

**VICTORIA UNIVERSITY OF WELLINGTON**

*Te Whare Wānanga o te Ūpoko o te Ika a Māui*



School of History, Philosophy, Political Science and International Relations

---

# **Reviving Hedonism about Well-Being: Refuting the Argument from False Pleasures and Restricting the Relevance of Intuitive ‘Evidence’**

---

**Daniel Michael Turton**

A thesis submitted to Victoria University of Wellington  
in fulfilment of the requirements for the degree of  
Master of Arts in Philosophy

2008

---

## Acknowledgements

---

Heartfelt thanks first and foremost to Associate Professor Nick Agar for his invaluable guidance and genuine encouragement.

Thanks also to Professor Kim Sterelny for insightful comments on Chapter 2 and to Dr Ramon Das for some useful discussions very early on. Much gratitude to Alan, Alice, Dennis, Frieder, Matt, Nick, and especially Tony for being my thought experiment guinea pigs (and occasionally co-experimenters). Much appreciation also to the numerous graduate and postgraduate students around Australasia and especially Victoria University of Wellington for enduring, and providing useful feedback on, too many of my hedonistic seminars. Indeed, many thanks to the whole Philosophy Department at Victoria University of Wellington for your friendship, support, and for the office on the sunny side.

A million ‘thank you’s to Dad and Joe for the many tedious hours you both spent providing incredibly useful and amusing editorial feedback.

And, finally, a hefty debt of gratitude to Mum and Linda for your specialist emotional support and encouragement.

---

## Abstract

---

Throughout the vast majority of its history, hedonism about well-being has been perennially unpopular (Feldman 2004). The arguments in this essay take steps towards reviving the plausibility of hedonism about well-being. The main argument currently used to refute hedonism about well-being, the Argument from False Pleasures, is shown to lack sufficient evidence to be compelling. The main evidence provided for the Argument from False Pleasures comes in the form of two thought experiments, the Experience Machine (Nozick 1974) and the Deceived Businessman (Kagan 1998). These thought experiments typically produce strong intuitive responses, which are used to directly support the Argument from False Pleasures.

This essay investigates how theories of well-being are currently evaluated by moral philosophers, with a specific focus on the place our intuitions have in the process. Indeed, the major role that moral intuitions play in evaluating theories of well-being, despite their sometimes dubious epistemic credentials, leads to an in-depth enquiry into their inner workings and potential for containing normatively significant information. The investigation, which draws on the work of Woodward and Allman (2007), concludes that intuitions about unrealistic thought experiments should not play an important role in evaluating theories of well-being. Rather, they should only act as a warning sign, highlighting moral propositions for further analysis. Based on these findings, a new method for assessing theories of well-being is suggested and applied to a specific internalist account of hedonism about well-being to show how the Deceived Businessman and Experience Machine thought experiments lack normative significance, leaving the Argument from False Pleasures without sufficient evidence to be compelling. Indeed, this essay concludes that the Argument from False Pleasures should no longer be thought to provide any good reason to believe that hedonism about well-being is implausible. This result is only one step on the road to reviving hedonism about well-being, but it is a very important one.

---

# Contents Page

---

<b>Introduction.....</b>	<b>1</b>
Argument Summary.....	1
Chapter Outline.....	3
 <b>Chapter 1: What Hedonism about Well-Being is and Why it Needs Reviving....</b>	<b>8</b>
Introduction.....	8
Defining Hedonism about Well-Being .....	9
The Argument from False Pleasures.....	23
The Old Method of Assessing Theories of Well-Being .....	28
Conclusion .....	41
 <b>Chapter 2: Towards a Better Understanding of Intuitions and the Role They Should Play in Evaluating Theories of Well-Being .....</b>	<b>42</b>
Introduction.....	42
How Moral Intuitions Really Work.....	43
Recommendations for the Use of Intuitions in Evaluating Theories of Well-Being .....	57
The New Method of Assessing Theories of Well-Being.....	71
Conclusion .....	77
 <b>Chapter 3: Refuting the Argument from False Pleasures.....</b>	<b>78</b>
Introduction.....	78
Evaluating Hedonism about Well-Being .....	79
Deflating the Deceived Businessman.....	86
Extinguishing the Experience Machine .....	104
Conclusion .....	124
 <b>Conclusion.....</b>	<b>127</b>
Argument Summary.....	127
Chapter Outline and Implications .....	128
 <b>References .....</b>	<b>134</b>

---

# Introduction

---

## Argument Summary

Virtually since its conception, hedonism about well-being, the theory that only pleasure makes your life go well and only pain makes it go badly,<sup>1</sup> has been perennially unpopular (Blake 1926; Crisp 2006a; Feldman 2004; Savery 1934; Silverstein 2000).<sup>2</sup> Going against this trend, this essay argues that the main argument currently causing hedonism about well-being's unpopularity is not nearly as convincing as it is often believed to be. Specifically, it is argued that the Argument from False Pleasures should not be used to cavalierly reject hedonism about well-being. The most commonly used pieces of evidence provided for the Argument from False Pleasures, versions of Nozick's (1974) Experience Machine and Kagan's (1998) Deceived Businessman thought experiments, are also discussed in great detail. These thought experiments typically produce strong intuitive responses, which are used to directly support the Argument from False Pleasures. However, this essay argues that these intuitive responses do not constitute reliable or relevant evidence for evaluating hedonism about well-being.

