counter-reasons operating in the opposite direction. Yet it seems that trade- and aid-type incentives play quite different roles within the matrix of reasons an agent has.

Health behaviours are often very complicated: for instance, dieting involves a number of other behaviours, such as what food one buys in the supermarket, or selects in the canteen, or whether one chooses to cook for oneself or eat in a restaurant with friends. It could be that an agent is motivated to eat healthily when selecting food from the canteen or cooking for herself, but when dining out with friends in the evening she typically selects unhealthy food options. The introduction of an incentive might successfully change her behaviour so that, even when dining out, she tends to select healthier options. For this agent, it would seem that most of the time the reasons guiding her behaviour are those normative reasons relating to things she values about eating healthily, aside from the incentive.

The time sensitivity of reasons seems significant here as well. Where agents struggle to alter health behaviours, this is often because their preferences appear to alter depending on whether they are tired, hungry, emotional, or a whole variety of other factors. This may be due to the increased influence of the impulsive system over an agent's behaviour in certain circumstances. Whatever the underlying behavioural control mechanism might be, again, it seems that incentives provide the right sort of reason to motivate action at the important time.

At the very least, this complex involvement of incentives with an agent's normative and motivating reasons and her actual actions complicate criticisms of incentives based on the idea that agents act for the 'wrong reasons'; that being motivate by incentives will corrupt an agent's integrity or undermine her dignity in some way. Because of their more central role as both normative and

motivating reasons, trade-type incentives will more frequently be exposed to these sorts of criticisms than aid-type incentives.

There is still the concern (mentioned earlier) that providing incentives will actively reduce the tendency for agents to be motivated by other reasons they have to act. This could result either in an overall decrease in motivation, or just decrease motivation from a certain set of (non-incentive-based) reasons. I will turn to this question in the next section.

'Intrinsic' Versus 'Extrinsic' Motivation

The Wrong Message argument from the media (Parke et al. [2011]) included the claim that 'rewards undermine intrinsic motivation.' This will be a negative feature of incentives if and when rewards undermine intrinsic motivation for performing the healthy (desired) behaviour, resulting in an overall decrease in motivation and reduction in the likelihood that the agent will adopt the healthier behaviour. Alternatively, an argument could be made that intrinsic motivation is somehow preferable to extrinsic motivation, thus, even if overall motivation (and likelihood that the desired behaviour will be performed) increases, this may not fully compensate for the loss of intrinsic motivation (in terms of what is a desirable outcome).

The former argument is slightly more straightforward, and turns on whether or not incentives actually succeed in increasing the adoption of healthy behaviours. The complexities here lie in measuring this over time to see if there are long term effects on motivation; effects on motivation in other areas of life (for example, the target behaviour may be successfully promoted, but other health related behaviours may suffer); the wider effects on the motivational tendencies of other individuals, and so on.

The latter argument requires that intrinsic motivation be shown to be more valuable than extrinsic motivation in some way. It must also be shown that such 'valuable' intrinsic motivation is lost when extrinsic motivators are introduced, and that this is not compensated for by gains in healthy behaviour. This latter argument relates to the criticism that to take incentives as a reason to act is not to act for the 'right reasons': that the reasons we would prefer people to be motivated by when they act are eroded by the effects of introducing motivating reasons such as incentives.

Quite a lot of empirical research has been conducted which seeks to explore and explain the effects of extrinsic motivators on intrinsic motivation. Social psychologists and behavioural economists have established a tradition of study that looks at the phenomenon of 'motivation crowding.' Deci and Ryan, and colleagues have developed what they describe as "Self-Determination Theory" (Deci and Ryan [1985]) which provides a framework for understanding how different sorts of motivations can be based on different reasons for action. This all begins with the observation that, contrary to what is predicted by the 'relative price effect' (where increase in reward simply leads to increase in effort), people sometimes respond differently to incentives:

> The Motivation Crowding Effect suggests that external intervention via monetary incentives or punishments may undermine, and under different identifiable conditions strengthen, intrinsic motivation.
>
> Frey and Jegen (2001: 589)

'Intrinsic' motivation here "refers to doing something because it is inherently interesting or enjoyable", whilst 'extrinsic' motivation "refers to doing something because it leads to a separable outcome" (Ryan and Deci [2000: 55]). Further, Ryan and Deci (2000: 56) assert that: "when intrinsically motivated a person is moved to act for the fun or challenge entailed rather than because of external prods, pressures, or rewards."

The term 'intrinsic' motivation is often used more freely in the literature than this definition would allow. In fact, many discussions of motivation crowding use the term 'intrinsic motivation' to refer to what should probably more accurately be described as 'more autonomous extrinsic motivation' (and 'extrinsic motivation' is often used to refer to 'less autonomous extrinsic motivation'). This is because, as described by Ryan and Deci (2000), the vast majority of human action is extrinsically motivated: generally, we do not act for the pleasure of the activity itself. A spectrum of autonomy is thus used to distinguish between different sorts of extrinsic motivations.[47]

[47] Autonomy here is taken to be the psychological construct, for which there are a particular set of measures used to assess autonomy empirically. For a discussion of the merits of some different measures of autonomy in psychology, see Lifton (1983).

Ryan and Deci provide a diagrammatic taxonomy of extrinsic human motivation which shows *external regulation* to be the least autonomous form of motivation, where behaviours are performed "to satisfy an external demand or to obtain an externally imposed reward contingency." (2000: 61) At the other end of the spectrum is *identification*, where "the person has identified with the personal importance of behaviour and has thus accepted its regulation as his or her own." (2000: 62)

To what extent should we take 'intrinsic,' or 'more autonomous extrinsic motivation' to be preferable to 'extrinsic' or 'less autonomous extrinsic motivation'? Further, to what extent do the former sorts of motivation overlap with being motivated by the 'right' sort of reasons, and the latter with the 'wrong' sort of reasons? I propose that there is nothing in-built to these classes of motivation that makes one always preferable to the other. Intrinsic motivation could easily be directed towards undesirable ends (for instance, if one derived pleasure in torturing others). There is nothing inherent in intrinsic motivation that ensures it will always be a good thing that people are motivated intrinsically. Thus, we should not suppose that agents who are motivated intrinsically are more likely to be motivated by the sorts of reasons described in the first part of this chapter, factors like the value of health, the need to avoid harming others, a desire to honour solidaristic obligations and to act as a competent decision-maker. Agents may identify with these reasons, and be motivated by them in a 'more autonomous' way, or they may not.

It may yet be preferable that agents not be motivated to act by external factors they experience as alien to them. For instance, influences such as coercion, or those which preclude rational deliberation over one's actions (see Chapter Four). Ryan and Deci claim that intrinsic (or more autonomous extrinsic) motivation "is associated with greater engagement…, better performance…, less dropping out…, higher quality learning…, and greater psychological well-being (2000: 63). Thus, more autonomous extrinsic motivation may be a preferable way of promoting a desired behaviour because of these instrumental benefits. This might mean it is better if agents are able to endorse the reasons motivating them, and identify those reasons as being consistent with their general character.

These different sorts of motivation may also interact in peculiar ways. Motivation crowding theory seeks to explain phenomena where 'intrinsic' motivation is supposedly undermined (eroded) by 'extrinsic' motivation. This is described as 'crowding out.' The opposite effect may occur, whereby an extrinsic motivator enhances the effects of an intrinsic motivator, and this is called 'crowding in.' Frey and Jegen outline the conditions under which motivation crowding of these sorts are supposed to occur:

> External interventions *crowd out* intrinsic motivation if the individuals affected perceive them to be *controlling*. In that case, both self-determination and self-esteem suffer, and the individuals react by reducing their intrinsic motivation in the activity controlled.
>
> External interventions *crowd in* intrinsic motivation if the individuals concerned perceive it as *supportive*. In that case, self-esteem is fostered, and individuals feel that they are given more freedom to act, thus enlarging self-determination.

Frey and Jegen argue that the effects of motivation crowding are "empirically well-founded and have been observed in many different and important areas of the economy and society" (2001: 591). They provide a number of examples where motivation crowding is said to be identifiable. The case of daycare centres, mentioned earlier as an example often used by Sandel, is provided by Frey and Jegen as a case of crowding out. Recall, fines were introduced to discourage parents from arriving late to pick up their children. However, the perverse effect is seen whereby the number of parents arriving late to collect their offspring actually *increases*, after the fine is introduced. The explanation given by Frey and Jegen is this:

> The introduction of a monetary fine transforms the relationship between parents and teachers from a non-monetary into a monetary one. As a result, the parents' intrinsic motivation to keep to the time schedules is reduced or is crowded out altogether; the feeling now is that the teachers are 'paid' for the disamenity of having to stay longer
>
> Frey and Jegen (2001: 602)

In their concluding remarks, Frey and Jegen assert that strong empirical evidence exists to show that motivation crowding occurs in a wide range of circumstances, including "children's learning behaviour; patients' readiness to take prescribed medication; monetary and symbolic rewards for undertaking various laboratory tasks; the tendency to reciprocate in the laboratory setting, reflecting work conditions in a firm; the amount of trust exhibited in a laboratory situation of incomplete contracts; the reaction of managers to various forms of supervision by their superiors; the readiness to offer voluntary work; the observation of time schedules in daycare centres; the on-time flight performance in the airline industry; the readiness to accept nuclear waste repositories (and other locally unwanted sites); and the amount of civic virtue exhibited, in particular with respect to fulfilling one's tax obligations (tax morale)." (2001: 606)

The empirical evidence for motivation crowding theory is disputed, however. Meta-analyses by Cameron and Pierce (1994), and Cameron, Banko, and Pierce (2001) oppose the conclusions made in a meta-analysis conducted by Deci, Koestner and Ryan (1999).

> Results indicate that, overall, reward does not decrease intrinsic motivation... [V]erbal praise produces an increase in intrinsic motivation. The only negative effect appears when expected tangible rewards are given to individuals simply for doing a task. Under this condition, there is a minimal negative effect on intrinsic motivation as measured by the time spent on task following the removal of reward.
>
> Cameron and Pierce (1994: 363)

> Our results suggest that, in general, rewards are not harmful to motivation to perform a task. Rewards given for low-interest tasks enhance free-choice intrinsic motivation. On high-interest tasks, verbal rewards produce positive effects on free-choice motivation and self-reported task interest. Negative effects are found on high-interest tasks when the rewards are tangible, expected (offered beforehand), and loosely tied to level of performance... Overall, the pattern of results indicates that reward contingencies do not have pervasive negative effects on intrinsic motivation.
>
> Cameron, Banko, and Pierce (2001: 1)

> As predicted, engagement-contingent, completion-contingent, and performance-contingent rewards significantly undermined free-choice intrinsic motivation… as did all rewards, all tangible rewards, and all expected rewards.
>
> Deci, Koestner, and Ryan (1999: 627)

Both groups suggest crowding out may occur when an expected, tangible reward is provided for engaging in a high interest task. None of these papers, however, consider the effects of incentives on health behaviours. Features of the behaviour being rewarded and the incentive on offer will affect the extent to which motivation crowding is expected. For example, key factors will include the level of interest in the behaviour in the absence of a reward; the type and size of the reward; who the reward is offered by; the time frame over which the reward is offered; and so on. It is thus difficult to extrapolate from the conclusions based upon experiments on different behaviours in different contexts to predictions about motivation crowding in the case of health incentives.

Incentives would fall into the category of 'extrinsic' motivators. However, so too would many other reasons motivating health behaviour change, including those discussed earlier (thought by some to be the 'right reasons' for adopting healthy behaviours). None of the reasons described look to motivate the agent to adopt healthier behaviours for the sheer pleasure of so acting; all motivate the agent through the prospect of some 'separable outcome.' The question of pertinence, then, seems to be whether incentives fall in with the *more* autonomous forms of extrinsic motivation, or the *less* autonomous forms.

If incentives are 'less autonomous motivators,' they may be perceived as controlling and liable to crowd out other (more autonomous) motivations. Thus, agents who have normative reasons such as a wish to fulfil solidaristic obligations may be less likely to be motivated by such reasons when 'less autonomous' motivators like incentives are introduced. However, if instead incentives lie at the supportive, 'more autonomous' end of the spectrum, they may actually crowd in intrinsic motivation from such already present normative reasons, and make those reasons more likely to motivate the agent in the future.

Trade-Type and Aid-Type Incentives

The different sorts of reasons provided by trade- and aid-type incentives for agents to adopt healthy behaviours may mean that they typically have different effects on agents' motivation (and so, too, different effects on motivation crowding). Aid-type incentives operate as a psychological aid to behaviour change in individuals who already have normative reasons, which they wish to be motivating reasons, to alter their behaviour. Thus, aid-type incentives are supportive of agents' pre-existing, intrinsic / more autonomous extrinsic motivations to alter their behaviour. It seems more likely that such a form of motivation is quite likely to crowd in, rather than crowd out, agents' existent motivations.

Trade-type incentives operate where the agent has little or no wish to alter her behaviour; she does not have or does not acknowledge normative reasons to avoid unhealthy behaviours. Trade-type incentives do not operate to support pre-existing reasons, but rather to provide a new, normative and motivating reason for behaviour change. These incentives may then be experienced as more controlling than supportive, and therefore more likely to crowd out motivation. However, the overall reduction in motivation this results in may be minimal, considering by their very nature, the agents we are talking about here have minimal motivation for behaviour change in the first place.

The lack of empirical research in this area makes it difficult to predict what the effects of incentives on motivation will be, particularly in the long term (weeks, months, or years after the incentive is removed). It may also be the case that a detrimental effect on motivation due to crowding can be compensated for, perhaps by increasing the size of the incentive to further boost the extrinsic motivation. Motivation crowding may thus have inefficiency effects, but these are not straightforward.

Concluding Remarks

In this chapter, I have sought to explore the claim that health incentives can have a corrupting effect, that they provide the wrong sort of reasons for adopting healthy behaviours, and that they undermine agents' pre-existing motivations to alter their behaviour.

I began with the criticism as presented by Sandel, who argues that valuable social norms can be crowded out when incentives are used to motivate

behaviour in certain situations. This can lead to a decrease in the desired behaviour, and also to the degradation of things that we value. The first, efficiency worry about motivation is picked up later in the chapter where I discuss intrinsic and extrinsic motivation. The second concern is that the commodification of certain things causes erosion of part of their value. This can occur either because those things are *incommensurable* with monetary value, and so characterising them in these terms will always miss a large part of the value they hold. It may also occur because the mere involvement of money in some situations taints the thing of value and *diminishes* the value it holds.

I suggested that an analysis of the agent's reasons for acting might help to fill out the criticism that incentives can be corrupting, particularly where it is not clear what is supposed to be 'commodified' or how this takes place. In acting for a reason, agents express particular attitudes, about how the good towards which their actions are directed should be treated; about how they wish to present themselves to others; about what appropriate justifications for action are; and so on. The expression of indifference towards particular reasons for action, where those reasons are based upon values such as health, avoiding harm, solidarity, and choice, may corrupt those values by indicating the agent's lack of interest in being motivated to act by them.

I argued that the values of health, avoiding harm, solidarity and choice are not clearly corrupted by the introduction of incentives. One overarching point is that, in the absence of incentives, such values would not act as motivating reasons for agents anyway. Thus, it is not clear that the addition of incentives makes the chance of corruption more urgent. In both the health and harm avoidance cases, I argued that the instrumental value of these goods is more important than the intrinsic value they may hold. Thus, if incentives are able to boost contributions to whatever valuing health and harm avoidance would normally contribute to (such as well-being), then this is likely to compensate some loss due to the corruption of the intrinsic value of health and harm avoidance.

Solidarity and choice look more vulnerable to corruption via health incentives. These goods derive value through the agent being motivated by them (or perceived as being motivated by them), rather than by external rewards such as incentives. In the case of solidarity, we might point to the fact that agents changing their behaviour due to incentives are already not motivated to adopt

healthy behaviours due to solidarity, though there may be some loss in the value attributed to solidaristic actions due to the public nature of incentives, and some individuals perceiving others as not acting solidaristically, and therefore giving up any solidaristic tendencies they might hold.

The value of choice may also be undermined by the operation of incentives, if agents are less able to make choices that reflect their characters (though incentives could actually help with this, where weakness of the will is at play), or where agents are perceived as not competent decision makers due to being offered incentives, or when they feel as though they are treated as less than fully competent in this regard.

It does seem that incentives could corrupt some values, such as those of solidarity, and certain values derived from making decisions independently. However, as with all the criticisms against incentives and other health interventions, such losses must be fitted into a bigger picture, and it must be considered how severe we feel those losses will be. For instance, if we refrain from using health incentives to encourage pregnant women to quit smoking on the basis that it may undermine solidaristic obligations or detract from the symbolic value of choice, we are assuming that such losses are unacceptable, even when we consider the health benefits that would potentially go to the unborn child (as well as further benefits to the mother and wider society). It is certainly worthwhile to look outside the narrow parameters of well-being and cost harms and benefits, but a sense of proportion must be maintained at all times.

The final part of the chapter considered whether incentives are likely to have a deleterious impact on motivation: undermining 'intrinsic' motivation by introducing 'extrinsic' motivators. First, I qualified the claim that intrinsic motivations are always preferable to extrinsic motivations, arguing that there is nothing inherently better about motivations described as 'intrinsic' (or 'more autonomous extrinsic'). However, where agents experience instrumental benefits from acting from intrinsic motivations rather than extrinsic ones, it may generally be preferable to encourage the former motivations for action rather than the latter.