This essay investigates how theories of well-being are currently being evaluated by moral philosophers, with a specific focus on the place our intuitions have in the

---

<sup>1</sup> A more precise definition is given later.

<sup>2</sup> That is, except for the brief period of British Empiricism, when the likes of Bentham (1789) and Mill (1861) were giving detailed accounts of hedonism (Sumner 1996).

process. Indeed, the major role that moral intuitions play in evaluating theories of well-being, despite their sometimes dubious epistemic credentials, leads to an in-depth enquiry into their inner workings and potential for containing normatively significant information. The investigation, which draws on the work of Woodward and Allman (2007), concludes that moral intuitions about unrealistic thought experiments should not play as important a role as they traditionally have in evaluating theories of well-being. Rather, they should only act as a warning sign; highlighting moral propositions for further analysis. Based on these findings, a new method for assessing theories of well-being is suggested and then applied to hedonism about well-being with the Argument from False Pleasures and the Deceived Businessman and Experience Machine thought experiments in mind. This re-evaluation of hedonism about well-being reveals how the Deceived Businessman and Experience Machine thought experiments lack the normative significance they are widely held to possess, leaving the Argument from False Pleasures without sufficient evidence to be compelling. Consequently, this essay concludes that the Argument from False Pleasures should no longer be thought to provide any good reason to believe that hedonism about well-being is implausible.

Of course, there are many other arguments against hedonism about well-being. However, it is assumed that, as a consequence of the arguments in this essay, more attention will be focused on the other objections and that they might also be rejected as scientific developments continue to shed light on how our minds work. This essay removes a major obstacle from hedonism about well-being's path to plausibility and should encourage more path-clearing. This could result in a full revival of hedonism about well-being in the future, allowing it to regain equal plausibility with the other popular views about well-being.

Taking steps towards reviving the plausibility of hedonism about well-being is a useful project because it can help us to better understand what well-being might

actually be. Indeed, if this essay helps to delineate and find truth about well-being, then it will be very useful in ongoing discussions in normative ethics and moral philosophy in general. Well-being is a central component of moral philosophy (Feldman 2004, p14; Kagan 1998) and having a better understanding of it is valuable to everyone. Indeed, Raz (2004) notes that a better understanding of well-being is important to both individuals and societies because it helps us to know what makes a life good for the one living it and, therefore, provides insight into what we might want our life or the lives of others to be like.<sup>3</sup> This is especially so if consequentialism is the preferred view of normative ethics because well-being makes for an excellent end goal of actions (Feldman 2004, p14). Philosophical theories of well-being do not, as a general rule, preach how to live one's life. This is, perhaps, particularly the case with defenses of hedonism about well-being (e.g. Feldman 2004) because of the Paradox of Hedonism: that the direct pursuit of pleasure often fails to produce the greatest overall pleasure for an individual (Crisp 2006b; Mill 1969, p85-86;<sup>4</sup> Sidgwick 1913, p48). However, accounts of well-being can be used as a desired endpoint for egoistic or egalitarian normative theories, especially by welfarists (who believe that only well-being has value (Crisp 2006b)), providing guidance of sorts for what people might want to strive towards. Indeed, even if there is no 'meaning' to life, we still have to decide how to live and how to treat others, something that a better understanding of well-being is certainly helpful for.

## Chapter Outline

With the importance and aim of this essay established, what follows is an outline of what each chapter sets out to achieve and how that relates to the overall aim of this essay. Chapter 1 begins by defining well-being, hedonism about well-being, and a specific internalist theory of hedonism about well-being to be used in this essay. The

---

<sup>3</sup> Scanlon also notes this as a possibility but attempts to create an argument against such a use of well-being and even argues against the concept of well-being itself (1998, esp. Chapter 3).

<sup>4</sup> Mill revealed this in his *Autobiography* (1873/1969, p85-86; c.f. Silverstein 2000, p294).

specific choice of theory, and especially the internal nature of pleasure that it prescribes, is then defended against other options. The theory, referred to as Hedonism about Well-Being (with capitals), is as follows:

All pleasure, and only pleasure, intrinsically contributes positively to well-being and all pain, and only pain, intrinsically contributes negatively to well-being, where pleasure and pain supervene on mental states such that pleasure is the intrinsically good part of mental states that we are consciously aware of and pain is the intrinsically bad part of mental states that we are consciously aware of.

Following the establishment of Hedonism about Well-Being as the theory to be discussed in this essay, its history of unpopularity, and its current philosophical status as a dead theory, is then discussed in order to motivate the aim of this essay. The Argument from False Pleasures and the considerable damage that it is doing to Hedonism about Well-Being's plausibility is then discussed. The two thought experiments that are usually used as evidence for the Argument from False Pleasures: Nozick's (1974) Experience Machine and Kagan's (1998) Deceived Businessman, are described, including explanations of how they supposedly support the Argument from False Pleasures. This discussion provides sufficient detail on the Argument from False Pleasures to set it as the target for refutation in order to help revive Hedonism about Well-Being. Then, drawing on the work of Timmons (2006), Kagan (1998), and Griffin (1986), how theories of well-being are evaluated by moral philosophers is discussed in detail, noting the major role that most philosophers afford to moral intuitions, despite their apparent drawbacks.

The approach moral philosophers take when evaluating theories of well-being is then formalised into the Old Method of Assessing Theories of Well-Being, which is explained with a focus on the role that intuitions play in this methodology. The role of these intuitions is then analysed further with guidance from Rachels (2003). This discussion of the Old Method of Assessing Theories of Well-Being, and particularly the role that intuitions play in it, is presented as questionable, highlighting the need for the thorough scientific analysis of intuitions that follows in Chapter 2.



Chapter 2 discusses psychological, neuroscientific and philosophical research on intuitions, and moral intuitions in particular, with the goal of providing a clear understanding of how they are formed and what that should mean for their use in evaluating theories of well-being. First, Woodward and Allman's (2007) account of moral intuitive cognition is explained in detail and corroborated by the research of other leaders in the field of cognitive science. Then, this clear understanding of how moral intuitions work is used to assess Woodward and Allman's (2007) recommendations for the appropriate use of moral intuitions. One of Woodward and Allman's (2007) important recommendations is then expanded on and made more relevant to the aim of this essay by adapting it into an argument for restricting the role of intuitions in evaluating theories of well-being. The conclusion of this argument: that intuitions about unrealistic well-being-related thought experiments are an unreliable source of evidence for evaluating theories of well-being and, therefore, should not be used if other types of evidence are available, is discussed and supported. With this argument in mind, the Old Method of Assessing Theories of Well-Being is found to be in need of at least slight amendment.

The New Method for Assessing Theories of Well-Being is then proposed and explained, including a discussion of how it deals with intuitions about unrealistic thought experiments more appropriately than the Old Method of Assessing Theories of Well-Being. This explanation includes a guide on how to apply the New Method for Assessing Theories of Well-Being and is followed by noting its implications for well-being-related thought experiments, such as the Deceived Businessman and Experience Machine, and theories of well-being in general. These implications open the door for further investigation into the normative significance of the Deceived Businessman and Experience Machine thought experiments for Hedonism about Well-Being, which is discussed in Chapter 3.

With the New Method for Assessing Theories of Well-Being sufficiently explained and supported in Chapter 2, Chapter 3 uses it to assess Hedonism about Well-Being. First, a brief case will be made for the *prima facie* plausibility of Hedonism about Well-Being before the main criticism against it is considered. Since a complete assessment of all of the arguments against Hedonism about Well-Being is not possible within the space of this essay, specific attention is focussed on the most popular one: the Argument from False Pleasures. The bulk of Chapter 3 is dedicated to assessing the normative significance that the Deceived Businessman and the Experience Machine thought experiments hold for Hedonism about Well-Being.

Each of the Deceived Businessman and the Experience Machine thought experiments is analysed in turn, revealing that, even when interpreted as charitably as possible, neither proves to have the normative significance that the proponents of the Argument from False Pleasures suppose. The power of the Deceived Businessman thought experiment is found to rest purely on a strong but misguided intuition that should not have been consulted in the first place. Furthermore, it is argued that any attempt to rework the Deceived Businessman thought experiment could not avoid this criticism of irrelevance. Likewise, the strength of the Experience Machine thought experiment is also found to rest on misguided intuitions, although less obviously so. The Experience Machine is then reinterpreted into what must be its strongest reincarnation to date; the Pleasure Machine. Even when given this new look, no good evidence to support the Argument from False Pleasures is produced.

Based on the irrelevance of the Deceived Businessman and Experience Machine thought experiments, it is then concluded that the Argument from False Pleasures lacks the evidence it requires to refute Hedonism about Well-Being. Importantly, the process used to evaluate the evidence that the Deceived Businessman and the Experience Machine thought experiments are supposed to provide will be so thorough that it is compatible with the New Method of Assessing Theories of Well-

Being *and* any approaches that place more importance on intuitive evidence. This fact makes the conclusions of Chapter 3 significant even if the arguments of Chapter 2 are not considered persuasive and the New Method of Assessing Theories of Well-Being is not accepted. The implications of this are then briefly considered, resulting in the conclusion that Hedonism about Well-Being might soon be shown to be at least as plausible as the other two major types of well-being; the desire-fulfilment and objective list theories. The major findings of this essay and their implications are then reiterated in the conclusion.

---

# Chapter 1:

## What Hedonism about Well-Being is and Why it Needs Reviving

---

### Introduction

This chapter begins with a setting of the scene for this essay by defining the theory to be defended. First well-being, the life that is good for the one living it, is disentangled from other notions of the good life. Then, a specific internalist experiential theory of hedonism about well-being, referred to as Hedonism about Well-Being, is defined and defended against other alternatives. Following this, Hedonism about Well-Being's history of unpopularity, and its current philosophical status as a dead theory, are argued for in order to show why Hedonism about Well-Being needs reviving. The objection that is currently doing the most damage to Hedonism about Well-Being's plausibility, the Argument from False Pleasures, is then discussed. Following this, the two thought experiments that are usually used as evidence for this argument: the Experience Machine and the Deceived Businessman, are described, including explanations of how they support the Argument from False Pleasures.

Then, drawing on the work of Timmons (2006), Kagan (1998) and Griffin (1986), how theories of well-being are evaluated by moral philosophers is discussed in detail, noting the major role that most philosophers afford to moral intuitions, despite their apparent drawbacks. The approach moral philosophers take when evaluating theories

of well-being is then formalised into the Old Method of Assessing Theories of Well-Being, which is explained with a focus on the role intuitions play in this methodology. The role of evaluative intuitions is then analysed further with guidance from Rachels (2003), including a discussion of reflective equilibrium. However, the conclusion Rachels (2003) reaches about the normative significance of intuitions is challenged, highlighting the need for the thorough analysis of moral intuitions that follows in Chapter 2.