The main claim, however, is that extrinsic motivation can actually have inefficient effects by crowding out intrinsic motivation and decreasing overall motivation. The evidence for this is scarce in the context of health incentives,

and this makes it very difficult to judge what the likely motivation crowding effects of incentives would be, as this is so dependent on context and the way incentives are experienced by the recipient.

I did offer some tentative thoughts, however. In particular, given that aid-type incentives motivate individuals who already have a preference to change their behaviour, such extrinsic motivation seems more likely to be perceived as supportive than controlling, and thus to crowd in rather than crowd out motivation. Trade-type incentives, on the other hand, seem more likely to be perceived as controlling, and thus to crowd out. This may be undesirable, but the impact on overall motivation may be limited because these individuals lack much intrinsic motivation prior to the offer of the incentive.

The main points from this chapter are that, once again, incentives that are intended to act as a psychological aid to behaviour change, supporting agents who already have a wish to alter their behaviour, are less likely to corrupt other things that we value (and that might otherwise motivate behaviour), and are less likely to have deleterious effects on agent motivation for adopting healthy behaviours.

CONCLUSION

I began this thesis by describing the huge significance behavioural choices have for population health. Too much food, alcohol and cigarettes, and not enough exercise, combined with non-adherence to treatment regimens and a failure to utilise healthcare resources such as disease screening and vaccination programs, result in diseases which are both unpleasant for those affected and costly to the state in general. Yet, some unhealthy behaviours have historically proven to be surprisingly resistant to change, with many efforts directed at informing people of the risks of unhealthy lifestyles and encouraging healthier alternatives failing to deliver significant improvements in health. The challenge posed by health behaviour change provides the context within which health incentives arise. Rewards for healthy behaviour could provide another means by which healthcare providers might reduce the disease burden and improve health.

The proposal of using financial incentives to reward healthy behaviour raises a number of hard questions, some of which demand ethical analysis of the underlying issues. It has been my task, here, to address some of these ethical issues, as part of the wider project of CSI Health to consider the viability of health incentives more generally. To this end, I began with the public discussion of health incentives in the news media. Of interest to me were those arguments relating to health incentives that could provide a starting point for identifying those concerns that this thesis should address.

In Chapter One, I discussed how the media can act as a 'source of arguments' relating to health incentives, and some of the special features this source possesses. I emphasised that particular pressures operate on media content due to the interests of those parties responsible for creating copy. To some extent, these pressures are thought to cause media content to resemble the opinions held by its consumers. For this reason, it is proposed that media content can act as a proxy for Public Opinion.

My intention has not been to present an ethical analysis of those arguments *most concerning to the public*. Rather, I have sought to combine a sensitivity to such arguments with a more conventional philosophical approach which tackles those principles of philosophical concern typically arising in the context of bioethical debates about healthcare interventions. In abstracting the specific tokens of arguments featuring in the media from their original settings and grouping these together into types of arguments (as done in the media analysis), important contextual information is lost. 'Public opinion,' as represented by the media arguments, has been further polluted by splicing and combining these arguments into general themes, more amenable to philosophical discussion.

The result is an analysis that begins with the media but swiftly moves away from it. I hope the end result, with four chapters relating to four general themes of concern, still covers most of the arguments arising in the media, although not in the form in which they originally arose. This has proved to be a more manageable way of structuring the discussion, and I hope it has created a clearer narrative than might have been possible were the discussion to have focused more strictly on the discrete arguments in the media.

Thus, the general themes relate to: issues of efficiency and the capacity for incentives to effectively promote well-being; the extent to which incentives are consistent with ideas of justice and the fair treatment of people; concerns about the harmful effects on the recipient of pressuring, perhaps coercing, people to change their behaviour; and finally, the harmful effect on values that we hold to be important that might arise from the involvement of money. When looking at the philosophical implications of these themes in relation to incentives, some aspects of concern seemed to fade away, whilst new worries emerged.

Chapter Three, then, begins the substantive discussion of health incentives by taking up some of those issues that tend to concern political philosophers about state action: ideas relating to efficiency, well-being and what the state ought to be promoting through healthcare, and the limitations of state interference. Although not the most obvious 'ethical' issue, efficiency was by far the most frequently referenced aspect of incentives in the media analysis. I argued that 'efficiency' should not be considered simply a calculative problem for economists to resolve (although this is certainly one aspect of it). Rather, much of the difficulty with determining the relative efficiency of different interventions relates to the kinds of positive and negative effects that should be counted, and the importance and value of these effects given the healthcare context in which they exist. As such, much debate about 'efficiency' turns out to depend on numerous value-laden judgements, rather than on 'objective' calculations.

I suggested that, sometimes, efficiency seems to be used as a smokescreen for other concerns which are typically more moralised and involve assumptions about the 'right' way to spend money. The apparently uncontroversial nature of a need for 'efficiency' might make this an appealing frame for raising other arguments. Moreover, in the case of incentives, the explicit involvement of money encourages a focus on the costs and savings this particular health intervention will involve. Such factors could explain why there is a strong tendency to refer to efficiency-related issues when discussing heath incentives. It is, however, important to emphasise the complexity of arguments taking place in this area.

This chapter was therefore concerned with some of the ways in which these kind of disputed value judgements involved in 'efficiency' inform the acceptability of health incentives. In the context of what the state ought to do, I suggested that one of the clearest limitations on state interference is set by the requirement to respect agent liberty. All state action, including efforts to improve public health, will therefore balance the desirable outcomes that may result from interventions affecting agent freedom, and the need to maintain those freedoms. I went on to discuss this in terms of the legitimacy of perfectionist and neutral state policy, arguing first, that moderate perfectionism should be acceptable, and second, that health incentives should be considered legitimate according to such a position.

I also went on to give a brief discussion of some allocative issues that arise in the context of state action, and in particular healthcare. The distribution of healthcare resources is a key function of the state, and various theories within political philosophy propose ways in which this distribution (or the distribution of social goods more widely) should be directed. To give an indication of the distributive implications of incentives, without being able to properly assess the actual impact incentives would have on resource distribution, I discussed a few example theories: Rawlsian liberalism, luck egalitarianism, the capability approach, and communitarianism. My aim here was to indicate that incentives need not be ruled out by these approaches (with the probable exception of communitarianism). As an intervention, incentives are flexible enough to direct resources towards particular populations, and they are consistent with efforts to ensure that the worst off are prioritised. This does, however, come with the caveat that more affluent individuals are generally more likely to benefit from healthcare interventions: such an effect may also be seen in the case of incentives.

In seeking to provide an efficient system of healthcare, public organisations such as the NHS must balance the interests both of the public and of the individual. A focus on the rights of the individual is typical of medical ethics literature, which emerged from a need to protect patients who were perceived as vulnerable in the face of the expertise and power of physicians. Hence the preoccupation within bioethics with individual freedom, to find ways of delivering healthcare that ensure the individual may exercise her autonomy, whilst acknowledging the expertise of healthcare professionals, and also protecting the interests of society as a whole.

Chapter Four thus considered the influence incentives might have on the individual and her psychological freedom, beginning with one of the strongest forms of interference: coercion. This chapter was split into three parts: Part One introduced the concept of coercion and discussed a number of different approaches to describing coercion in the philosophical literature. I sought to indicate the variety of approaches to describing this concept, and the kind of features of the situation in which proposals are made, and the features of the recipient of those proposals, which are important to judgements of coercion. Two key criteria linked to coercion, either in combination or independently,

were the thought that a coercive proposal makes the recipient *worse off* in some way, and the thought that a coercive proposal places the recipient under a certain (significant) degree of *pressure*.

In Part Two I then sought to apply this theory to the case of health incentives, and consider if incentives can coerce. I described how particular psychological idiosyncrasies might be exploited to design effective incentive schemes, and how this could result in individuals being made to experience the non-payment of an incentive as a 'loss,' and therefore, how they might be made 'worse off' by an incentive proposal. Yet, as I argued, it still seems unlikely that incentives will have sufficient pull over an agent's will to be considered coercive.

Finally, in Part Three I extended the discussion to include some other means of influence: persuasion, manipulation, and nudging. After briefly characterising these methods of influence I considered the extent to which incentives might be thought to utilise them in changing people's behaviour and, importantly, whether this kind of influence should be considered unacceptable on the grounds that it interferes with agent liberty and autonomy.

As with coercion, I argued that the analysis of other means of influence should give us little reason to think that incentives will have liberty-undermining effects on recipients. However, sometimes it seems that incentives could have autonomy-undermining effects where they manipulate, through distraction, deception, or by encouraging action based upon sub-optimal or faulty reasoning. Yet the extent to which we should be concerned about these effects can only be measured with a clear eye on the value of autonomy in these contexts.

To further consider the significance (and possibility) of 'autonomous' action, I provided a description of the reflective and impulsive determinants of behaviour developed in social psychological research. This describes how much behaviour (including in the domain of health-relevant activities) is not controlled by reflective, conscious, value-sensitive mechanisms, but rather by impulsive, automatic, environmentally-cued processes. Considering what kind of autonomy is possible when behaviour is controlled in this way suggests that the effects of incentives might be less worrisome than at first it seemed.

I placed a lot of emphasis on the importance of agent experience and individual psychology in determining the effect of incentives on agent autonomy, and when this will be harmful, and when unjustified. Thus,

contextual information and facts about the particular agent under consideration will be fundamental to judgements of this sort. However, certain generalisable 'risk factors' for incentives more likely to coerce, or otherwise negatively affect autonomy, might be identified. I suggested these included incentives which target 'vulnerable' agents; where recipients have little money or other resources; where incentives are of a large magnitude; where recipients experience the non-receipt of incentives as a loss from the status quo; where incentives promote others' interests above the agent's; where incentives exploit faulty reasoning processes or employ deception. In these instances, it seems more likely that incentives could exert increased pressure on the recipient's will, or manipulate her in some unacceptable way.

Chapter Five considered some issues relating to justice. Rather than seeking to consider justice as a whole (a rather daunting prospect), I focused on the aspect relating to desert, and whether it is fair to furnish individuals with rewards when they have previously failed to adopt healthy behaviours (which some argue citizens are obliged to adopt in the first place). This is the concern often articulated as incentives 'rewarding bad behaviour.'

I discussed a number of ways in which the desert-based criticism might be fleshed out, but suggested that on the whole, these were unconvincing. At the very least, far more conceptual work on the part of the critics would be needed to establish the bases for such criticisms. In particular, I argued that evidence from research into the social determinants of health and the psychology of health behaviour undermines common assumptions about the robust link between an agent and her (health-related) behaviour. Generally, it is taken that moral responsibility is a necessary prerequisite for blameworthiness, and moral responsibility is usually taken to require control over one's actions. However, the empirical evidence I summarised suggests that health behaviour has less to do with freely chosen decisions, expressing our preferences about how we wish to behave, and more to do with the circumstances we find ourselves in, and factors external to our control. This makes it problematic to identify agents as morally responsible, and thus blameworthy, for their behaviour.

Moral responsibility can be tricky to judge, and it also has connotations that go beyond praise and blame (denying an agent is 'morally responsible' may be to deny her the status of being a full moral agent in some sense). I used

Pettit's theory of freedom as fitness to be held responsible as another means of identifying what features must be present in order to consider agents fully free and thus, fully fit to be held responsible. Desert is generally taken as a feature of a particular individual, yet my primary concern here is with the desert of groups of people; those (potentially) in receipt of some form of health incentive. Such individuals are connected in the minimal sense that they are candidates for the incentive, and they may share salient features due to being part of this group (for instance, they may be smokers or they may live in a particular area). Whilst not all of those who possess such features will have their freedom (in Pettit's sense) undermined by the kinds of forces described in Chapter Five, it is possible that this may happen some of the time; where incentives target agents with particular features, the very features they hold in common may detract from the extent to which they should be considered fully free. On the basis of responsibility-dependent desert, then, it looks like withholding healthcare (in this case, health incentives) will not be justified.

In the final chapter, I considered those criticisms that propose that the use of health incentives to change behaviour - particularly where this takes the form of money - is likely to corrupt values that are important to us. As discussed, this criticism is rather diffuse, and often seems appealing without being very specific about what, exactly, is corrupted, how this occurs, and why this is undesirable. First of all, then, I sought to clarify what the concern seems to be. As presented by Sandel (one of the key proponents here), the introduction of money into certain areas of life results in commodification and, sometimes, corruption.

Given the claim that money represents the 'wrong' reason for acting in certain circumstances (such as adopting a healthy behaviour) I considered what it means for something to be the 'right' reason for action, and outlined four values that could be candidates for providing us with 'good' reasons for acting, and which could plausibly be corrupted via the kinds of processes described. These were the intrinsic value of health; the need to avoid harming others; the value of solidarity; and the value associated with choosing for oneself. I argued that it is not clear that such values will ordinarily be corrupted by the use of health incentives, and that more would need to be done to show that such corruption is a real threat, and moreover, one that is worth avoiding (that is, that compensatory effects of incentives would be insufficient). The social value

of solidarity and being seen to be a free and competent 'chooser' may be harmed by the introduction of health incentives as a (perceived) motivator. Yet once again, the significance of this loss must be balanced against the corresponding gains.

The final section considered the distinction between 'intrinsic' and 'extrinsic' motivation. It is sometimes claimed that it is better for agents to act from intrinsic motivation than from extrinsic motivation, both because extrinsic motivation can 'crowd out' intrinsic motivation and result in overall lower levels of motivation, and also because intrinsic motivators are somehow preferable. The 'preferable' sense of intrinsic motivation seems to relate to features like those 'good' reasons discussed. Again, I argued that it is problematic to evaluate the precursors of action in this way without more conceptual work being done. Further, there is little evidence to directly support the assertion that health incentives crowd out intrinsic motivation, with most research being conducted in different contexts, making it hard to extrapolate to the health behaviour case. Whilst it may be that 'supportive' external motivators should be preferred to 'controlling' ones, it is not clear that incentives fall on the wrong side of this division.

Having summarised the discussion in the main part of this thesis, it is worth making a few more general points that have emerged from this analysis. I will do this in the final concluding sections.

Key Features

Trade-Type and Aid-Type Incentives

There are a few general features that I have identified as making a scheme more or less likely to look ethically dubious. Most of these seem to fall along the trade-type / aid-type distinction I have described. The former, trade-type incentives, act by providing an extra motivation that 'tips the balance' of preferences in favour of the incentivised behaviour. The archetypical recipient here is one who has no prior interest in altering her behaviour, but who may be induced to do so by the offer of an incentive. In this case, we may think the agent is 'paid to change her behaviour.' In the latter case, aid-type incentives support an agent to better match her behaviour to her consistent, long-term preferences about how she wishes to act.

These are the idealised versions of trade- and aid-type incentives, where both the recipient and the provider intend and experience the incentive in a particular way (as either a trade or an aid). Yet in the real life case, things will rarely be so neat: recipients' preferences may fluctuate, making it hard to assess if an incentive acts to induce or support behaviour change; those providing incentives may misjudge the recipients, intending to offer a supportive, aid-type incentive but in fact offering something that has the effects of a trade-type incentive; and so on. Still, I think the distinction is helpful, and there are some general remarks to be made about the 'purest' forms of these two types of incentive.

In the context of the state's role in promoting health efficiently and justly, features of less or more permissible incentives may match up with features reminiscent of trade- and aid-type incentives. For instance, I suggested that a moderately perfectionist strategy should be acceptable, if a government intervention to promote health is to be deemed acceptable. Interventions that respect agents' conceptions of the good life, and support them to succeed in pursuing such lives will be more consistent with moderate perfectionism than interventions which seek to alter those conceptions, or to get people to act against them. The former, more acceptable form of intervention sounds more akin to an aid-type incentive (supportive of prior preferences); the latter style of intervention seems to operate as a trade-type incentive (indifferent regarding agents' initial preferences and interested only in the goodness of the outcome).

When considering worries about the potential for incentives to pressure the agent, to override her will, or to manipulate her in some way, we may also think that aid-type incentives are more likely to be benign. This is because aid-type incentives are aimed at supporting the agent's preferences, not altering them. The will of the agent is what aid-type incentives are intended to promote, rather than the interests of some third party.

In terms of their affect on autonomy, the features of aid-type incentives would appear more promoting than stifling: here the incentive helps an agent to act consistently in ways she prefers. This may relate to her ability to exert reflective control over her behaviour, rather than being driven by more impulsive desires which respond more to environmental cues than endorsed values. Trade-type incentives look more problematic, with the potential to

create, rather than resolve, conflict between an agent's two systems of behavioural control.

Where the concern with incentives is whether or not the recipient is deserving of them, I argued the trade / aid distinction could be significant. This is because there is a moral difference between denying someone access to a trade that she would prefer to have access to, and denying someone access to a form of healthcare that she would otherwise be entitled to (this requires certain assumptions about the access to incentives prior to judgements about desert). I argued that in the latter case, the justification (based upon blameworthiness of the agent) must be stronger than in the former case. It may be that we should *never* accept blameworthiness as justification for removing someone's entitlement to healthcare (even where we can reliably judge people as blameworthy).