## Defining Hedonism about Well-Being

In order to take any steps towards reviving Hedonism about Well-Being, it first needs to be clearly defined. This section defines well-being in general and then Hedonism about Well-Being specifically. Although well-being is associated with the good life, there are certain types of good life that well-being is not related to. These less relevant types of good life are identified and discussed before going on to an examination of exactly how well-being does relate to the good life. The point of this discussion is to avoid confusion over the different ways in which a person's life could be considered good.

Well-being does not refer to the aesthetically good life; a life that would make a good piece of art, such as a story (Feldman 2004, p 9). For example, while the life of Anne Frank makes for a great story,<sup>5</sup> we certainly do not wish that we have a traumatic life like hers. Well-being also does not refer to the good example of a life; the life that would be documented and reported on in a museum exhibit showing what the life of the average modern-day human is like (Feldman 2004, p9). The average life, although probably lived by many of us, certainly does not make for a very appealing version of the good life. Some lives, such as the life of Mother Teresa, are considered to be

---

<sup>5</sup> See: *The Diary of Anne Frank, The Critical Edition*, by Anne Frank and the Netherlands State Institute for War Documentation (1989).

causally good. A causally good life brings about good, as exemplified by Mother Teresa's work with the sick (Feldman 2004, p8-9). However, living a causally good life is no guarantee that the life will be good *for the one living it*, as again exemplified by Mother Teresa, whose life was full of depression and loneliness.<sup>6</sup> Well-being is also often thought of as the healthy life; the life you might be trying to achieve if you visit a well-being centre (Crisp 2006b). Although a well-being related account of the good life might involve being healthy, the philosophical notion of well-being is generally broader than that (Crisp 2006b). While these distinctions between different types of good lives are quite obvious, the morally good life is not quite as easy to disentangle from the type of good life that well-being is referring to: the prudentially good life. The next few paragraphs are dedicated to this more subtle task.

The type of good life that well-being refers to is the prudentially good life, or the life that is good for the one living it (Crisp 2006b; Feldman 2004, p8-12). There are many definitions of well-being in the literature on psychology and moral philosophy but most are fairly synonymous. For example, psychologists are becoming increasingly interested in subjective well-being, which many of them see as a preferable alternative to economic accounts of objective well-being (Angner 2005), such as Gross Domestic Product (Pigou 1960) or indicators of real income or purchasing power (Rutherford 2002). Economic accounts of well-being are too abstract to pursue here, but psychologists' accounts are somewhat closer to what philosophers mean by well-being, making them more useful to discuss. Most psychologists tend to endorse a happiness account, a life-satisfaction account, or a combined account of well-being. Those who equate well-being with happiness attempt to measure well-being with direct questions about happiness (e.g. Bradburn 1969), monitoring smiling (e.g. Ekman, Friesen & O'Sullivan 1988; Ekman & Rosenberg 1997), averaging the data compiled by constantly paging participants and getting them to report their happiness at that moment (e.g. Kahneman 1999; Kahneman *et al.* 1999; Kahneman *et al.* 2004),

---

<sup>6</sup> See *Mother Teresa: Come Be My Light* (Doubleday, 2007), a compilation of her writing and commentary by Brian Kolodiejchuk.

and even using neuroimaging techniques (e.g. Davidson 1992; 2000; Davidson, Jackson & Kalin 2000). Psychologists who equate well-being with life satisfaction: being satisfied with how one's life is going (e.g. Seligman & Csikszentmihalyi 2000; Pavot & Diener 1993), or combination accounts (e.g. Urry *et al.* 2004), use similar measures but with more of an emphasis on self-reporting. However, all of these psychological accounts of well-being are designed with measurement in mind, which increases the chances of them being restricted in a way that philosophical definitions do not have to be.

Philosophers are slightly more consistent with their definitions, describing well-being as: what is “non-instrumentally or ultimately good *for* a person” Crisp (2006b, np), “quality of life” or “thriving” (Nussbaum & Sen 1993, p1), “a life going well” (Moore & Crisp 1996, p599), “a person's welfare” Sumner (1996, p1), “what makes your life go well” or what is in “your best interests” (Keller 2004, p39; closely paraphrased from Parfit 1984, p493), “the life that is good for the person whose life it is” (Raz 2004, p269), and “what it is for a given individual to be better... off” (Kagan 1998, p41). Several concepts are also used as equivalent to well-being by some philosophers, although they are generally seen as “near-equivalents”, such as “...happiness, flourishing, eudaimonia, and utility” (Moore & Crisp 1996, 599). The most important themes represented in these definitions are the subjective and prudential nature of well-being. It should also be noted that the term well-being can refer to both positive and negative lives for the person living them, such that the addition of certain experiences to a life could make it better for the person living it, thereby increasing her well-being, or worse for the person living it, thereby decreasing her well-being (Nagel 1979, p2). This clear description of well-being makes it easier to differentiate it from the morally good life.