Finally, when considering corruption, the trade / aid distinction may do the most work. The main complaint here relates to the commodification of behaviour, which seems uniquely applicable to the use of trade-type incentives. Where agents accept incentives only as a form of psychological aid, to help them achieve their behaviour change goals, they do not appear to 'commodify' behaviour in the required way. In the trade-type case, it may be that there is a monetary exchange taking place whereby the thing of value (and which, in this instance, is valued in terms of money) is the behaviour of the recipient. Commodification (and corruption) is more plausible here, although this by no means guarantees it will occur.

I sought to flesh out this distinction a little more in terms of reasons, and the different role trade-type and aid-type incentives play as explanations and justifications for action. I suggested that, in the trade-type case, incentives play a key role in both explaining and justifying the agent's behaviour (they provide both motivating and normative reasons). In the aid-type case, however, incentives play a smaller role in motivating the agent's behaviour, and they may not be necessary to justify it; the story we tell about why the agent changed her behaviour is more likely to refer to her prior reasons for adopting a healthier lifestyle than to the incentive. That said, the complete picture would need to include reference to the incentive as well.

In the context of intrinsic and extrinsic motivation, it might further be of relevance whether incentives take a trade- or an aid-type form. Although, as I

argued, evidence on the applicability of motivation crowding theory to health incentives is scarce, as a minimum we might expect that supportive, aid-type incentives are less likely to result in the crowding-out of intrinsic motivation (and more likely to crowd-in this motivation) than are trade-type incentives. The latter, where these are perceived as controlling, may have the crowding-out effects that critics are concerned about.

If aid-type incentives are less likely to have morally objectionable effects, then it may be wise to try to restrict the use of health incentives to this form, and to avoid the use of trade-type incentives where they look likely to be contentious. Yet it will not be straightforward to reliably restrict incentives to one or other type because of the importance of contextual factors, including the agent's individual psychological experience. However, some basic guidelines might be feasible.

Targeting agents with a prior desire to change their behaviour seems like a helpful way of restricting the actions of health incentives to the aid-type form. This will emphasise the importance of processes of recruitment of agents onto the incentive scheme. For instance, offering incentives to agents who have gone out of their way to seek assistance in quitting smoking will target those incentives towards people with prior preferences for this behaviour change. Offering incentives to people one sees smoking on the street is less likely to successfully target incentives towards those who already wish to quit.

Such methods will not be infallible: word may get out that an incentive scheme, initially scrupulously targeted at those already wishing to change behaviour, is operating. This may attract others who seek the incentives more than the behaviour change, and it may not be possible to screen such participants out. Thus, those who would experience incentives as trade-type (not aid-type) could become involved.

Incentive type and size may also play a role here. Offering very large incentives may have the effect of attracting far more people, many of whom have no prior preference to alter their behaviour but would very much like to receive the reward for doing so. Similarly, grocery vouchers may be less appealing to those with little interest in healthy eating. It is not clear where the cut-off will be between aid-type incentives of a sufficient size to effectively change recipients' behaviour, and trade-type incentives large enough to

indiscriminately recruit people. More by way of empirical evidence will be needed here in order to draw the correct boundaries.

Beyond the Trade / Aid Distinction

Efforts to target incentives carefully and to pay close attention to who is recruited, as well as to the design of incentive schemes, will have ethical significance beyond the trade / aid distinction. I noted in Chapter Four that where incentives combine elements such as being very large; offered to vulnerable agents (such as those suffering from mental illness); experienced in some sense as a 'loss'; offered by powerful members of authority (including physicians); and so on, they may be more likely to place significant pressure on the recipient's will. Such incentives could have undermining effects on agent liberty and autonomy.

Incentive schemes may also be used in isolation or alongside other methods of behaviour change. The supplementation of an incentive scheme with a support group, for instance, may increase success rates, and may also foster the kinds of values that critics of incentive schemes worry get lost, particularly in relation to the physician-patient relationship (such as trust, benevolence, and respect). More work needs to be done to understand how incentive schemes and other behaviour change interventions could complement each other, but it could be that where these act in combination, the benefits are increased and the harms mitigated.

I have not given much consideration to the merits of promoting particular behaviours. A range of health-related behaviours have been discussed, and some more gestured at. I have allowed the assumption that most of these behaviours will be relatively uncontroversial, and likely to improve health (and very unlikely to harm it). This might be true for a very general account, but at the level where a particular behaviour is described (such as quitting smoking or being tested for chlamydia) it should be possible to be more specific about the likely costs and benefits, both to the individual and to wider society, of that behaviour. Further than this, when we consider the *specific individual's* likelihood of benefitting or experiencing harm from a behaviour change intervention, both in terms of her health and other contributors to her well-being, the picture may change yet again.

We might also wish to identify schemes which, for one reason or another, attract particularly hostile media coverage. In the analysis discussed in Chapter Two, those schemes targeted at improving the health of pregnant women were described as attracting the most positive coverage, whilst those aimed at illicit drug users or overweight populations received more criticism in the press (Parke et al. [2011: 10]). Whether or not this hostility appears warranted upon closer inspection, it may nonetheless prove obstructive to the effective implementation of particular incentive schemes, and healthcare providers would be wise to take such factors into account.

Differences Between Schemes and Individual Experiences Will Matter

It is, thus, highly significant that, within the broad description of 'incentives to encourage healthy behaviour,' a huge range of variations exist. I have narrowed my focus somewhat to consider only *positive*, *personal*, *conditional*, *financial*, *health* incentives. Yet even within these limits incentives can still come in a variety of different forms: cash payments, lottery tickets, luxury goods vouchers, grocery vouchers, or anything with a discernible monetary value. This value can be large or small, or represent a small chance of a large reward, or large chance of a small reward. Incentives can be offered by physicians, economists, psychologists, politicians, some alliance of these disciplines, or anyone else who wishes to change behaviour. They may target smoking cessation, diet, physical activity, vaccination uptake, screening for STIs and cancer, adherence to medication, or something else thought to promote health. The incentives may be transferred in a lump sum, at weekly or monthly intervals, or may even be repossessed if the agent fails to meet her targets. Schemes could last for any amount of time: days, weeks, months, or even years.

All of these factors, and more besides, can affect how incentives should be evaluated. At times I have referred to different sorts of schemes in order to illustrate how a particular scheme might look in whatever case I am discussing. Sometimes it is useful to discuss the most controversial schemes, such as those which pay pregnant women to quit smoking, or mentally ill people to take their medication. But the way we assess the fairness, or coercive influence, or corruptive nature of one scheme may differ a lot from another. Hence, my aim has been to give a more abstracted account, and to indicate what general

features will be important, rather than seek to evaluate individual instances of health incentive use.

Psychology is also central to some of the discussion in this thesis. For instance, when considering coercion, I emphasised the importance of the pressure placed on the agent's will. I argued that fundamental to the ethical significance of a particular application of (potentially coercive) pressure on the agent's will is the experience of the agent. This will clearly vary between different agents depending on characteristics unique to those individuals.

Other features important to this ethical analysis will also depend greatly on the psychological experience of the agent. These include well-being, moral responsibility, freedom to act, and blameworthiness. Such things may be thought to include subjective, agential factors as well as independently identifiable factors, and the degree to which each feature requires reference to agential and non-agential factors will depend on the theory adopted in each case. Yet it seems likely that on the majority of accounts, at least some information about the agent and some information about the situation will be needed.

Sometimes it is unhelpful to place an emphasis on what occurs within somebody's head, as this can be fiendishly difficult to measure. If, however, this seems to be where the main action is taking place, then there is little to be gained by avoiding this conclusion. It need not preclude assessment by third parties: many methods have been developed within psychology precisely for the purpose of estimating what goes on within the human mind. Where it is difficult to know with much certainty the experiences of an agent, educated guesswork may be necessary. It still seems preferable, however, to make uncertain assessments of what we actually care about than to take accurate measurements of something that misses the point.

Limitations and Further Research

As with any piece of research, the range and depth of issues covered within this thesis is necessarily restricted. I have only been able to consider some of those criticisms of incentives of ethical interest, not all. Nor does this consideration permit one to draw conclusions about the overall permissibility of using incentives to encourage healthy behaviour: as described in the introduction, the

negative approach I have adopted here to responding to certain criticisms is insufficient, lacking as it does any positive account in favour of the use of health incentives.

It is probably also the case that any one of the chapters in this thesis might have been extended significantly, perhaps to the point of constituting a thesis in its own right. Most of the discussions have been necessarily circumscribed both in their dealing with the philosophical background and development of theory, the implications of such theories for health incentives, and the normative conclusions we might draw from these discussions regarding the practical implementation of incentives. The efficiency of healthcare interventions and the relevance of well-being; the legitimate role of the state and the justifiability or otherwise of perfectionist policies; the best theories of distributive justice and how healthcare ought to be allocated; the effect of incentives on liberty and autonomy; the relevance of desert in determining healthcare; the extent to which agents can be considered morally responsible for their behaviour; the processes thought to result in corruption through commodification; the kinds of reasons for action that promote the proper valuation of goods: all of these could be discussed more deeply with relevance to the ethics of health incentives and perhaps philosophical research more widely.

More use could be made of the media, and other methods for eliciting the attitudes towards health incentives held be members of the public. A closer consideration of the criticisms raised here could prove fruitful, coming up with issues that I have overlooked. A more sociological approach to this work might provide a better understanding of how agents perceive themselves and others in relation to healthy and unhealthy lifestyles; what significance health has in this context; how healthcare should operate here; the special significance of money and how its operation is modified in different relationships; and many more issues besides.

Such a discussion also suggests the opportunity for a more sociological perspective on the issues that have been covered here. The most obvious example, to my mind, is the discussion of how the influence of incentives might place coercive, or otherwise undesirable pressure on the recipient. My focus has been undeniably agent-centric and psychology-focused, yet the kinds of influences and pressures operating on agents who exist in society is manifold and complex. The kinds of power imbalances that exist in every interaction

create their own pressures to act in particular ways, and sometimes these will be *experienced* by the agent (in that she *feels* pressure in some way), and sometimes they will go completely unnoticed. Such a discussion would require a subtle analysis of the wider relations between incentive recipients and providers, visible healthcare providers (physicians, nurses) and invisible ones, the media, family, friends and peers, influential businesses, the state, and other actors.

Another worthwhile extension would be to consider the use of health incentives in a more global context. My discussion has remained mostly rooted with the UK, yet incentives can (and are) used in countries all over the world. Different social, cultural, economic, and political situations will affect the ethical consideration of incentives elsewhere. Different behaviours are likely to be targeted depending on the healthcare needs of a particular country. Attitudes towards money may influence the acceptability and efficacy of incentives, and different sized incentives may prove effective or necessary. It may be that a structure roughly similar to that adopted in this thesis could be used in other global contexts, or it might be that the consideration of similar issues would prove unproductive for health incentives used elsewhere.

I have shown that, on the whole, the criticisms of health incentives considered here do not quite hit the mark. It has yet to be shown, however, whether health incentives possess the necessary features to justify their implementation. Are incentives even effective at improving health? Would they produce sustainable benefits? Could they be practically implemented? There is still a great deal that is unknown about the applicability of health incentives, leaving further empirical and theoretical work still to be done. It would be futile to implement ethically benign interventions that have no clinical impact. Thus, more work in psychology, economics, sociology, epidemiology, and medicine, as well as philosophy and bioethics will be necessary in order to answer the final question of whether and when health incentives *should* be used.

Despite these limitations, I hope this thesis has served to shed some light on the ethical implications of using incentives to encourage healthy behaviour. If nothing else, I have shown that the ethical implications are complex here, and that knee-jerk reactions of suspicion and hostility should not be accepted

without further scrutiny of the underlying moral principles at work. When we apply such scrutiny, often, some of those concerns seem misguided.

APPENDIX A
EXAMPLE INCENTIVE SCHEMES

Below are a few examples of incentive schemes that have operated in the past or continue to operate. This list is far from exhaustive, and is intended only to be illustrative of the kind of forms schemes can take.

Give It Up For Baby

Ballard and Radley (2009); NSMC (2010)

This is a smoking cessation scheme targeted at pregnant women in Tayside, Scotland. Women were offered an incentive of £12.50 per week for every week they didn't smoke (as demonstrated by taking a carbon monoxide breath test). The £12.50 was redeemable through a National Entitlement Card at ASDA supermarket, and could be spent on groceries but not alcohol or cigarettes.

Women were enrolled on the scheme for the duration of pregnancy and for three months after the birth. Free Nicotine Replacement Therapy was also provided to women, as was one-to-one support from a 'Give It Up For Baby' Development Worker.

Financial Incentives for Adherence Trial (FIAT)

Priebe et al. (2009)

The FIAT scheme uses financial incentives to encourage mentally ill patients (those diagnosed with schizophrenia, schizo-affective psychosis, or bipolar illness) to adhere to anti-psychotic depot medication treatments. The

medication is given in the form of an injection, up to once a week, for a period of 12 months. For every injection received, patients are paid £15 in cash.

Individuals recruited onto this trial come from a pool of previously poorly-adherent patients (missing 50% or more of prescribed depot medications over the last four months) for whom all other available methods of improving adherence have failed. Those with learning difficulties or a poor command of English are excluded, and participants must be capable of giving informed consent to take part in the study.

Wee for a Wii

NHS Cambridgeshire (2009); Cambstakeatest (2009)

The 'Wee for a Wii' campaign was run by NHS Cambridgeshire and NHS Peterborough and their respective community services from 2009-2010 (other similar schemes have been run in other regions as well). People aged 16-24 were encouraged to provide a urine sample or swab to be tested for chlamydia. Tests could be obtained via text message or online for free (from the website freetest.me), and returned via post.

Those returning a completed test kit were entered into a prize draw to win either a Nintendo Wii games console, cinema tickets, or HMV vouchers.

Project Prevention

Project Prevention (2011); BBC (2010)

Not 'health promoting' as such, but nonetheless an interesting (and ethically contentious) case of incentive use, Project Prevention is a charitable operation originating in the United States which offers alcohol and drug addicts money in return for complying with long term contraception or sterilisation. Those meeting the requirements of Project Prevention can receive a one off payment of $300 dollars in return for compliance with certain forms of birth control.

In 2010, Project Prevention began operating in the UK as well. Although the majority of Project Prevention participants are women, the first person in the UK to be recruited was a 38 year old man who underwent a vasectomy in exchange for £200.

Pounds for Pounds (Weight Wins)

Relton et al. (2011); Weight Wins (2013)

Weight Wins is a private enterprise that offers cash rewards to participants who successfully meet weight loss goals. The scheme is flexible, so participants can set their own targets and time frame, and the rewards they can earn will vary with these factors. Those joining up to Weight Wins must pay a fee to do so, and have monthly weigh-ins to show if they have lost, gained, or maintained weight.

Pounds for Pounds was a pilot scheme run by NHS Eastern and Coastal Kent, which operated through Weight Wins. Eligible adults needed a body mass index (BMI) over 22.5 kg/m^2 (a BMI of 25 or over is considered overweight [NHS {2013}]). The maximum length of plan was 13 months, and rewards ranged from £70 to £425 per year. Payments were credited to participants on a monthly basis, with a bonus payment of 50% of the total maximum reward paid at completion if the individual achieved her final target weight. Participants were also provided with a booklet of weight loss tips.

doi: 10.1111/j.1369-7625.2011.00719.x

Financial incentives to encourage healthy behaviour: an analysis of UK media coverage

Hannah Parke BMedSci,* Richard Ashcroft PhD,† Rebecca Brown BA,‡ Theresa M Marteau PhD§ and Clive Seale PhD¶

*Medical Student, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, † Professor, School of Law, Queen Mary University of London, London, ‡PhD student, School of Law, Queen Mary University of London, §Professor, Kings College London and ¶Professor, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK

Correspondence
Clive Seale
Professor of Medical Sociology
Barts and the London School of Medicine and Dentistry
Queen Mary University of London
2 Newark Street
London E1 2AT
UK
E-mail: c.seale@qmul.ac.uk

Accepted for publication
15 June 2011

Keywords: conditional cash transfer, health incentives, mass media, public opinion, public response

Abstract

Background Policies to use financial incentives to encourage healthy behaviour are controversial. Much of this controversy is played out in the mass media, both reflecting and shaping public opinion.

Objective To describe UK mass media coverage of incentive schemes, comparing schemes targeted at different client groups and assessing the relative prominence of the views of different interest groups.

Design Thematic content analysis.

Subjects National and local news coverage in newspapers, news media targeted at health-care providers and popular websites between January 2005 and February 2010.

Setting UK mass media.

Results The study included 210 articles. Fifteen separate arguments favourable towards schemes, and 19 unfavourable, were identified. Overall, coverage was more favourable than unfavourable, although most articles reported a mix of views. Arguments about the prevalence and seriousness of the health problems targeted by incentive schemes were uncontested. Moral and ethical objections to such schemes were common, focused in particular on recipients such as drug users or the overweight who were already stereotyped as morally deficient, and these arguments were largely uncontested. Arguments about the effectiveness of schemes and their potential for benefit or harm were areas of greater contestation. Government, public health and other health-care provider interests dominated favourable coverage; opposition came from rival politicians, taxpayers' representatives, certain charities and from some journalists themselves.

Conclusions Those promoting incentive schemes for people who might be regarded as 'undeserving' should plan a media strategy that anticipates their public reception.

Introduction

The use of monetary (and quasi-monetary) rewards to influence behaviour is commonplace in everyday life. Certain activities are encouraged through payment (performing one's job, for instance), whilst others are discouraged by making them more costly (by such means as speeding fines). In these cases, the offering or demanding of money is explicit, but other ways of delivering financial rewards and punishments can be more subtle. Taxes are raised on environmentally damaging means of transport, whilst more 'sustainable' practices receive tax breaks and subsidies. Similar methods have been used to influence behaviours important to health. For example, taxation is used to discourage smoking and alcohol consumption.[1,2] In the health field, incentive schemes involve giving people cash or quasi-cash (e.g. grocery vouchers) conditional upon an identifiable change in behaviour (e.g. quitting smoking and taking exercise).

Recent UK government initiatives have included health incentive schemes to tackle obesity,[3] encourage physical activity,[4] improve the health of pregnant women[5] and reduce illicit drug use.[6,7] Outside the government, other schemes have arisen, including the use of cash payments to improve treatment adherence for psychotic disorders[8,9] and a variety of local health service initiatives concerning smoking cessation[10,11] and testing for sexually transmitted infections (STIs).[12,13]

This article reports a study that aimed to describe the content of popular and professional media to assess the extent of favourable and unfavourable coverage of health incentive programmes within the UK, to compare schemes targeted at different groups and to compare the extent to which the views of differing interest groups gain prominence.

Background

Given the pervasiveness of money as a motivator for behaviour change, and the general acceptance of its use under certain circumstances for discouraging unhealthy behaviours (such as high taxation rates on tobacco and alcohol), it is perhaps surprising that financial incentive schemes attract much controversy. Yet, proposals to offer financial incentives to people encouraging them to adopt 'healthy' behaviours have been criticized by academicians[14,15] and, as the articles we retrieved in the media analysis reported here show, by politicians, patient representatives and a variety of others. Some criticism is targeted at the general idea of using financial incentives to change health behaviours, whereas other criticism targets specific incentive schemes (for example, in psychiatric treatment).[14,16,17]

Why is there any interest in proceeding with incentive schemes at all then, given the apparent hostility towards them? It cannot be said that there is overwhelming evidence for their efficacy.[18–21] Whilst there is evidence to suggest they could be useful in some areas,[22–25] rationales for incentive schemes in the UK are also influenced by social, cultural and political factors, one of which has been the 'choice agenda' of the UK government in the past decade. This focuses on individual choice and 'empowerment' as methods of promoting competition and quality of health, education and welfare services.[26–29]

Approaches to health care have become similarly individual-choice focused, with the major theme in bioethics in the last half of the twentieth century being autonomy.[30,31] Medical care has moved away from medical paternalism towards a system where patient choice is central.[32] Although autonomy and choice stand in a complex relationship, there is a tendency to identify them in practice and policy.[31]

This focus is rather in tension with another major theme in health care: an emphasis on public health and disease prevention. This typically involves non-specific interventions affecting large numbers of people. In some cases, those subjected to the intervention may not benefit directly and may have no option to refuse the intervention (for example, the introduction of fluoride into the water supply). Despite these worries, it is largely through such measures that the greatest improvements in public health have

been achieved.[33] Understandably, given growing concerns about the impact of chronic diseases that often result from lifestyle factors, preventive medicine has become a priority for UK governments. A new body, 'Public Health England', will have ring-fenced funding specifically for public health interventions[32] and will draw on the approach of the Nuffield Council on Bioethics in 2007 to tackle public health problems.[34] This approach tries to reconcile the population and individual perspectives, but there is a consistent emphasis on individual autonomy and freedom.[35]

Financial incentives should therefore be viewed in the context of the current trend for promoting individual choice and a need to address the potentially catastrophic effects of unhealthy lifestyles. Thaler and Sunstein[36] propose ways in which behavioural economics and social psychology can inform interventions aimed at behaviour change, largely by altering the physical or social environment in which the behaviour occurs. They propose to offer a way of 'improving' peoples' decisions without restricting choice. Although popular in policy circles, skepticism about the effectiveness of 'nudging' people in this way has been expressed elsewhere,[37] and financial incentives sit somewhere close to such nudges. However, financial incentives more explicitly alter the nature of the choices available to individuals.

The UK government in the first part of 2010 indicated a continuing interest in using incentive schemes to improve health.[32] Understanding the response of mass media to health incentive schemes is therefore important. In the UK, the National Institute for Health and Clinical Excellence (NICE) launched a consultation exercise to assess public opinion to inform NICE policy and guidance on health incentives.[38] This reflects concerns raised about the acceptability of the UK health incentive schemes listed earlier.

Mass media

Popular media both reflect and shape public opinion. Early studies of mass media audiences focused on measuring the effects that media messages were felt to have, treating these as propaganda that media producers were attempting to 'inject' into the minds of their audiences.[39] The view that audiences play a more active and sometimes oppositional role in relation to media messages subsequently grew in popularity. The growth of the Internet, though, has seen a breakdown of strict divisions between media producers and media audiences, as access to the production of messages using this medium has become increasingly democratized. This has been accompanied, in the case of British newspapers, by a significant trend towards 'tabloidization' and popularism, whereby journalists seek to reconcile their claims to act as the 'voice of the people' with their obligation to satisfy the business interests and political preferences of their owners.[40]

In addition, journalists depend on 'sources' for stories, or for injecting controversy into stories where the existence of differing viewpoints is an important element,[41] as they are in stories about health incentives. Much day-to-day journalistic practice depends on reporting the content of press releases issued by sources that journalists consider authoritative, such as the government, other politicians, leaders of professional or lobbying groups. On controversial topics, competition between sources for access to mass media coverage can be intense, so that the media becomes an important site for struggles over policy.[42]

In a media landscape marked by the proliferation of news and entertainment channels across different technological platforms, it is clear that the Internet, as well as democratizing access to production, has added to the ways in which conventional broadcasters and print media journalists reach their publics. Coupled with the effects of search engines in focusing attention on the sites of major institutions, this has ensured that the views of conventional 'big media' producers are highly prominent on the Web[43,44]

It is clear, too, that study of the more specialized media targeted at health professionals is important if we are to understand the messages that circulate around this constituency, who are the front line of workers seeking to alleviate

health problems that incentive schemes are designed to prevent or solve. Professional magazines targeted at doctors, nurses, health service managers and other health-care workers, the news sections of health research journals, and websites aimed at health professionals are relevant here.

Methods

Search terms (see Appendix S1; Supporting Information) were used to retrieve all UK national and local newspaper articles about health incentive schemes from the Nexis database, an archive of UK newspapers, published between 1/1/2005 and 18/2/2010. The terms were used in different combinations until new combinations retrieved no new articles. Prior to 2005, there was little coverage of UK health incentive schemes, and all of the schemes eventually identified were introduced after 2005.

The BBC online news archive (http://www.bbc.co.uk/news) and the online archives of *Pulse*, *BMJ*, *Lancet* and *Nursing Times* were searched for news items. We searched a further 19 websites providing medical and nursing news (for example, http://www.rcn.org.uk; http://www.medicalnewstoday.com), identified from the top 20 Google hits for 'medical news' and 'nursing news.'

Articles were included if they contained a minimum of one paragraph discussing the use of health incentives in the UK. News articles, editorials, letters, features and comment pieces were included unless they were from media targeted at health-care professionals (such as the *BMJ*) where only news items were included so that academic discussions of schemes were excluded. The final sample consisted of 210 articles.

Articles were first categorized according to the six types of health behaviour they discussed [weight control, health-promoting behaviour in pregnancy, antipsychotic medication adherence, illicit drug use reduction, smoking cessation and testing for sexually transmitted infections (STIs)]. A seventh category covering articles discussing incentive schemes in general was added (see Table 1 for results relating to this exercise).

A coding scheme was then developed to identify themes within articles. This firstly identified and categorized different arguments for and against incentive schemes (see Appendix S2 for a list of these categories with definitions and examples). Articles were then categorized into those that contained solely favourable arguments, solely unfavourable and a mixture ('mixed') (see Table 2). Finally, text reporting the views of the different sources used by journalists (for example, government spokespeople, opposition politicians, doctors) was categorized into the groups shown in Table 3.

Where statistics reporting the frequency of coding categories are reported in the text later, a count of the number of times a coded segment occurred is first given, followed by the proportion of articles in which at least one such coded segment occurred, expressed as raw numbers and then a percentage. For example, 'used 63 times in 47/210 (22%) articles' means that the coded item was found 63 times at various points in the text, that it occurred in 47 of the 210 articles and that 47/210 = 22%.

A kappa statistic of 0.8, indicating substantial agreement between independent coders (HP and CS), was achieved for a sample of 10 articles involving 370 coding decisions. Coding and retrieval of coded segments were carried out using *NVivo* software, which enabled statistical patterns to be identified.

Results

A breakdown of the 210 articles included in our analysis shows that articles in national (76 articles) and local (78) newspapers were more common than in media targeted at health- and social care professionals (11 articles). The BBC website produced 28 articles, with 17 on other websites. Nationally announced or government-sponsored schemes (health in pregnancy (60), weight control (37) and illicit drug use reduction (27)) received more coverage than local or wholly non-governmental initiatives [antipsychotic medication adherence (10) and STI testing (7)], with the exception of smoking cessation

Table 1 Number of articles: publication by type of incentive scheme (Supporting Information)

	Weight control	Health in pregnancy	Antipsychotic medication adherence	Illicit drug use reduction	Smoking cessation	STI testing	General	Total
Popular								
Local newspapers	8	30	1	7	26	4	2	78
Daily Mail	6	2	0	4	2	0	0	14
Mirror	3	6	0	0	3	0	0	12
The Sun	1	1	1	1	1	0	4	9
Daily Telegraph	3	1	1	4	1	1	0	11
The Times	4	2	1	1	1	0	3	12
Guardian	4	5	1	1	0	0	2	13
Independent	2	1	1	0	0	0	1	5
Total popular	31	48	6	18	34	5	12	154
Professional								
BMJ	1	0	1	1	0	0	0	3
Nursing Times	1	0	0	1	1	0	1	4
Practice Nurse	0	0	0	0	1	0	0	1
Pulse	0	1	0	0	0	0	0	1
Lancet	0	0	0	1	0	0	1	2
Total professional	2	1	1	3	2	0	2	11
Websites								
BBC	3	4	2	4	10	2	3	28
Other websites*	1	7	1	2	2	0	4	17
Total websites	4	11	3	6	12	2	7	45
Total all	37	60	10	27	48	7	21	210

*5 of 19 produced hits: http://www.healthcarerepublic.com; http://www.medicalnewstoday.com; http://www.news-medical.net; http://www.staffnurse.com; http://www.rcn.org.uk.

Table 2 Arguments by type of scheme (number of articles)

	Unfavourable	Mixed	Favourable	All
Weight control	10	22	5	37
Health in pregnancy	3	22	35	60
Antipsychotic medication adherence	0	9	1	10
Illicit drug use reduction	8	18	1	27
Smoking cessation	2	32	14	48
STI testing	0	6	1	7
General	5	13	3	21
Total articles	28 (13%)	122 (58%)	60 (29%)	210 (100%)

schemes (48), which were local (see Table 1 for a full breakdown).

Table 2 shows the overall argument of each article by the type of incentive scheme. The table shows that there were more favourable articles (29%) than unfavourable ones (13%), but the majority (58%) presented mixed coverage. Coverage of health in pregnancy was notably favourable (35/60 = 55% wholly in favour). Coverage of weight control schemes (10/37 = 27% wholly unfavourable) and illicit drug use reduction (8/27 = 30% wholly unfavourable) were the most critical of the use of incentives.

Whether a target group of recipients were deemed 'deserving' was a factor influencing support or opposition. For example, a politician who criticized the provision of a cash benefit to

pregnant women was vilified in terms that would not have been possible without the underlying assumption that this category of person was deserving of help:

> Callous Tory Peter Lilley has astonishingly attacked benefits given to pregnant women. The Mirror, June 2009.

Thirty-four conceptually different arguments were identified and coded. Fifteen were arguments in favour of incentives, and 19 were unfavourable towards them. These arguments are presented separately with examples in Appendix S2, but the major ones are described here in groups, which relate to common overarching themes. These concerned arguments about the problems the schemes were designed to solve, their effectiveness, benefits and harms for both participants in schemes and for society as a whole and the moral or ethical issues raised by the schemes.

The problems schemes are designed to solve

It was common for proponents of schemes to describe the disease or health problem that schemes were designed to solve (used 63 times in 47/210 (22%) articles) or to emphasize how widespread and serious this problem was (117 times in 78/210 (37%) articles). Examples are as follows:

> Low birth weight babies and premature deliveries are much more common in mothers who smoke. BBC (website), February 2008.

> Experts say that by 2050 at least 60 per cent of the UK population will be obese – so fat their health is in danger. Daily Mail, November 2008

The view that incentive schemes are best targeted towards the hard to engage, disadvantaged or vulnerable as a last resort was used 45 times in 32/210 (15%) articles, as in the following:

> Financial incentives might be a treatment option for a high-risk group of non-adherent patients with whom all other interventions to achieve adherence have failed. BBC (website), January 2007.

No arguments were found that took an opposite view to these.

Effectiveness

By contrast, arguments both supporting and opposing the view that schemes were effective were advanced. Supporters of incentives were commonly shown (130 times in 82/210 (39%) articles) presenting positive evidence, or good reason, to believe that they worked, citing research evidence, experts' opinion or individual testimonials:

> Such schemes have been used in the US with research showing participants stay drug-free for twice as long as those not taking part in incentive schemes. BBC (website), June 2008.

> Mrs Belcher of Whimple, Devon, who weighed 11 stone when she started, said, "It's a little bonus that kept me determined to finish." The Times, October 2009.

Arguments proposing that incentives were ineffective were also presented. Sometimes it was simply stated that there was a lack of evidence and sometimes that there was evidence of ineffectiveness (29 times in 19/210 (9%) articles). For example:

> There is however little research that shows that a financial incentive, combined with nutritional advice, is enough to persuade mothers from the most deprived areas to change their lifestyle. News-Medical (website), September 2007.

> Administrator Betty Reed, 46, who smokes 30 cigarettes a day, says "maybe the bribe would have a good effect. But for me, when you are spending £8 a day on cigarettes, £12.50 is really nothing." The Herald, June 2008.

The view that schemes would not offer long-term solutions, or did not address the root of the problem, was presented 46 times in 24/210 (11%) articles, as in:

> This is no kind of long-term solution- a temporary financial incentive won't stop people putting the weight back on once they have got the cash. Daily Mail, January 2009.

Occasionally [6 times in 4/210 (2%) articles], the opinion was asserted (unsupported by any evidence) that schemes simply would not work:

> If it were just a question of money, they would have stopped eating years ago. After all, if they ate less, they would be richer. The Daily Telegraph, January 2008.

Benefits and harms

Very occasionally (just 3 times in 2/210 (1%) articles), the view that no harm can come to people involved in incentive schemes was expressed, as if the author of this idea imagined that potential for harm was a possible criticism of the scheme being proposed:

> There is no harm intended or caused- the service users can revoke the offer at any time. The Metro, January 2007.

More commonly (124 times in 94/210 (45%) articles), the view that an incentive would help people to do what was in their best interests (including mentions of the health benefits to the individual) and the view (50 times in 34/210 (24%) articles) that incentives would help relieve some of the financial pressures on the recipient were mentioned:

> The one-off payment is intended to help pregnant mums stay fit and healthy in the run-up to the birth. South Wales Echo, April 2009.

> Stephen Timms, Financial Secretary to the Treasury, said: "We understand that the run-up to a birth is an expensive time for families." The Western Mail, January 2009.

These are benefits to recipients, but benefits to others, including to society as a whole, were also mentioned. There was the view that incentives would lead to cost savings to the health service (79 times in 54/210 (26%) articles), and other benefits to society such as a reduction in crime committed by illicit drug users (26 times in 19/210 (9%) articles) were presented. For example:

> The National Centre for Health and Clinical Excellence says its plans – to be piloted in up to six centres – will save the NHS money in the long run. BBC (website), July 2007.

> If it works to keep people in treatment there would be considerable benefits to the public. The Daily Telegraph, January 2007.

> Even small incentives could make a real difference not only to patients' lives, but also to the lives of those around them. BBC (website), July 2007.

More rarely mentioned were the views that incentives could have a beneficial effect in addressing health inequalities (9 times 6/210 (3%) articles) and that an incentive scheme would introduce a more positive way of relating to patients, it being more honest and improving the doctor–patient relationship (8 times in 8/210 (4%) articles):

> The scheme by Tayside Health Board aims to break the link between low income and high levels of nicotine dependency. The Herald, June 2008.

> It provides a much better and positive way of relating to drug users than sometimes we have done in the past. BBC (website), January 2007.

Against this last argument, although similarly quite rarely expressed (9 times in 9/210 (4%) articles) was the view that an incentive scheme may have a detrimental effect on the doctor–patient relationship, or other relationships between clients and professionals because joint decision making was undermined. As one article put it:

> It undermines the therapeutic alliance the doctor and patient have- something crucial for long-term health care. Medical News Today (website), August 2007.