Well-being does not refer directly to the morally good or virtuous life: the life that an individual should lead just because it is the morally right way to live (Arneson 1999,

p1; Feldman 2004, p8; Railton 1989; p155; Silverstein 2000, p280). It is fairly common to conflate the morally good life with well-being, an error that can occur for various reasons. For example, the word 'good' can be taken to mean 'morally good', as opposed to 'prudentially good'. This conflation can also arise when moral intuitions do not allow a person to imagine any kind of life that she considers to be immoral to be enjoyable from the inside, even by someone with different preferences to herself. Such a lack of empathy and imagination can cause some people to unintentionally conflate accounts of well-being with their personal versions of the morally good life, which leads them to unfairly reject many accounts of well-being. A very similar type of conflation, however, is not necessarily erroneous. Some moral philosophers, such as Aristotelian Virtue Ethicists (e.g. Hursthouse 1999), propose positive arguments that well-being (the prudentially good life) is the same as, or at least very close to, the virtuous life. This kind of equating of the morally good life with the prudentially good life is reasonable as long as it is explained *how* the ultimate bearer or bearers of value in a life make that life better for the person living it.

It should now be clear that a person's well-being concerns how well his life goes for him, although it might not yet be obvious what it *means* for a life to be going well for the person living it. Many different accounts of well-being have been developed over the last 2,000 or so years, but they nearly all fall into one of three categories outlined by Derek Parfit (1984, p493): either hedonistic theories, desire-fulfilment theories, or objective list theories. Objective list theories, such as Hurka's (1993) account of Perfectionism and Moore's account of Ideal Utilitarianism (1903; 1912), propose a list of things that make lives go better or worse for the individual living them (Parfit 1984, p493). These theories may include certain mental states, such as pleasure, or the satisfaction of preferences in the list of things that make lives go better. However, they cannot consist of solely one or the other because that would make the account of well-being collapse into one of the other two major categories.



Desire-fulfilment accounts of well-being hold that our life goes well to the extent that our preferences are satisfied (Parfit 1984, p493). There are several more specific varieties of desire-fulfilment accounts, but the more considered ones, such as Railton's (1986) very complex account,<sup>7</sup> tend to recommend that only the satisfaction of certain types of desires actually improve well-being. The usual candidates are more fundamental desires, rationally formed desires, or desires about which the subject of the life in question is relevantly informed (Brandt 1979; Griffin 1986, p11).

Hedonistic accounts of well-being, as illustrated by Parfit (1984), describe the best life for someone as the happiest one. However, Hedonism about Well-Being is really just one type of mental state or experiential account, albeit the dominant one (Kagan 1998, p34). A mental state account of well-being purports that only certain mental states, such as being in pleasure or pain, contribute to well-being and experiential accounts maintain that only our experiences, such as the experience of pleasure or pain, contribute to our well-being. The only notable difference between the two views is that it is possible to have a certain type of mental state without experiencing it and it might be possible to have an experience<sup>8</sup> that doesn't affect your mental state at all, such as someone making a rude gesture at you behind your back. However, as this essay argues in more detail later, the best accounts of hedonism are only concerned with the intersection of the two: mental states that we experience. A more detailed explanation of hedonism about well-being and a defense of internalist experiential mental state accounts of hedonism about well-being follow.

Before giving a precise definition of hedonism about well-being, however, the general notion of hedonism about well-being needs to be clarified by distinguishing it from two other types of hedonism. The first type of hedonism that needs to be differentiated from hedonism about well-being is folk hedonism. Folk hedonism is

---

<sup>7</sup> Railton states that "an individual's good consists in what he would want himself to want, or to pursue, were he to contemplate his present situation from a standpoint fully and vividly informed about himself and his circumstances, and entirely free of cognitive error or lapses of instrumental rationality" (1986, p16).

<sup>8</sup> More on this point follows when externalism is discussed.

the way that the word 'hedonism' is used in everyday language and non-philosophical literature. It commonly refers to a way of life that involves pursuing risqué pleasures, especially pleasure from sex, drugs and food, with scant or no regard for the future consequences to others or even the hedonist himself (Feldman 2004, p21). For example, Hedonism Resorts invites anyone over eighteen years old to purchase a "daring day pass" or a "naughty night pass" and "Experience a world where almost anything goes".<sup>9</sup> And, from the novel *Drop City*, by T.C. Boyle (2003, p416): "What was his goal in life? Pan had asked [Joe] one night... To have a good time. To get drunk, get laid, raise some hell and answer to nobody. 'So you're a hedonist, then?' Dale had put in. 'Bet your ass I am,' Joe said." Also, for Feldman (2004, p21) an unnamed but easily identifiable "former publisher of a slightly scandalous girly magazine" is the paradigm folk hedonist, with his "beviess of voluptuous young women" who regularly "reveal extensive amounts of tanned flesh".

Hedonism about well-being generally makes no distinction between different types of pleasure, such that sex is not necessarily chosen over a conversation with a good friend.<sup>10</sup> However, by not making a distinction between intellectual and bodily pleasures (or higher and lower pleasures as Mill referred to them),<sup>11</sup> most philosophical hedonists were attacked by other philosophers for endorsing a moral theory that encouraged profligacy (Feldman 2004, p21). For example, Hedonism has been referred to as the philosophy of swine (c.f. Crisp 2006a, p630; 2006b, ss4.1), because it doesn't discriminate between pleasure from morally irreproachable acts and pleasure from morally questionable acts, such as sex with animals. Indeed, this view has been held by many important philosophers, who all refuse to grant equal value to pleasures from acts that are disgraceful, base, malicious, or undesirable in

---

<sup>9</sup> For information on Hedonism Resorts, see [http://www.superclubs.com/brand\\_hedonism/index.asp](http://www.superclubs.com/brand_hedonism/index.asp). Accessed on 22/02/2008.