More commonly raised as a potential harm to participants was the view that incentives could coerce patients into making decisions they may not otherwise have made (33 times in 27/210 (13%) articles), as in:

> The option of being paid to take a drug treatment could unduly influence people's decision making over whether the treatment is right for them. BMJ, October 2009.

Negative health consequences because of such coercive effects, including drug side-effects, were also a concern for some (12 times in 7/210 (3%) articles):

> The mental health charity MIND says that paying people could coerce people into taking drugs that are known to have serious side effects. BMJ, October 2009.

Finally, in terms of harm to participants, on just one occasion, it was mentioned that incentive schemes might stigmatize participants:

> As well as risking further stigma of people suffering from mental illness. The Times, September 2007.

In terms of harm to others or to society as a whole, and going against the view that health incentives schemes could save money, there was a commonly expressed view (111 times in 61/210 (29%) articles) that the money would be better spent elsewhere:

> Is NHS cash going to be channelled into dance lessons and vouchers for fatties when people need cancer drugs and better end-of-life care? Belfast Telegraph, November 2008.

Moral and ethical concerns

In the last quote (and in the example of Peter Lilley earlier), it can be seen that moral concerns about the degree to which recipients deserved to be helped fuelled the objection. In other arguments, moral and ethical concerns were more prominent, and these exclusively contained objections to schemes. First, it was sometimes stated (18 times in 13/210 (6%) articles) that such schemes were plainly unethical, with no further explanation as to why:

> Three quarters of respondents said they had concerns about using financial incentives, most of whom said the practice would be unethical. BBC (website), January 2007.

The view that participants might misuse the rewards or lie to get them was a common objection (39 times in 29/210 (14%) articles), and the idea that they reward the unhealthy or undeserving (30 times in 20/210 (10%) articles) was also expressed. A further moral objection (19 times in 16/210 (8%) articles) was the idea that rewarding healthy behaviour sends out the wrong message because being healthy should be its own reward. Incentive schemes were also said to undermine personal responsibility for health (16 times in (11/210 (5%) articles) and to be given away too easily (10 times in 7/210 (3%) articles). Examples of these arguments are given below:

> Some charities have criticised the lack of measures to ensure the cash is actually spent on healthy food. BBC (website), November 2007.

> Why is this society so hell bent on rewarding the least deserving? Aberdeen Evening Express, January 2007.

> Staying healthy should be enough of an incentive for people to come in for testing, they shouldn't need to be bribed by the opportunity to win high-end electrical goods. Milton Keynes Citizen, July 2009.

> What is this great country coming to? Free gifts and handouts for junkies and failed asylum seekers. The Sun, July 2007.

Related to these moral objections was the view that such schemes represented another excess of the 'nanny state', the government's aim to right every wrong (12 times in 8/210 (4%) articles) and the view that a universal benefit (like the Health in Pregnancy grant) was unfair in not targeting only people in need (4 times in 2/210 (1%) articles):

> And why can't they admit that this is absolutely none of their business anyway? That's what never ceases to astonish me about this Government: its unshakable belief in both the duty and the power of the state to right everything that's wrong with our lives. Daily Mail, January 2008.

> Why wasn't it aimed at those women more in need, rather than being given to everyone, irrespective of their income? The Times, April 2009.

Somewhat related to moral objections, but also identifying an outcome that might arise if people lacked the capacity to resist being incentivized to behave in health-damaging ways, was the view (18 times in 14/210 (7%) articles) that schemes provided perverse incentives:

> Upgrade from being merely chubby to Rubenesque and the Government will help out. Break the scales and take up two bus seats and ministers will subsidise your fare. Consume four pizzas a day and they will pay for your gym membership; force-feed your children Haribos and ministers may cough up for their after-school carrot sticks. The Times, November 2008.

Other arguments

A variety of other arguments, both for and against, were identified. Firstly, there was the simple view that the schemes were either praiseworthy or to be regarded critically, without any reasoning given. Praise of this sort was given rarely (2 times in (2/210 (1%) articles), whereas criticism of this sort was offered more often (21 times in 21/210 (10%) articles). Examples include:

> Others feel it's an effective way to cut the problem. The Mirror, July 2009.

> Drug workers described the proposals as 'ridiculous'. Daily Mail, January 2007.

The views that incentive schemes were new (19 times in 15/210 (9%) articles) and not new (10 times in 10/210 (5%) articles) were both put forward as arguments for supporting schemes:

> Here at NHS Rotherham we want to be at the forefront of groundbreaking schemes which help encourage mums to quit. BBC (website), May 2009.

> Professor Priebe argued that financial incentives to influence healthy behaviour already existed, such as higher taxes on cigarettes and alcohol. BMJ, October 2009.

Finally, there was the view (4 times in 3/210 (1%) articles) that such schemes were an exercise for those in charge of things to provide evidence of activity, but with potentially system-damaging results:

> We would be concerned if incentives were used by poor-performing treatment services to mask problems and hit government targets. BBC (website), January 2008.

Sources

Table 3 shows how sources divided in their support for health incentive schemes. The majority (61%) of source quotes were favourable towards incentive schemes. Government spokespeople, public health representatives, services allied to medicine, academics and doctors were all prominent in coverage and largely favourable towards incentive schemes. These sources were sometimes supported in stories by interviews with participants in schemes speaking positively about their benefits or those representing business interests.

Table 3 Support for schemes by different sources (number of times a source was quoted)

Source type	Unfavourable	Mixed or neutral	Favourable	All
Government*	0	9	50	59
Public health†	2	3	40	45
Services allied‡	3	6	28	37
Charities§	14	7	18	39
Academics	1	9	17	27
Doctors	6	1	15	22
Participants	0	1	12	13
Business	0	0	11	11
Other lay person	4	4	6	14
Opposition politicians	18	6	2	26
Lobbyists¶	15	1	0	17
'Critics'**	8	0	0	8
Think tank††	2	3	0	5
Celebrity	1	0	0	1
Total	75 (23%)	50 (15%)	199 (61%)	324 (100%)

*Includes politicians and civil servants.
†Public health experts and NHS managers.
‡For example, the Royal College of Midwives.
§For example, the National Childbirth Trust.
¶For example, the Taxpayers Alliance.
**Unnamed people referred to as 'critics of the scheme'.
††For example, the New Economics Foundation.

Those representing charities were somewhat prominent but were more evenly split between support and opposition. Opposition politicians and lobbyists from such organizations as the Taxpayer's Alliance were largely critical of the schemes.

Discussion

This analysis shows that UK media coverage of incentive schemes has been more often favourable than unfavourable, although most articles reported a mix of views. For some issues, arguments and counter arguments were made in the overall coverage, most notably in relation to effectiveness, with schemes at various points being described as effective, ineffective or harmful. On other issues, arguments were put forward on one side, which were not contested by the other side. For example, supporters of schemes commonly put forward arguments that stress the seriousness and recalcitrant nature of the health problems addressed by schemes. No evidence was found that critics of schemes disagreed with these perceptions.

By contrast, the moral and ethical objections to schemes, which clearly drive a considerable amount of the opposition to incentives, many of which have been identified and debated in discussions of incentives in academic publications,[45–48] were not significantly contested in media coverage by supporters of schemes. A significant strand of moral objection starts from the view that participants in schemes are undeserving and that providing incentives could be interpreted as rewards for bad behaviour, rather than motivation for good behaviour. In addition, the response to individual schemes was somewhat related to the perceived moral standing of the proposed participants in the different incentive schemes.

Popular stereotypes about those considered deserving or undeserving of help came into play here. For example, the reporting of a scheme to provide cash grants to pregnant women was particularly favourable. By contrast, schemes aimed at overweight people or illegal drug users had a more critical reception, such people being regarded as having behaved irresponsibly in bringing about their health problems. Smoking cessation schemes received a mixed press, again reflecting a degree of moral opprobrium directed at those who have acquired this habit. Attempts to counteract the view that participants in schemes were morally blameworthy and therefore did not deserve help were not made by proponents of schemes, unless one regards the justification that some groups, such as illicit drug users, are otherwise difficult to influence by other means is some kind of recognition of a special moral status.

Underlying these popular judgements about the moral dimensions of incentive schemes are deep-rooted reactions to the fact that incentives appear to involve a taboo trade-off. Fiske and Tetlock[49] point out that moral outrage of this sort is a common reaction when such proposed trade-offs violate the integrity of elementary models that people use to think about their social relationships. In this case, incentives may represent an attempt to put a price on something that many feel ought to be priceless. Propositions to do such things as measure the monetary worth of one's children, loyalty to one's country, or acts of friendship, evoke similarly condemnatory responses. Kahan[50] too has noted the influence of such 'cultural cognition' on judgements people make about scientific and policy issues.

It seems that if proponents of schemes are going to successfully overcome objections that circulate in popular media, then moral arguments about who is deserving of help may need to be addressed more explicitly than hitherto. This will inevitably involve arguing for a special moral status for some groups, on the grounds of their relative lack of competence to behave in the way the moral majority prefer. This itself, of course, could contribute to the stigma that such groups already attract.[15]

The fact that most coverage was favourable partly reflects the fact that it was often stimulated by the announcement of government initiatives in which government spokespeople and their allies in the NHS and public health promoted the virtues of the schemes concerned.

The nature of this story is such that journalists are likely to have relied heavily on press releases from those announcing the launch of schemes. This ensures prominence of the views of sources favourable to the schemes, such as government and health service spokespeople. This should not be allowed to obscure the fact that objections to schemes were considerable and took the particular form we have described.

Given that media present significant opposition to incentive schemes, this study suggests that those proposing incentive schemes will more easily gain public support if evidence is presented for effectiveness and cost-effectiveness, if it exists. The latter may be particularly important in influencing the views of those who consider participants in schemes as undeserving: learning that it reduces the tax burden may be more persuasive than learning the scheme enables those with low self-regulatory capacity to change their behaviour.

Conflict of interest

None.

Sources of funding

This study was completed under the auspices of the Centre for the Study of Incentives in Health, funded by a strategic award from the Biomedical Ethics Programme of the Wellcome Trust (PI Marteau: 086031/Z/08/Z).

Acknowledgement

The Wellcome Trust funds Rebecca Brown's PhD studentship as part of the strategic award funding the Centre for the Study of Incentives in Health.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Search terms.

Appendix S2. Arguments favourable and unfavourable towards incentive schemes, with number of times and in number of articles the argument occurred.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

References

1 Jochelson K. Nanny or steward? The role of government in public health. *Public Health*, 2006; **120:** 1149–1155.
2 Chaloupka F, Grossman M, Saffer H. The effects of price on alcohol consumption and alcohol-related problems. *Alcohol Research & Health*, 2002; **26:** 22–34.
3 Department of Health. *Healthy Weight, Healthy Lives: A Cross Government Strategy for England.* London: HM Government, 2008.
4 Department of Health. *Healthy Weight, Health Lives One Year On: A Cross Government Strategy for England.* London: HM Government, 2009.
5 H.M. Government. *Health and Social Care Act. (c.14) Health in Pregnancy Grant. Part 4: 131-8.* London: HM Government, 2008.
6 NICE. Drug misuse: psychosocial interventions. National Clinical Practice Guideline Number 51. National Collaborating Centre for Mental Health. Commissioned by the National Institute for Health and Clinical Excellence, 2008.
7 H.M. Government. *Drugs: Protecting Families and Communities.* The 2008 drug strategy. London: HM Government, 2008.
8 Classen D, Fakhoury W, Ford R, Priebe S. Money for medication: financial incentives to improve medication adherence in assertive outreach. *Psychiatric Bulletin*, 2007; **31:** 4–7.
9 Priebe S, Burton A, Ashby D *et al.* Financial incentives to improve adherence to anti-psychotic maintenance medication in non-adherent patients - a cluster randomised controlled trial (FIAT). *BMC Psychiatry*, 2009; **9:** 61.
10 NHS Brighton and Hove. Give it a break stop smoking challenge. Press release, 2006. Available at: http://www.brightonandhovepct.nhs.uk/about/news/2006/giveitabreak.asp, accessed 28 April 2010.
11 Christie B. Scottish NHS offers cash to get smokers to quit. *BMJ*, 2009; **338:** b1306.
12 NHS Great Yarmouth and Waveney. Five pounds up for grabs to get tested. 2009. Available at: http://

www.gywpct.nhs.uk/news_item.asp?fldID=274, accessed 28 April 2010.
13 NHS Haringey. Wee for a wii competition. Press release, 2010. Available at: http://www.haringey.nhs.uk/communications/press_room/press_releases/2010/wee_for_a_wii_competition.shtm, accessed 28 April 2010].
14 Szmukler G. Financial incentives for patients in the treatment of incentives. *Journal of Medical Ethics*, 2009; **35:** 224–228.
15 Popay J. Should disadvantaged people be paid to take care of their health? No. *BMJ*, 2008; **337:** a594.
16 Priebe S, Sinclair J, Burton A *et al.* Acceptability of offering financial incentives to achieve medication adherence in patients with severe mental illness: a focus group study. *Journal of Medical Ethics*, 2010; **36:** 463–468.
17 Shaw J. Is it acceptable for people to be paid to adhere to medication? No. *BMJ*, 2007; **335:** 233.
18 Cahill K, Perera R Competitions and incentives for smoking cessation. Cochrane Database of Systematic Reviews 2008, Issue 3. Art. No.: CD004307. doi: 10.1002/14651858.CD004307.pub3.
19 Johnston M, Sniehotta F. Financial incentives to change patient behaviour. *Journal of Health Services Research and Policy*, 2010; **15:** 131–132.
20 Kavanagh J, Oakley A, Harden A, Trouton A, Powell C. Are incentive schemes effective in changing young people's behaviour? A systematic review. *Health Education Journal*, 2011; **70:** 192–205.
21 Marteau TM, Ashcroft RE, Oliver A. Using financial incentives to achieve healthy behaviour. *BMJ*, 2009; **338:** b1415.
22 Lagarde M, Haines A, Palmer N. Conditional cash transfers for improving uptake of health interventions in low and middle-income countries: a systematic review. *JAMA*, 2007; **298:** 1900–1910.
23 Lussier JP, Heil SH, Mongeon JA, Badger GJ, Higgins ST. A meta-analysis of voucher-based reinforcement therapy for substance use disorders. *Addiction*, 2006; **101:** 192–203.
24 Charness G, Gneezy U. Incentives to exercise. *Econometrica*, 2009; **77:** 909–931.
25 Volpp KG, Troxel AB, Pauly MV, Glick HA, Puig A, Asch DA. A randomized, controlled trial of financial incentives for smoking cessation. *New England Journal of Medicine*, 2009; **360:** 699–709.
26 Department of Health. *Choice of Hospital: Guidance for PCTs, NHS Trusts and SHAs on Offering Patients Choice of Where They are Treated*. London: Department of Health Publications, 2003.
27 Department of Health. *Equity and Excellence: liberating the NHS*. Norwich: The Stationary Office, 2010.
28 LeGrand J. *The Other Invisible Hand: Delivering Public Services through Choice and Competition*. Princeton: Princeton University Press, 2007.
29 Oliver A. Reflections on the development of health inequalities policy in England. *Health Care Analysis*, 2010; **18:** 402–420.
30 Faden R, Beauchamp T. *A History and Theory of Informed Consent*. Oxford: Oxford University Press, 1986.
31 O'Neill O. *Autonomy and Trust in Bioethics*. Cambridge: Cambridge University Press, 2002.
32 The Stationary Office. *HM Government Healthy Lives, Healthy People: Our Strategy for Public Health in England*. Norwich: The Stationary Office, 2010.
33 Dawson A, Verweij M (eds) *Ethics, Prevention, and Public Health*. Oxford: Clarendon Press, 2007.
34 Nuffield Council on Bioethics. *Public Health: The Ethical Issues*. London: The Nuffield Council on Bioethics, 2007.
35 Rose N. *Powers of Freedom: Reframing Political Thought*. Cambridge: Cambridge University Press, 1999.
36 Thaler RH, Sunstein CR. *Nudge: Improving Decisions about Health, Wealth and Happiness*. New Haven: Yale University Press, 2008.
37 Marteau TM, Ogilvie D, Roland M, Suhrcke M, Kelly MP. Judging nudging: can 'nudging' improve population health? *BMJ*, 2011; **342:** d228.
38 Wise J. NICE Citizens Council debates incentives for healthy behaviour. *BMJ*, 2010; **340:** c2747.
39 Seale C. *Media and Health*. London: Sage, 2002.
40 McQuail D. *McQuail*'s Mass Communication Theory, 5th edn. Newbury Park: Sage, 2005.
41 Conrad P. Use of expertise: sources, quotes and voice in the reporting of genetics in the news. *Public Understanding of Science*, 1999; **8:** 285–302.
42 Miller D, Kitzinger J, Williams K. *The Circuit of Mass Communication: Media Strategies, Representation and Audience Reception in the AIDS Crisis*. London: Sage, 1998.
43 Introna L, Nissenbaum H. Shaping the web: why the politics of search engines matters. *The Information Society*, 2000; **16:** 1–17.
44 Seale C. New directions for critical internet health studies: representing cancer experience on the web. *Sociology of Health and Illness*, 2005; **27:** 515–540.
45 Long JA, Helweg-Larsen M, Volpp KG. Patient opinions regarding "pay for performance for patients." *Journal of General Internal Medicine*, 2008; **23:** 1647–1652.
46 Claassen D. Financial incentives for antipsychotic depot medication: ethical issues. *Journal of Medical Ethics*, 2007; **33:** 189–193.
47 Cookson R. Should disadvantaged people be paid to take care of their health? Yes. *BMJ*, 2008; **337:** a589.
48 Marteau TM, Oliver A, Ashcroft RE. Changing behaviour through state intervention. *BMJ*, 2008; **337:** a2543.

49 Fiske AP, Tetlock PE. Taboo trade-offs: reactions to transactions that transgress the spheres of justice. *Political Psychology*, 1997; **18:** 255–297.

50 Kahan D. Fixing the communications failure. *Nature*, 2010; **463:** 296–297.

Appendix 1: Search terms (Supporting Information)

Case study 1 – *Weight control*

Nhs	brib!	Cash	Fat
Diet!	obes!	Overweight	Incentive!
Stick	Weight	weight wins	reward!

e.g. nhs AND cash AND diet! AND obes!

Case study 2 – *Health in pregnancy*

190	pregnan!	Grant	payment
health!	Mum	Cash	

e.g. mum AND 190 AND cash

Case study 3 – *Antipsychotic medication adherence*

Cash	psycho!	schizophreni!	incentiv!
Drug	Claassen	Medication	treatment

e.g. cash AND drug AND psycho!

Case study 4 – *Illicit drug use reduction*

addict!	reward!	test!	government
Drug	Voucher	Negative	

e.g. addict! AND drug AND reward! AND voucher

Case study 5 – *Smoking cessation*

smok!	Reward	quit!
incentiv!	pregnan!	

e.g. smok! AND pregnan! AND quit! AND incentive!

Case study 6 – *STI testing*

Chlamydia	Wee	Wii	Prize	test
Sexual	Infection	transmitted	reward	

e.g. Chlamydia AND wee AND wii

Case study 7 – *General*

Government	Reward	health
Conditional	Cash	transfer

e.g. conditional AND cash AND transfer

Appendix 2: Arguments favourable and unfavourable towards incentive schemes, with number of times and in number of articles the argument occurred (Supporting Information)

Arguments in favour of incentives

- **POSITIVE EVIDENCE (130 times; 82 articles)**

Theory or empirical evidence which suggest the effectiveness of incentives, also personal stories of success and 'experts' who believe the schemes will work.
'Such schemes have been used in the US with research showing participants stay drug-free for twice as long as those not taking part in incentive schemes' BBC (website), June 2008.

- **BEST INTERESTS (124 times; 94 articles)**

Incentives help people to do what is in their best interests, including mentions of the health benefits to the individual.
'The one-off payment is intended to help pregnant mums stay fit and healthy in the run-up to the birth' South Wales Echo, April 2009.

- **PREVALENCE (117 times; 78 articles)**

The prevalence of the health issue is cited.
'The UK is in the grip of an obesity epidemic with a quarter of all adults and one in five children now obese' Daily Mail, January 2008.

- **MONEY SAVING (79 times; 54 articles)**

Incentive schemes will save money in the long run, they are cost effective.
'The trust says the plan could save the NHS thousands of pounds by preventing conditions such as heart disease' BBC (website), October 2008.

- **DISEASE (63 times; 47 articles)**

Disease related to the health behaviour is mentioned, or more general health problems relating to the health behaviour are mentioned.
'Low birth weight babies and premature deliveries are much more common in mothers who smoke' BBC (website), February 2008.

- **FINANCIAL BURDEN (50 times; 34 articles)**

Incentives will help relieve some of the financial pressures on the recipient.
'The financial stress associated with impending motherhood can now be eased by a £190 cash boost for women due to give birth on or after April 6' Belfast Telegraph, January 2009.

- **LAST RESORT (45 times; 32 articles)**

Ideal as a last resort, for the hard to engage, disadvantaged and vulnerable.
'Financial incentives might be a treatment option for a high-risk group of non-adherent patients with whom all other interventions to achieve adherence have failed' BBC (website), January 2007.

- **BENEFITS SOCIETY (26 times; 19 articles)**

Benefits to the rest of society, not just the participant in the incentive scheme.
'If it works to keep people in treatment there would be considerable benefits to the public' The Daily Telegraph, January 2007.

- **INNOVATIVE (19 times; 15 articles)**

Incentive schemes are a new and innovative idea.
'This is a very important aspect of Scotland's battle for better public health and it is worth being brave and innovative' The Herald, June 2006

- **OTHER ORGANISATIONS (11 times; 9 articles)**

Other organisations are looking to follow suit with health incentive schemes, or other organisations are advocating the use of incentive schemes.
'The Welsh Office said it was interested in pursuing a similar scheme' BBC (www), June 2008.

- **NOT NOVEL (10 times; 10 articles)**

Incentives are not a new and controversial idea, but an already established practice.
'Professor Priebe argued that financial incentives to influence healthy behaviour already existed, such as higher taxes on cigarettes and alcohol' BMJ, October 2009.

- **IMPROVES INEQUALITY (9 times; 6 articles)**

Incentives may help to close the gap in health inequalities in our society.
'The scheme by Tayside Health Board aims to break the link between low income and high levels of nicotine dependency' The Herald, June 2008.

- **POSITIVE (8 times; 8 articles)**

A more positive way of relating to patients, it is more honest and can improve the doctor-patient relationship.
'It provides a much better and positive way of relating to drug users than sometimes we have done in the past' BBC (website), January 2007.

- **NO HARM (3 times; 2 articles)**

No harm can come to people involved in incentive schemes.
'There is no harm intended or caused- the service users can revoke the offer at any time' The Metro, January 2007.

- **PRAISE (2 times; 2 articles)**

Praise for incentives without any reasoning given.
'Others feel it's an effective way to cut the problem' The Mirror, July 2009.

Arguments unfavourable towards incentives

- **BETTER SPENT (111 times; 61 articles)**

The money being spent on incentive schemes could be better spent elsewhere, includes mentions of the cost of the schemes and references to taxpayers' money.
'Is NHS cash going to be channelled into dance lessons and vouchers for fatties when people need cancer drugs and better end-of-life care?' Belfast Telegraph, November 2008.

- **NOT LONG TERM (46 times; 24 articles)**

The scheme won't offer any long term solutions, it does not address the roots of the problem.
'This is no kind of long-term solution- a temporary financial incentive won't stop people putting the weight back on once they have got the cash' Daily Mail, January 2009.

- **MISUSE (39 times; 29 articles)**

People may abuse the system, fraud may be a problem. Or, the rewards may be misused by the recipient.
'But they are expected to be free to spend the cash as they see fit, even on unhealthy products like cigarettes and alcohol' Birmingham Post, September 2007.

- **BRIBE (33 times; 27 articles)**

Incentives coerce patients into making decisions they may not have otherwise made.
'The option of being paid to take a drug treatment could unduly influence people's decision making over whether the treatment is right for them' BMJ, October 2009.

- **REWARDS UNHEALTHY (30 times; 20 articles)**

The unhealthy or undeserving are rewarded whilst the healthy miss out.
'Why is this society so hell bent on rewarding the least deserving?' Aberdeen Evening Express, January 2007.

- **NEGATIVE EVIDENCE (29 times; 19 articles)**

A lack of evidence to support the effectiveness of the incentive scheme, or evidence that shows the schemes to be ineffective.
'There is however little research that shows that a financial incentive, combined with nutritional advice, is enough to persuade mothers from the most deprived areas to change their lifestyle' News-Medical (website), September 2007.

- **CRITICISM (21 times; 18 articles)**

An incentive scheme is criticised without any reasoning provided.
'Drug workers described the proposals as 'ridiculous'' Daily Mail, January 2007.

- **WRONG MESSAGE (19 times; 16 articles)**

Rewarding healthy behaviour sends out the wrong message. Being healthy should be its own reward, incentives shouldn't be the motivation. Rewards undermine intrinsic motivation.
'Staying healthy should be enough of an incentive for people to come in for testing, they shouldn't need to be bribed by the opportunity to win high-end electrical goods' Milton Keynes Citizen, July 2009.

- **ENCOURAGE (18 times; 14 articles)**

A perverse effect of incentive schemes may be that they encourage people to adopt the unwanted behaviour.
'We have to be careful we don't appear to be encouraging people to take up smoking in order to reward them for giving up' The Metro, February 2009.

- **UNETHICAL (18 times; 13 articles)**

Where the ethics of the scheme are raised without any further explanation.
'Three quarters of respondents said they had concerns about using financial incentives, most of whom said the practice would be unethical' BBC (website), January 2007.

- **RESPONSIBILITY (16 times; 11 articles)**

People should take responsibility for their own health, it is within their own control.
'I question very much whether the NHS should be directing its resources to weight loss-something within people's control' Daily Mail, January 2009.

- **NANNY STATE (12 times; 8 articles)**

Incentives cited as another excess of the nanny state, the government's aim to right every wrong.
'Time and time again this molly-coddling government thinks it can right what is fundamentally wrong with our lives' Sunday Mirror, January 2008.

- **GIVEAWAY (10 times; 7 articles)**

Rewards are handed out too easily.
'What is this great country coming to? Free gifts and handouts for junkies and failed asylum seekers' The Sun, July 2007

- **HEALTH HARM (12 times; 7 articles)**

The behaviour being encouraged may have negative health consequences on the participant, including drug side-effects.
'The mental health charity MIND says that paying people could coerce people into taking drugs that are known to have serious side effects' BMJ, October 2009.

- **DOCTOR-PATIENT RELATIONSHIP (9 times; 9 articles)**

Incentives may have a detrimental effect on the doctor-patient relationship, or other relationships between clients and professionals. The relationship should be about joint decision making.

'It undermines the therapeutic alliance the doctor and patient have- something crucial for long-term health care' Medical News Today (website), August 2007.

- **CASH WON'T WORK (6 times; 4 articles)**

Money won't cause a behaviour change in people.
'If it were just a question of money, they would have stopped eating years ago. After all, if they ate less, they would be richer' The Daily Telegraph, January 2008.

- **POLITICAL (4 times; 3 articles)**

Just another political exercise.
'We would be concerned if incentives were used by poor-performing treatment services to mask problems and hit government targets' BBC (website), January 2008.

- **UNIVERSAL (4 times; 2 articles)**

The drawbacks of universal benefits as opposed to targeted ones.
'Why wasn't it aimed at those women more in need, rather than being given to everyone, irrespective of their income?' The Times, April 2009.

- **STIGMA (1 time; 1 article)**

Negative effects on the individual, including stigmatisation, loss of autonomy and loss of dignity.
'*As well as risking further stigma of people suffering from mental illness'* The Times, September 2007.

BIBLIOGRAPHY

Ackerman, Bruce (1980) *Social Justice in the Liberal State*, London: Yale University Press

Anderson, Elizabeth S. (1990) 'Is Women's Labor a Commodity?', *Philosophy and Public Affairs*, Vol. 19, No. 1, pp. 71-92

Anderson, Elizabeth S. (1993) *Value in Ethics and Economics*, Cambridge MA: Harvard University Press

Appiah, Kwame Anthony (2008) *Experiments in Ethics*, Cambridge MA: Harvard University Press

Aristotle (1984) *The Complete Works of Aristotle, Oxford Revised Translation*, Barnes [ed.], Princeton: Princeton University Press

Arneson, Richard J. (1990) 'Primary Goods Reconsidered', *Noûs*, Vol. 24, No. 3, pp. 429-454

Arneson, Richard J. (1992) 'Commodification and Commercial Surrogacy', *Philosophy and Public Affairs*, Vol. 21, No. 2, pp. 132-164

Arneson, Richard J. (2000) 'Perfectionism and Politics', *Ethics*, Vol. 111, No. 1, pp. 37-63

Ashcroft, Richard E. (2011) 'Personal Financial Incentives in Health Promotion: Where do they Fit in an Ethic of Autonomy?', *Health Expectations*, Vol. 14, No. 2, pp. 191-200

Bains, Namrata; Pickett, William; Maundry, Brian; and Mecredy, Darlene (2000) 'Predictors of Smoking Cessation in an Incentive-Based

Community Intervention', *Chronic Diseases in Canada*, Vol. 21, No. 2, pp. 54-61

Ballard, Paul, and Radley, Andrew (2009) 'Give it Up for Baby - A Smoking Cessation Intervention for Pregnant Women in Scotland', *Cases in Public Health Communication and Marketing*, Vol. 3, pp. 147-160

Bandura, Albert (1978) 'Self-Efficacy: Toward a Unifying Theory of Behavioral Change', *Advances in Behaviour Research and Therapy*, Vol. 1, No. 4, pp. 139-161

BBC (2010) 'Charity Offers UK Drug Addicts £200 to Be Sterilised', [Online], *BBC News*, Available at: http://www.bbc.co.uk/news/uk-england-london-11545519, [Accessed 1st May 2013]

Beauchamp, Tom L., and Childress, James F. (2001) *Principles of Biomedical Ethics*, Oxford: Oxford University Press

Bechara, Antoine (2005) 'Decision Making, Impulse Control and Loss of Willpower to Resist Drugs: A Neurocognitive Perspective', *Nature Neuroscience*, Vol. 8, No. 11, pp. 1458-1463

Becker, Gary S., and Elias, Julio J. (2007) 'Introducing Incentives in the Market for Live and Cadaveric Organ Donations', *The Journal of Economic Perspectives*, Vol. 21, No. 3, pp. 3-24

Berlin, Isaiah (1969) 'Two Concepts of Liberty', in *Four Essays on Liberty*, Berlin, Oxford: Oxford University Press

Berman, Mitchell (2002) 'The Normative Functions of Coercion Claims', *Legal Theory*, Vol. 8, No. 1, pp. 45-89

Bernier, Michel, and Avard, Jacqueline (1986) 'Self-Efficacy, Outcome, and Attrition in a Weight-Reduction Program', *Cognitive Therapy and Research*, Vol. 10, No. 3, pp. 319-338

Blumenthal-Barby, J. S., and Burroughs, Hadley (2012) 'Seeking Better Health Care Outcomes: The Ethics of Using the "Nudge"', *The American Journal of Bioethics*, Vol. 12, No. 2, pp. 1-10

Bovens, Luc (2008) 'The Ethics of Nudge', in *Preference Change: Approaches from Philosophy, Economics and Psychology*, Grüne-Yanoff and Hansson (eds.), New York: Springer

Brown, Alexander (2009) *Personal Responsibility: Why it Matters*, London: Continuum

Burnyeat, M. F. (1980) 'Aristotle on Learning to be Good', pp. 69-92 in *Essay's on Aristotle's Ethics*, Rorty [ed.], London: University of California Press

Burton, Alexandra; Marougka, Stamatina; and Priebe, Stefan (2010) 'Do Financial Incentives Increase Treatment Adherence in People With Severe Mental Illness? A Systematic Review', *Epidemiologia E Psichiatria Sociale*, Vol. 19, No. 3, pp. 233-242

Buyx, Alena M. (2008) 'Personal Responsibility for Health as a Rationing Criterion: Why We Don't Like it and Why Maybe We Should', *Journal of Medical Ethics*, Vol. 84, pp. 871-874

Cahill, K., and Perera, R. (2009) 'Competitions and Incentives for Smoking Cessation', *Cochrane Database of Systematic Reviews*, Issue 3, pp. 1-37

Cambstakeatest (2009) 'Wee for a Wii - Take a Chlamydia Test for a Chance to Win!', [Online], *Peterborough and Cambridgeshire NHS Trusts Chlamydia Screening Programme*, Available at: http://www.cambstakeatest.com/wii.html, [Accessed 1st May 2013]

Cameron, Judy, and Pierce, W. David (1994) 'Reinforcement, Reward, and Intrinsic Motivation: A Meta-Analysis', *Review of Educational Research*, Vol. 64, No. 3, pp. 363-423

Cameron, Judy; Banko, K, and Pierce, W. David (2001) 'Pervasive Negative Effects of Rewards on Intrinsic Motivation: the Myth Continues', *The Behavior Analyst*, Vol. 24, pp. 1-44

Cave, Eric M. (2007) 'What's Wrong with Motive Manipulation', *Ethical Theory and Moral Practice*, Vol. 10, No. 2, pp. 129-144

Chan, Joseph (2000) 'Legitimacy, Unanimity, and Perfectionism', *Philosophy and Public Affairs*, Vol. 29, No. 1, pp. 5-42

Chapple, A.; Ziebland, S.; and McPherson, A. (2004) 'Stigma, Shame, and Blame Experienced By Patients With Lung Cancer: Qualitative Study', *British Medical Journal*, Vol. 328, No. 7454, pp. 1470-1475

Coggon, John (2012) *What Makes Health Public?*, Cambridge: Cambridge University Press

Cohen, G. A. (1989) 'On the Currency of Egalitarian Justice', *Ethics*, Vol. 99, No. 4, pp. 906-944

Cohen, G. A. (2001) *If You're an Egalitarian, How Come You're So Rich?*, London: Harvard University Press

Coupland, Carol; Harcourt, Sally; Vinogradova, Yana; Smith, Gillian; Joseph, Caro; Pringle, Mike; and Hippisley-Cox, Julia (2007) 'Inequalities in Uptake of Influenza Vaccine By Deprivation and Risk Group: Time Trends Analysis', *Vaccine*, Vol. 25, No. 42, pp. 7363-7371