<sup>10</sup> The hedonism of Aristippus is the only well-known example of a philosophical account of hedonism that is close to folk hedonism (c.f. Feldman 2004, p30-33).

<sup>11</sup> See Mill's (1861, p12-14) *Utilitarianism*.

some way (Feldman 2004, p 38-39).<sup>12</sup> It seems that these philosophers had a moral problem with hedonism based on a conflation of the morally good life and the prudentially good life, as discussed above. Presumably, many anti-hedonistic philosophers have been heavily influenced by folk hedonism, and its pervasive image of profligacy, in this way, motivating them to argue ferociously against it.

Another form of hedonism that needs to be distinguished from hedonism about well-being is psychological hedonism; an explanatory theory about our behaviour, which claims that all of our actions are necessarily either subconsciously or consciously pleasure-seeking. Psychological hedonism has been endorsed by some historically important thinkers, such as William James (1902, p78) and perhaps Darwin and Spencer (c.f. Badcock 2000, p125-126), and has been used as evidence for hedonism about well-being by many of its earlier supporters, such as Bentham (1789), Epicurus (c.f. Feldman 2004, p5), and Mill (1861). However, psychological hedonism's claim that everything we do is pleasure-seeking is not required to support hedonism about well-being and so will not be defended here. This is fortunate because psychological hedonism would be very hard to defend, particularly in the light of examples Feldman (2004, p5) provides, such as a moral saint who seeks the pleasure of others as the goal of his actions, or perhaps seeks to do what is morally right just because he believes it to be the right thing to do. Furthermore, consider performing habitual actions, such as running one's hand through one's hair, making a clicking sound with one's tongue, or sacrificing something for one's child. Providing reasoned justifications as to how these actions could be solely based on the pursuit of pleasure is a daunting, if not impossible task.<sup>13</sup> The reason that psychological hedonism is not useful for a defense of hedonism about well-being is that it makes a descriptive claim about how we function, as opposed to a theoretical claim about what kind of life would be best for us. Nevertheless, the distinction is important because hedonism about well-being has

---

<sup>12</sup> Feldman cites Aristotle (Nicomachean Ethics, X, 3), Brandt (1979, p316), Brentano (1969, p90) and Broad (1930, p53-54).

<sup>13</sup> Some examples adapted from Irvine (2006).

been rejected in the past due to its ignoble ties to psychological hedonism (e.g. Dewey 1922; c.f. Blake 1926, p3).

Several philosophical dictionaries are also misleading on the topic of what the term hedonism refers to. Simon Blackburn (2005, p161) and Gregory Pence (2005, p25) both severely limit hedonism by describing it as a specifically normative theory, which claims that pursuing *our own* pleasure should be the aim of all our actions without mentioning any other types of hedonism. This normative account of hedonism is better known as hedonistic egoism and stands opposed to hedonistic utilitarianism (Moore 2004), which states that the morally good action is the one that is likely to maximise net pleasure *for everyone* (Brink 2006, p 381). In a similarly misleading way, Vesey and Foulkes (1990) take the opposite approach and only discuss hedonistic utilitarianism and psychological hedonism. Furthermore, while Gosling (1995, p337) acknowledges the two different normative branches of hedonism, and psychological hedonism, in his philosophical dictionary entry, he, like Blackburn (2005), Pence (2005), and Vesey and Foulkes (1990), fails to specifically mention hedonism about well-being. These omissions are inexcusable because it is clear that normative versions of hedonism *rely* on hedonism about well-being as a justification.<sup>14</sup> Such narrow descriptions of hedonism could lead a reader to assume that hedonism is just a normative theory, which is simply not the case.

Some of the first records of philosophical discussion, as captured by Plato in *Protagoras* and *Philebus*, mention hedonism about well-being. Furthermore, some philosophers have based all of their work around hedonism about well-being, such as Epicurus, whose account of hedonism is discussed in Book X of *The Lives of the Eminent Philosophers* by Diogenes Laertius (c.f. Feldman 2004, p91). Perhaps the most detailed definition of hedonism about well-being is provided by Feldman (2004, p25-

---

<sup>14</sup> However, Gosling's mention of rationalizing hedonism, which he describes as "pleasure is the only object that makes a pursuit rational" comes close (1995, p337). Furthermore, that particular philosophical dictionary, *The Oxford Companion to Philosophy* (edited by Honderich 1995), also has an entry on Ancient Hedonism (p337-339), which does discuss hedonism about well-being, or the good life.

30), who refers to it as Default Hedonism. Feldman's (2004, p27) Default Hedonism can be summarised as follows:

Every episode of pleasure is intrinsically good. Every episode of pain is intrinsically bad. An individual's well-being is entirely determined by the intrinsic values of the episodes of pleasure and pain contained in that life, such that as an individual's net pleasure (total pleasures minus total pains) increases, so does their well-being.<sup>15</sup>

Feldman's (2004, p27) Default Hedonism is a fairly general account of hedonism about well-being (which is in its favour because it captures the commonalities between most of the variations of hedonism). A similar definition is provided by David Sobel (2002, p240), who describes hedonism about well-being as "the thesis that pleasure is the only intrinsic benefit and pain the only intrinsic harm an agent's well-being can receive." Sobel's (2002, p240) definition does seem to capture the most important components of hedonism about well-being, although, it remains silent on whether *all* pleasure and pain contribute to well-being. This omission counts against the definition because if *not all* pleasure and pain intrinsically contribute to well-being, then the theory becomes a lot more flexible, allowing it to avoid all objections by claiming that certain kinds of pleasure and pain do not intrinsically contribute to well-being for some reason. Such flexibility allows for hedonism about well-being to be bent out of shape, creating a theory that is potentially quite unlike what we really have in mind when we discuss it. Therefore, a combination of Sobel's (2002) and Feldman's (2004) definitions will be used, resulting in the following very simple definition of hedonism about well-being:

Hedonism about well-being is the theory that all pleasure, and only pleasure, intrinsically contributes positively to well-being and that all pain, and only pain, intrinsically contributes negatively to well-being.