Cribb, Alan, and Owens, John (2010) 'Whatever Suits You: Unpicking Personalization for the NHS', *Journal of Evaluation in Clinical Practice*, Vol. 16, No. 2, pp. 310-314

Cummins, S.; McKay L; and Macintyre S. (2005) 'McDonald's Restaurants and Neighbourhood Deprivation in Scotland and England', *American Journal of Preventive Medicine*, Vol. 29, No. 4, pp. 308-310

Danjoux, Ilan (2007) 'Revealing the Public: A Media Analysis Approach to Public Opinion Research', [Online], *Centre for International Politics Working Paper Series*, No 25, University of Manchester, Available at: http://www.socialsciences.manchester.ac.uk/disciplines/politics/about/themes/cip/publications/documents/Danjoux-CIP-2007.pdf, [Accessed 21st September 2012]

Davidson, Donald (2001) *Essays on Actions and Events: Second Edition*, Oxford: Clarendon Press

Dawson, Angus J. (2010) 'The Future of Bioethics: Three Dogmas and a Cup of Hemlock', *Bioethics*, Vol. 24, No. 5, pp. 218-225

Deci, Edward L.; Koestner, Richard; and Ryan, Richard M. (1999) 'A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation', *Psychological Bulletin*, Vol. 125, No. 6, pp. 627-668

Deci, Edward L., and Ryan, Richard M. (1985) *Intrinsic Motivation and Self-Determination in Human Behavior*, New York: Plenum Press

Department for Communities and Local Government (2011) *English Indices of Deprivation 2010: Guidance Document*, London: Department for Communities and Local Government

Dickert, Neal, and Grady, Christine (1999) 'What's the Price of a Research Subject? Approaches to Payment for Research Participation', *New England Journal of Medicine*, Vol. 341, No. 3, pp. 198-203

Dolan, Paul; Hallsworth, M.; Halpern, David; King, Dominic; Metcalfe, R.; and Vlaev, I. (2012) 'Influencing Behaviour: The Mindspace Way', *Journal of Economic Psychology*, Vol. 33, pp. 264-277

Dolan, Paul, and Kahneman, Daniel (2007) 'Interpretations of Utility and Their Implications for the Valuation of Health', *The Economic Journal*, Vol. 118, No. 525, pp. 215-234

Doris, John M. (1998) 'Persons, Situations, and Virtue Ethics', *Noûs*, Vol. 31, No. 4, pp. 504-530

Doris, John M. (2002) *Lack of Character: Personality and Moral Behavior*, Cambridge: Cambridge University Press

Doubleday, Richard, and Wilsdon, James (eds.) (2013) *Future Directions for Scientific Advice in Whitehall*, London: Centre for Science and Policy

Drummond, Michael; Brixner, Diana; Gold, Marthe; Kind, Paul; McGuire, Alistair; and Nord, Erik (2009) 'Toward a Consensus on the QALY', *Value in Health*, Vol. 12, Supplement 1, pp. S31-S35

Duxbury, Neil (1996) 'Do Markets Degrade?', *The Modern Law Review*, Vol. 59, No. 3, pp. 331-348

Dworkin, Gerald (1988) *The Theory and Practice of Autonomy*, Cambridge: Cambridge University Press

Dworkin, Ronald (1981a) 'What is Equality? Part 1: Equality of Welfare', *Philosophy and Public Affairs*, Vol. 10, No. 3, pp. 185-246

Dworkin, Ronald (1981b) 'What is Equality? Part 2: Equality of Resources', *Philosophy and Public Affairs*, Vol. 10, No. 4, pp. 283-345

Dworkin, Ronald (1985) *A Matter of Principle*, Oxford: Oxford University Press

Dworkin, Ronald (2000) *Sovereign Virtue: The Theory and Practice of Equality*, London: Harvard University Press

Dworkin, Ronald (2003) 'Equality, Luck and Hierarchy', *Philosophy and Public Affairs*, Vol. 31, No. 2, pp. 190-198

Faden, Ruth R., and Faden, Alan I. (1978) 'The Ethics of Health Education as Public Health Policy', *Health Education and Behavior*, Vol. 6, No. 2, pp. 180-197

Feinberg, Joel (1970) *Doing and Deserving*, Princeton: Princeton University Press

Feinberg, Joel (1984) *The Moral Limits of the Criminal Law, Vol. 1: Harm to Others*, Oxford: Oxford University Press

Feinberg, Joel (1986) *The Moral Limits of the Criminal Law, Vol. 3: Harm to Self*, Oxford: Oxford University Press

Frankfurt, Harry (1988/2007) 'Coercion and Moral Responsibility', pp. 26-46 in *The Importance of What We Care About*, Frankfurt, Cambridge: Cambridge University Press

Frey, Bruno, and Jegen, Reto (2001) 'Motivation Crowding Theory', *Journal of Economic Surveys*, Vol. 15, No. 5, pp. 549-611

Gerbner, George (1998) 'Cultural Analysis: An Overview', *Mass Communication and Society*, Vol. 1, No. 3, pp. 175-194

Gigerenzer, Gerd, and Selten, Reinhard (eds.) (2002) *Bounded Rationality: The Adaptive Toolbox*, Cambridge MA: MIT Press

Gillon, Raanon (2003) 'Ethics Needs Principles, Four Can Encompass the Rest, and Respect for Autonomy Should Be "First Among Equals"', *Journal of Medical Ethics*, Vol. 29, No. 5, pp. 307-312

Giuffrida, Antonio, and Torgerson, David J. (1997) 'Should We Pay the Patient? Review of Financial Incentives to Enhance Patient Compliance', *British Medical Journal*, Vol. 315, No. 7110, pp. 703-707

Gneezy, Uri, and Rustichini, Aldo (1000) 'A Fine is a Price', *Journal of Legal Studies*, Vol. 29, pp. 1-17

Gonzalez, Richard, and Wu, George (1999) 'On the Shape of the Probability Weighting Function', *Cognitive Psychology*, Vol. 38, No. 1, pp. 129-166

Goodyear-Smith, Felicity, and Buetow, Stephen (2001) 'Power Issues in the Doctor-Patient Relationship', *Health Care Analysis*, Vol. 9, No. 4, pp. 449-462

Goold, Susan D. (1998) 'Money and Trust: Relationships Between Patients, Physicians, and Health Plans', *Journal of Health Politics, Policy and Law*, Vol. 23, No. 4, pp. 687-695

Goold, Susan D., and Lipkin, Mack (1999) 'The Doctor-Patient Relationship', *Journal of General Internal Medicine*, Vol. 14, No. S1, pp. 26-33

Gorr, Michael (1986) 'Toward a Theory of Coercion', *Canadian Journal of Philosophy*', Vol. 16, No. 3, pp. 383-406

Grant, Ruth W., and Sugarman, Jeremy (2004) 'Ethics in Human Subjects Research: Do Incentives Matter?', *Journal of Medicine and Philosophy*, Vol. 29, No. 6, pp. 717-738

Griffin, James (1986) *Well-Being: Its Meaning, Measurement and Moral Importance*, Oxford: Oxford University Press

Growing, Nik (1994) 'Real Time Television Coverage of Armed Conflicts and Diplomatic Crises: Does it Pressure or Distort Foreign Policy Decisions?', *The Joan Shorenstein Barone Center on the Press, Politics and Public Policy*, Working Paper, Vol. 94, No. 1

Guyer, Paul (2000) *Kant on Freedom, Law, and Happiness*, Cambridge: Cambridge University Press

HM Government (2010) *Healthy Lives, Health People: Our Strategy for Public Health in England*, London: The Stationary Office

Haidt, Jonathan; Roller, Silvia H.; and Dias, Maria G. (1993) 'Affect, Culture, and Morality, or Is It Wrong to Eat Your Dog?', *Journal of Personality and Social Psychology*, Vol. 65, No. 4, pp. 613-628

Harcombe, Zoe (2009) 'We'll Pay You to Lose Weight', [Online], *Zoe Harcombe*, Available at: http://www.zoeharcombe.com/2009/07/we'll-pay-you-to-lose-weight/, Accessed 30th April 2013

Harman, Gilbert (1999) 'Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error', *Proceedings of the Aristotelian Society*, Vol. 99, pp. 315-331

Harman, Gilbert (2000) 'The Nonexistence of Character Traits', *Proceedings of the Aristotelian Society*, Vol. 100, pp. 223-226

Harris, John (1997) 'The Rationing Debate: Maximising the Health of the Whole Community. The Case Against: What the Principal Objective of the NHS Should Really Be', *British Medical Journal*, Vol. 314, No. 7081, pp. 667-672

Hausman, Daniel M. (2006) 'Valuing Health', *Philosophy and Public Affairs*, Vol. 34, No. 3, pp. 246-274

Hausman, Daniel M., and Welch, Brynn (2010) 'Debate: To Nudge or Not to Nudge', *Journal of Political Philosophy*, Vol. 18, No. 1, pp. 123-136

Haybron, D. (2008) *The Pursuit of Unhappiness*, Oxford: Oxford University Press

Health England (2009) *Incentives for Prevention: Health England Report No. 3*, London: Health England

Herman, Barbara (1981) 'On the Value of Acting from the Motive of Duty', *The Philosophical Review*, Vol. 90, No. 3, pp. 359-382

House of Lords (2011) *Science and Technology Select Committee Second Report of Session 2010-12: Behaviour Change*, London: Stationary Office

Kahneman, Daniel, and Tversky, Amos (1979) 'Prospect Theory: An Analysis of Decision Under Risk', *Econometrica*, Vol. 47, No. 2, pp. 263-291

Kahneman, Daniel; Knetsch, J.; and Thaler, Richard (1991) 'Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias', *Journal of Economic Perspectives*, Vol. 5, No. 1, pp. 193-206

Kahneman, Daniel, and Tversky, Amos (2000) *Choices, Values, and Frames*, Cambridge: Cambridge University Press

Kahneman, Daniel (2012) *Thinking Fast and Slow*, London: Penguin Books

Kamtekar, Rachana (2004) 'Situationism and Virtue Ethics on the Content of Our Character', *Ethics*, Vol. 114, No. 3, pp. 458-491

Kant, Immanuel (1785/1993) *Grounding for the Metaphysics of Morals: On a Supposed Right to Lie because of Philanthropic Concerns*, translated by Ellington, Indianapolis: Hackett

Kivetz, Ran, and Simonson, Itamar (2002) 'Earning the Right to Indulge: Effort as a Determinant of Customer Preferences Toward Frequency Program Rewards', *Journal of Marketing Research*, Vol. 39, No. 2, pp. 155-170

Korsgaard, Christine M. (1996) *Creating the Kingdom of Ends*, Cambridge: Cambridge University Press

Korsgaard, Christine M. (1999) 'Self-Constitution in the Ethics of Plato and Kant', *Journal of Ethics*, Vol. 3, No. 1, pp. 1-29

Kotz, D. and West, R. (2009) 'Explaining the Social Gradient in Smoking Cessation: It's Not in the Trying, but in the Succeeding', *Tobacco Control*, Vol. 18, No. 1, pp. 43-46

Kraut, R. (2007) *What is Good and Why*, Cambridge MA: Harvard University Press

Kupperman, Joel J. (2001) 'The Indispensability of Character', *Philosophy*, Vol. 76, No. 296, pp. 239-250

Lamont, Julian (1994) 'The Concept of Desert in Distributive Justice', *The Philosophical Quarterly*, Vol. 44, No. 174, pp. 45-64

Lagarde, Mylene; Haines, Andy; and Palmer, Natasha (2009) 'The Impact of Conditional Cash Transfers on Health Outcomes and Use of Health Services in Low and Middle Income Countries', *Cochrane Database of Systematic Reviews*, No. 4, pp. 1-50

Larmore, Charles (1987) *Patterns of Moral Complexity*, Cambridge: Cambridge University Press

Lexmond, Jen, and Reeves, Richard (2009) *Building Character*, London: Demos

Lifton, P. D. (1983) 'Measures of Autonomy', *Journal of Personality Assessment*, Vol. 47, No. 5, pp. 514-523

Lussier, Jennifer; Heil, Sarah; Mongeon, Joan; Badger, Gary; and Higgins, Stephen (2006) 'A Meta-Analysis of Voucher-Based Reinforcement Therapy for Substance Use Disorders', *Addiction*, Vol. 101, No. 2, pp. 192-203

Macdonald L.; Cummins S.; and Macintyre S. (2007) 'Neighbourhood Fast Food Environment and Area Deprivation - Substitution or Concentration?', *Appetite*, Vol. 49, No. 1, pp. 251-254

MacIntyre, Alasdair C. (1981) *After Virtue: A Study in Moral Theory*, Indiana: University of Notre Dame Press

MacIntyre, Alasdair C. (1988) *Whose Justice? Which Rationality?*, London: Duckworth Press

Maheswaran, Ravi; Pearson, Tim; Jordan, Hannah; and Black, David (2006) 'Socioeconomic Deprivation, Travel Distance, Location of Service, and Uptake of Breast Cancer Screening in North Derbyshire, UK', *Journal of Epidemiology and Community Health*, Vol. 60, No. 3, pp. 208-212

Mantzari, Eleni; Vogt, Florian; and Marteau, Theresa M. (in submission, personal communication) 'Does Incentivising Pill-Taking 'Crowd-Out' Risk-Information Processing? Evidence from a Web-Based Experiment', *Social Science and Medicine*

Marmot, Michael; Atkinson, Tony; Bell, John; Black, Carol; Broadfoot, Patricia; Cumberlege, Julia; Diamond, Ian; Gilmore, Ian; Ham, Chris; Meacher, Molly; and Mulgan, Geoff (2010) *Fair Society, Healthy Lives: The Marmot Review*, London: The Marmot Review

Marteau, Theresa M.; Kinmonth, Ann-Louise; Thompson, Simon; and Pyke, Stephen (1996) 'The Psychological Impact of Cardiovascular Screening and Intervention in Primary Care: A Problem of False Reassurance?', *British Journal of General Practice*, Vol. 46, No. 411, pp. 577-582

Marteau, Theresa M.; Ashcroft, Richard; and Oliver, Adam (2009a) 'Using Financial Incentives to Achieve Healthy Behaviour', *British Medical Journal*, Vol. 338, No. 7701, pp. 983-985

Marteau, Theresa M.; Oliver, Adam; and Ashcroft, Richard E. (2009b) 'Changing Behaviour Through State Interventions: When Does an Acceptable Nudge Become and Unacceptable Shove?', *British Medical Journal*, Vol. 338, No. 7687, pp. 121-122

Marteau, Theresa M. (2010) 'Changing Behaviour to Improve Population Health', pp. 96-107 in *Health Innovations: More for Less in Healthcare*, Churchill (ed.), London: The Smith Institute

Marteau, Theresa M.; Ogilvie, David; Roland, Martin; and Suhrcke, Marc (2011) 'Judging Nudging: Can Nudging Improve Population Health?', *British Medical Journal*, Vol. 342, No. 3, pp. 263-265

Marteau, Theresa M.; Hollands, Gareth J.; and Fletcher, Paul C. (2012) 'Changing Human Behavior to Prevent Disease: The Importance of Targeting Automatic Processes', *Science*, Vol. 337, No. 6101, pp. 1492-1495

Mason, Andrew (1990) 'Autonomy, Liberalism and State Neutrality', *The Philosophical Quarterly*, Vol. 40, No. 161, pp. 433-452

McCaffery, K.; Wardle, J.; Nadel, M.; and Atkin, W. (2002) 'Socioeconomic Variation in Participation in Colorectal Cancer Screening', *Journal of Medical Screening*, Vol. 9, No. 3, pp. 104-108

McCombs, Maxwell, and Shaw, Donald (1972) 'The Agenda-Setting Function of Mass Media', *Public Opinion Quarterly*, Vol. 36, No. 2, pp. 176-187

Mendelsohn, Harold (1968) 'Which Shall it be: Mass Education or Mass Persuasion for Health', *American Journal of Public Health and the Nation's Health*, Vol. 58, No 1, pp. 131-137

Menzel, Paul; Gold, Marthe R.; Nord, Erik; Pinto-Prades, Jose-Louis; Richardson, Jeff; and Ubel, Peter (1999) 'Toward a Broader View of Values in Cost-Effectiveness Analysis of Health', *The Hastings Center Report*, Vol. 29, No. 3, pp. 7-15

Mill, John Stuart (1859/2006) *On Liberty and the Subjugation of Women*, London: Penguin Books

The Mirror (2009) 'Pay nurses, not these mollycoddled tubbies', [Online], *The Mirror*, Available at: http://daily-mirror.vlex.co.uk/vid/pay-nurses-these-mollycoddled-tubbies-71671808, [Pay wall], [Accessed 23rd September 2012]

Mitchell, Gregory (2005) 'Libertarian Paternalism is an Oxymoron', *Northwestern University Law Review*, Vol. 99, No. 3

Moller, Dan (2011) 'Wealth, Disability, and Happiness', *Philosophy and Public Affairs*, Vol. 39, No. 2, pp. 177-206

Montgomery, Scott M., and Ekbom, Anders (2002) 'Smoking During Pregnancy and Diabetes Mellitus in a British Longitudinal Birth Cohort', *British Medical Journal*, Vol. 324, No. 7328, pp. 26-27