This definition is very general, with different accounts of pleasure and pain allowing for many different variants of hedonism to be captured. And, it is also fairly self-explanatory, except perhaps for the use of 'intrinsically', which is explained below.

---

<sup>15</sup> Feldman's (2004, p27) actual definition is more detailed than this but the essential elements (as I see them) are repeated here.

For something to be intrinsically, or non-instrumentally, valuable, it has to be valuable in and of itself; not because of anything else that it might bring about (Feldman 2004, p58). Many things are instrumentally valuable in that they *provide* something that is intrinsically valuable. The common test used to discern if something is of intrinsic or instrumental value is to ask why the thing is valuable. If no reason can be given, then the thing in question is probably intrinsically valuable. However, if we can provide an answer, then the thing is likely to be only instrumentally valuable. Indeed, a particular event of apparent value might have to be reduced in this way several times before settling on the intrinsic value provided by this instrumentally valuable event. For example, finding \$100 in your pocket is *prima facie* valuable. However, when you ask yourself *why* it is valuable, you will likely have to ask yourself several times before you reach the non-reducible reason for why the \$100 is intrinsically valuable to you.

When the notion of intrinsic value is considered in the case of well-being, the exact question to ask about a thing of *prima facie* value is: ‘why does this thing make my life better for me?’, or ‘why does this thing increase my well-being?’ It should be noted that many things, such as finding \$100 can lead to intrinsically valuable and instrumentally valuable things at the same time. For example, if a hedonist found \$100 she might feel pleased about it (and the pleasure she felt would have intrinsic value for her) and then she might buy herself a delicious vegetarian curry (the money is instrumentally valuable in that it allowed her to buy the curry) and then feel pleasure from eating it (intrinsic value from the pleasure again). Interestingly, the answers that you find yourself providing for the question: ‘why does this thing increase my well-being?’, also provide insight into which of the three general accounts of well-being you intuitively support. For example, if you always end up on the answer ‘because it gives me pleasure’ and cannot provide something intrinsically valuable that pleasure instrumentally brings about, then your intuitions about well-being are hedonistic. Discerning what has intrinsic value is epistemologically prior to

discovering instrumentally valuable things, which only lead us to what is intrinsically valuable. This is why theories of well-being ultimately aim to account for what thing or things intrinsically affect well-being and how it or they achieve it.

The general definition of hedonism about well-being mentioned above is a broad one because it does not define pleasure in any particular way. Of course, exactly how a specific account of hedonism about well-being defines pleasure is vital to its success, since many objections against hedonism refer to particular notions of pleasure (e.g. Dennet 1988; Ryle 1954; Sprigge 1988). Indeed, certain accounts of pleasure, such as some of Feldman's (2004, p55-63) variations on his own Intrinsic Attitudinal Pleasure (which describe pleasure as a positive attitude about a state of affairs) appear to be specifically designed to avoid certain objections to hedonism. For example, Feldman's (2004, p112-114) Truth-Adjusted Intrinsic Attitudinal Hedonism, which allows the value pleasure contributes to well-being to be adjusted based on if the pleasure is based on truth or falsity, seems to be posited solely to refute the Argument from False Pleasures (discussed below).<sup>16</sup> Although this essay will not attempt a lengthy defense of a particular account of pleasure, an account will be stipulated for sake of consistency when dealing with the objections to hedonism about well-being.

Following Bentham (1789), Crisp (2006c) and Hume (1777), and to a lesser extent Moore (1903, p12-13) and Broad (1930), this essay proposes an internalist account of pleasure as an enjoyable experience, which places pleasure and, thereby, value for well-being solely within the mind of each individual. On this kind of account, pleasure supervenes on brain states, such that there can be no change in pleasure without a change of mental states. Internalist accounts of pleasure stand opposed to

---

<sup>16</sup> Feldman (2004, p112-114) rightly claims that his Truth-Adjusted Intrinsic Attitudinal Hedonism adequately deals with the Argument from False Pleasures. However, Truth-Adjusted Intrinsic Attitudinal Hedonism is not an easily recognisable form of Hedonism because of the theory of pleasure it posits. In fact, I have argued against Attitudinal Pleasure being hedonistic elsewhere (Turton 2007). Regardless of the final verdict on whether any of Feldman's (2004) variants of Hedonism satisfy the criteria for hedonistic theories about well-being (Feldman 2004, Chapter 8), his variants are certainly not as hedonistic as a theory of Hedonism about well-being that uses the account of pleasure that follows here.

externalist accounts of pleasure, which can share the value for well-being between the mind and the outside world (Moore 2004). Internalist accounts of pleasure are subject to the criticism that introspection reveals a wide variety of pleasures, many of which are enjoyable in different ways.<sup>17</sup> This is an empirical issue, however, and one that might be resolved in the not-too-distant future, as discussed briefly in Chapter 3. Regardless, this essay will continue on the assumption that internalism about pleasure is true, since a philosophical defense would require another essay and cognitive scientists will probably provide a more informative answer to this question in due course.