Nagel, Thomas (1991) *Equality and Partiality*, Oxford: Oxford University Press

Neal, David T.; Wood, Wendy; and Quinn, Jeffrey M. (2006) 'Habits - A Repeat Performance', *Current Directions in Psychological Science*, Vol. 15, No. 4, pp. 198-202

Nglazi, Mweete D.; van Schaik, Nienke; Kranzer, Katharina; Lawn, Stephen D.; Wood, Robin; and Bekker, Linda-Gail (2012) 'An Incentivized HIV Counseling and Testing Program Targeting Hard-to-Reach Unemployed Men in Cape Town, South Africa', *Journal of Acquired Immune Deficiency Syndromes*, Vol. 59, No. 3, pp. e28-e34

NHS (2013) 'NHS Health Check: Understanding Your BMI,' [Online], *NHS Choices*, Available at: http://www.nhs.uk/Planners/NHSHealthCheck/Pages/HealthyBMI.aspx, [Accessed 1st May 2013]

NHS (2012) *The NHS Constitution for England: The NHS Belongs to Us All*, London: Department of Health

NHS Cambridgeshire (2009) 'Wee for a Wii Console at Local NHS Roadshow', [Online], *Cambridgeshire NHS*, Available at: http://www.cambridgeshire.nhs.uk/downloads/press%20releases/2009/15_December_2009_-_Wee_for_a_Wii_Console_at_local_NHS_roadshow.pdf, [Accessed 1st May 2013]

NICE (2010) 'Who We Are', [Online], *National Institute of Health and Care Excellence*, Available at: http://www.nice.org.uk/aboutnice/whoweare/who_we_are.jsp], [Accessed 25th August 2012]

Nisbett, Richard E., and Ross, Lee (1980) *Human Inference: Strategies and Shortcomings of Social Judgment*, Michigan: Prentice Hall

Nord, Erik (1999) *Cost-Value Analysis in Health Care: Making Sense out of QALYs*, Cambridge: Cambridge University Press

Nord, Erik; Daniels, Norman; and Kamlet, Mark (2009) 'QALYs: Some Challenges', *Value in Health*, Vol. 12, Supplement 1, pp. S10-S15

Nozick, Robert (1974) *Anarchy, State, and Utopia*, Oxford: Blackwell Publishing

Nozick, Robert (1997) *Socratic Puzzles*, London: Harvard University Press

NSMC (2010) *Showcase: Give it Up for Baby*, London: The National Social Marketing Centre

Nuffield Council on Bioethics (2007) *Public Health: Ethical Issues*, London: Nuffield Council on Bioethics

Nussbaum, Martha C., and Sen, Amartya K. (1993) *The Quality of Life*, Oxford: Oxford University Press

Nussbaum, Martha C. (2000) 'Aristotle, Politics, and Human Capabilities: A Response to Antony, Arneson, Charlesworth, and Mulgan', *Ethics*, Vol. 111, No. 1, pp. 102-140

Nussbaum, Martha C. (2011) *Creating Capabilities*, Cambridge MA: Harvard University Press

Olds, David L.; Henderson, Charles R.; and Tatelbaum, Robert (1994) 'Intellectual Impairment in Children of Women Who Smoke Cigarettes During Pregnancy', *Pediatrics*, Vol. 93, No. 2, pp. 221-228

Olsbrisch, Mary E.; Benedict, Sharon M.; Haller, Deborah L.; and Levenson, James L. (2001) 'Psychosocial Assessment of Living Organ Donors: Clinical and Ethical Considerations', *Progress in Transplantation*, Vol. 11, No. 1, pp. 40-49

Olsaretti, Serena (2003) 'Distributive Justice and Compensatory Desert', pp. 187-204 in *Desert and Justice*, Olsaretti (ed.), Oxford: Clarendon Press

O'Neill, Onora (1975) *Acting on Principle*, New York: Columbia University Press

O'Neill, Onora (1984) 'Paternalism and Partial Autonomy', *Journal of Medical Ethics*, Vol. 10, No. 4, pp. 173-178

O'Neill, Onora (1989) *Constructions of Reason: Explorations of Kant's Practical Philosophy*, Cambridge: Cambridge University Press

O'Neill, Onora (2002) *Autonomy and Trust in Bioethics*, Cambridge: Cambridge University Press

Osborne, Thomas, and Rose, Nikolas (1999) 'Do the Social Sciences Create Phenomena?: The Example of Public Opinion Research', *British Journal of Sociology*, Vol. 50, No. 3, pp. 367-396

Owens, John (2012) 'Creating an Impersonal NHS? Personalization, Choice and the Erosion of Intimacy', [Online], *Health Expectations*, Early View

Available at: http://onlinelibrary.wiley.com/doi/10.1111/hex.12000/full, [Accessed 30th April 2013]

Paul-Ebhohimhen, V., and Avenell, A. (2007) 'Systematic Review of the Use of Financial Incentives in Treatments for Obesity and Overweight', *Obesity Reviews*, Vol. 9, pp. 355-367

Parke, Hannah; Ashcroft, Richard; Brown, Rebecca; Marteau, Theresa M.; and Seale, Clive (2011) 'Financial Incentives to Encourage Healthy Behaviour: An Analysis of UK Media Coverage', [Online], *Health Expectations*, Early View Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1369-7625.2011.00719.x/abstract, [Accessed 21st September 2012]

Pettit, Philip (2001) *Theory of Freedom*, Cambridge: Polity

Powlick, Philip J. (1995) 'The Sources of Public Opinion for American Foreign Policy Officials', *International Studies Quarterly*, Vol. 39, No. 4, pp. 427-451

Prainsack, Barbara, and Buyx, Alena (2011) *Solidarity: Reflections on an Emerging Concept in Bioethics*, London: Nuffield Council on Bioethics

Prendergast, Michael; Podus, Deborah; Finney, John; Greenwell, Lisa; and Roll, John (2006) 'Contingency Management for Treatment of Substance Use Disorders: A Meta-Analysis', *Addiction*, Vol. 101, pp. 1546-1560

Priebe, Stefan; Burton, Alexandra; Ashby, Deborah; Ashcroft, Richard E.; Burns, Tom; David, Anthony; Eldridge, Sandra; Firn, Mike; Knapp, Martin; and McCabe, Rose (2009) 'Financial Incentives to Improve Adherence to Anti-Psychotic Maintenance Medication in Non-Adherent Patients - A Cluster Randomised Controlled Trial (FIAT)', *BMC Psychiatry*, Vol. 9, No. 1, pp. 61-70

Project Prevention (2011) 'Project Prevention: Home', [Online], *Project Prevention*, Available at: http://www.projectprevention.org/, [Accessed 1st May 2013]

Radin, Margaret Jane (1987) 'Market Inalienability', *Harvard Law Review*, Vol. 100, No. 8, pp. 1849-1937

Radin, Margaret Jane (1993) *Reinterpreting Property*, Chicago: University of Chicago Press

Ranganathan, Meghna, and Lagarde, Mylene (2012) 'Promoting Healthy Behaviours and Improving Health Outcomes in Low and Middle Income

Countries: A Review of the Impact of Conditional Cash Transfer Programmes', *Preventive Medicine*, Vol. 55, pp. S95-S105

Rawls, John (1971/1999) *A Theory of Justice: Revised Edition*, Cambridge MA: The Belknap Press of Harvard University Press

Rawls, John (1988) 'The Priority of Right and Ideas of the Good', *Philosophy and Public Affairs*, Vol. 17, No. 4, pp. 251-276

Raz, Joseph (1986) *The Morality of Freedom*, London: Clarendon Press

Reilly, Mark P.; Roll, John M.; and Downey, Karen K. (2000) 'Impulsivity and Voucher Versus Money Preference in Polydrug-Dependent Participants Enrolled in a Contingency-Management-Based Substance Abuse Treatment Program', *Journal of Substance Abuse Treatment*, Vol. 19, No. 3, pp. 253-257

Relton, Clare; Strong, Mark; and Li, Jessica (2011) 'The 'Pounds for Pounds' Weight Loss Financial Incentive Scheme: An Evaluation of a Pilot in NHS Eastern and Coastal Kent', *Journal of Public Health*, Vol. 33, No. 4, pp. 536-542

Roberts, S. A.; Brabin, L.; Stretch, R.; Baxter, D.; Elton, P.; Kitchener, H.; and McCann, R. (2011) 'Human Papillomavirus and Social Inequality: Results from a Prospective Cohort Study, *Epidemiology and Infection*, Vol. 139, No. 3, pp. 400-405

Ross, Lee (1977) *The Intuitive Psychologist and His Shortcomings: Distortions in the Attribution Process*, pp. 173-220 in *Advances in Experimental Social Psychology*, Berkowitz (ed.), New York: Academic Press

Rossi, John, and Yudell, Michael (2012) 'The Use of Persuasion in Public Health Communication: An Ethical Critique', *Public Health Ethics*, Vol. 5, No. 2, pp. 192-205

Rudinow, Joel (1978) 'Manipulation', *Ethics*, Vol. 88, No. 4, pp. 338-347

Ryan, Richard M., and Deci, Edward L. (2000) 'Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions', *Contemporary Educational Psychology*, Vol. 25, pp. 54-67

Sadurski, Wojciech (1985) *Giving Desert its Due: Social Justice and Legal Theory*, Dordrecht: D. Reidel Publishing

Sandel, Michael (1998) 'What Money Can't Buy: The Moral Limits of Markets', pp. 87-122 in Vol. 21 of *The Tanner Lectures on Human Values*, Peterson (ed.), Salt Lake City: University of Utah Press

Sandel, Michael (2012) *What Money Can't Buy: The Moral Limits of Markets*, London: Allen Lane

Sass, Hans-Martin (1992) 'Introduction: The Principle of Solidarity in Health Care Policy', *The Journal of Medicine and Philosophy*, Vol. 17, pp. 367-370

Scanlon, T. M. (1986) 'The Significance of Choice', pp. 149-216 in Vol. 8 of *The Tanner Lectures on Human Values*, McMurrin (ed.), Cambridge: Cambridge University Press

Scanlon, T. M. (1998) *What We Owe to Each Other*, London: The Belknap Press of Harvard University Press

Scanlon, T. M. (2012) 'An Interview with T. M. Scanlon by Yascha Mounk', [Online], *The Utopian*, Available at: http://www.the-utopian.org/T.M.-Scanlon-Interview-1, [Accessed 23rd September 2012]

Scheffler, Samuel (1992) 'Responsibility, Reactive Attitudes, and Liberalism in Philosophy and Politics', *Philosophy and Public Affairs*, Vol. 21, No. 4, pp. 299-323

Scheffler, Samuel (2003) 'Distributive Justice and Economic Desert', pp. 68-92 in *Desert and Justice*, Olsaretti (ed.), Oxford: Clarendon Press

Schmidt, H. (2009) 'Just Health Responsibility', *Journal of Medical Ethics*, Vol. 35, pp. 21-26

Segall, Shlomi (2007) 'In Solidarity with the Imprudent: A Defense of Luck Egalitarianism', *Social Theory and Practice*, Vol. 33, No. 2, pp. 177-198

Sen, Amartya K. (1990) 'Justice: Means Versus Freedoms', *Philosophy and Public Affairs*, Vol. 19, No. 2, pp. 111-121

Sen, Amartya K. (1999) *Development as Freedom*, Oxford: Oxford University Press

Sen, Amartya K. (2009) *The Idea of Justice*, London: Penguin Books

Shaw, Joanne (2007) 'Is It Acceptable for People to be Paid to Adhere to Medication? No', *British Medical Journal*, Vol. 335, No. 7613, p. 233

Sher, George (1997) *Beyond Neutrality: Perfectionism and Politics*, Cambridge: Cambridge University Press

Slovic, Paul (2000) *The Perception of Risk*, London: Earthscan

Strack, Fritz, and Deutsch, Roland (2004) 'Reflective and Impulsive Determinants of Social Behavior', *Personality and Social Psychology Review*, Vol. 8, No. 3, pp. 220-247

Strawson, Peter (1962) 'Freedom and Resentment', *Proceedings of the British Academy*, Vol. 48, pp. 1-25

Sumner, L. W. (1996) *Welfare, Happiness, and Ethics*, Oxford: Oxford University Press

Sunstein, Cass (1998) 'Practical Reason and Incompletely Theorized Agreements', *Current Legal Problems*, Vol. 51, No. 1, pp. 267-298

Sutherland, Kim; Christianson, Jon B.; and Leatherman, Sheila (2008) 'Impact of Targeted Financial Incentives on Personal Health Behavior: A Review of the Literature', *Medical Care Research and Review*, Vol. 65, No. 6, pp. 36S-78S

Thaler, Richard (1980) 'Towards a Positive Theory of Consumer Choice', *Journal of Economic Behavior and Organization*, Vol. 1, No. , pp. 39-60

Thaler, Richard H. and Sunstein, Cass R. (2008) *Nudge: Improving Decisions About Health, Wealth, and Happiness*, London: Yale University Press

Thornton, V. (2009) 'Who Gets the Liver Transplant? The Use of Responsibility as the Tie Breaker', *Journal of Medical Ethics*, Vol. 35, pp. 739-742

Tverdek, Edward (2004) 'The Efficiency Imperative: Five Questions', *Science & Society*, Vol. 68, No. 4, pp. 447-474

Tversky, Amos, and Kahneman, Daniel (1974) 'Judgment Under Uncertainty: Heuristics and Biases', *Science*, Vol. 185, No. 4157, pp. 1124-1131

Tversky, Amos, and Kahneman, Daniel (1981) 'The Framing of Decisions and the Psychology of Choice', *Science*, Vol. 211, No. 4481, pp. 453-458

Tversky, Amos, and Kahneman, Daniel (1986) 'Rational Choice and the Framing of Decisions', *The Journal of Business*, Vol. 47, No. 2, pp. 263-291

Tversky, Amos, and Kahneman, Daniel (1991) 'Loss Aversion in Riskless Choice: A Reference-Dependent Model', *The Quarterly Journal of Economics*, Vol. 106, No. 4, pp. 1039-1061

Van Gorp, Baldwin (2007) 'The Constructionist Approach to Framing: Bringing Culture Back In', *Journal of Communication*, Vol. 57, No. 1, pp. 60-78

Vlaev, Ivo; King, Dominic; Halpern, David; Dolan, Paul; and Hallsworth, Michael (2010) *MINDSPACE: Influencing Behaviour Through Public Policy*, London: Cabinet Office/Institute for Government

Vohs, Kathleen D.; Mead, Nicole L.; and Goode, Miranda R. (2006) 'The Psychological Consequences of Money', *Science*, Vol. 314, No. 5802, pp. 1154-1156

Voigt, Kristin (2012) 'Incentives, Health Promotion and Equality', *Health Economics, Policy and Law*, Vol. 7, No. 3, pp. 263-283

Wall, Steven (1998) *Liberalism, Perfectionism and Restraint*, Cambridge: Cambridge University Press

Wall, Steven (2001) 'Neutrality and Responsibility', *The Journal of Philosophy*, Vol. 98, No. 8, pp. 389-410

Walzer, Michael (1983) *Spheres of Justice: A Defense of Pluralism and Equality*, New York: Basic Books

Ward, Corinne; Lewis, Sarah; and Coleman, Tim (2007) 'Prevalence of Maternal Smoking and Environmental Tobacco Smoke Exposure During Pregnancy and Impact on Birth Weight: Retrospective Study Using Millennium Cohort', *BMC Public Health*, Vol. 7, No. 1, pp. 81-87

Warwick, D. P., and Kelman, H. C. (1973) 'Ethical Issues in Social Intervention', in *Processes and Phenomena in Social Change*, Zaltman (ed.), New York: Wiley Interscience

Weale, Albert (1990) 'Equality, Social Solidarity, and the Welfare State', *Ethics*, Vol. 100, No. 3, pp. 473-488

Webb, Thomas, L., and Sheeran, Paschal (2006) 'Does Changing Behavioral Intentions Engender Behavior Change? A Meta-Analysis of the Experimental Evidence', *Psychological Bulletin*, Vol. 132, No. 2, pp. 249-268

Weight Wins (2013) 'Weight Wins: The diet rewards programme that PAYS you to lose weight', [Online], *Weight Wins*, Available at: http://www.weightwins.com/, [Accessed 19th January 2013]

Wertheimer, Alan (1987) *Coercion*, Oxford: Princeton University Press

WHO (2008) *Behaviour Change Strategies and Health: The Role of Health Systems*, Copenhagen: World Health Organization

WHO (2011) *Global Status Report on Noncommunicable Diseases 2010*, Geneva: World Health Organization

Wilson, James, and Dawson, Angus J. (2010) 'Giving Liberty its Due, But No More: Trans Fats, Liberty, and Public Health', *The American Journal of Bioethics*, Vol. 10, No. 3, pp. 34-36

Wilson, James (2011) 'Why It's Time to Stop Worrying About Paternalism in Health Policy', *Public Health Ethics*, Vol. 4, No. 3, pp. 269-279

Wolff, Jonathan (2005) *Railway Safety and the Ethics of the Tolerability of Risk*, London: Railway Safety and Standards Board

Wolff, Robert Paul (1970/1998) *In Defense of Anarchism*, London: University of California Press

Zimmerman, David (1981) 'Coercive Wage Offers', *Philosophy and Public Affairs*, Vol. 10, No. 2 pp. 121-145