One fixed definition of pleasure is certainly needed to make an account of hedonism about well-being complete and to avoid the main mistake made by Feldman (2004) in his defense of hedonism about well-being. Feldman's (2004) error was to employ a strategy that included using different definitions of pleasure to refute each different objection to hedonism. This strategy fails to provide a complete vindication for any particular account of hedonism about well-being because no particular account is run through the gauntlet of all of the important arguments against it. Clearly, for hedonism about well-being to become plausible in the face of a barrage of different objections, then one specific account of it must be able to survive each of the important objections (Turton 2007).

Feldman (2004), continuing a popular trend in recent times (Crisp 2006a), also used externalist accounts of pleasure in his defense of hedonism about well-being. As mentioned above, externalist accounts of pleasure put some of the value of pleasure out in the world, as opposed to internalist accounts, which hold that pleasure, and the value it provides, supervene on mental states only.<sup>18</sup> Externalist accounts of pleasure

---

<sup>17</sup> (e.g. Griffin 1986, p8; Gosling 1969, p37-40; Sprigge 1988, p130, Sumner 1996, p92-93; Feldman 1997 p8, 132; Bernstein 1998, p25; Carson 2000, p13-14; Sobel 2002, p241; c.f Crisp 2006c, p104).

<sup>18</sup> See Crisp (2006a; 2006b) and Sobel (2002) for more information on the distinction between internal and external pleasures and particularly Crisp (2006a) for a defense of internalism about pleasure being more suited to hedonism about well-being.



do not match well with the common conception of what pleasure is because they begin to merge the sensation of pleasure with the actions and objects in the world that do or should cause pleasure.<sup>19</sup> Internalist accounts, however, focus on the phenomenology of pleasure, how it feels to us on the inside. Internalist accounts of pleasure also completely comply with the requirements of experiential accounts of well-being, which is important because Nozick (1974, p42, 43) devised the Experience Machine thought experiment specifically to show that more than just how our experiences feel to us “from the inside” matters to us.<sup>20</sup> For these reasons, the following internalist account of pleasure (and pain) will be used henceforth:

Pleasure and pain supervene on mental states such that pleasure is the intrinsically good part of conscious mental states, and pain is the intrinsically bad part of conscious mental states, where the intrinsically good (or bad) part of a conscious mental state is the part that we find enjoyable (or the opposite of enjoyable).

This account of pleasure (and pain) is broad and undiscerning. The account is broad because *all* mental states that we are consciously aware of, and have a positive or

---

<sup>19</sup> Arguably, all externalist accounts of pleasure are not very hedonistic because of their inclusion of something *other* than how pleasure feels from the inside in what intrinsically contributes positively to well-being. The externalist move allows them to give varying weights (including no weight) to any pleasure depending on what causes it. This move seems opposed to a central tenet of hedonism; that all pleasure, *and only pleasure*, intrinsically contributes positively to well-being, because it allows for something other than pleasure to dictate how much value that pleasure has. Even if this pleasure-plus-qualifier type of externalism about pleasure can be passed off as pleasure, it is clear that it is not as hedonistic as an internalist account, which focuses solely on the pleasure. A classic example of externalist accounts being un-hedonistic can be seen in Feldman (2004, p55-63), who takes his externalism about pleasure to the extreme. His account of pleasure, which he describes as Intrinsic Attitudinal Pleasure, asserts that the only thing that contributes positively to well-being is being intrinsically pleased about some state of affairs – being pleased about something for itself and not because it leads to something else. This view of pleasure clearly causes his Intrinsic Attitudinal Hedonism to collapse into a desire-fulfillment account of well-being when one considers being pleased about something but feeling no pleasure about it. How can one know that one is pleased when one feels no pleasure? Indeed, it seems that one can have an opinion about something but as soon that opinion stops being neutral, one becomes pleased about it, which necessarily entails the feeling of at least some small amount of pleasure. The capacity to feel pleasure, allows us to differentiate between things that please us and things that neither please nor displease us. Indeed, with all of the phenomenology of pleasure removed, all that is left in Feldman’s Intrinsic Attitudinal Pleasure is the preference for certain states of affairs.

<sup>20</sup> Externalists will agree with Nozick, suggesting that the pleasures we would experience in the Experience Machine would not be valuable because they would fail to fulfill some criterion about accurately tracking the world outside of our heads. Unfortunately for externalists, that criterion is difficult to specify in a way that is not trivial. Arguably, the Experience Machine is the *real external* cause of anything we experience while we are in it. Furthermore, our experiences outside of the machine are dogged by various delusions, deceptions and, perhaps most importantly for the philosophically-minded, doubt about the very existence of the external world.

negative valence (as opposed to a neutral one), will intrinsically contribute to our well-being. It should be noted that mental states are often complex and that the positive or negative valence of a mental state comes from the intrinsically good or bad part of that mental state. The account is undiscerning because the *cause* of the mental state is unimportant. The intrinsically good or bad part of mental states directly contributes to well-being regardless of what caused those mental states. When construed in this way, our pleasure and pain can be thought of as the enjoyment and suffering that we are consciously aware of.<sup>21</sup> Feeding this account into the definition of Hedonism about Well-Being above, we can state a particular theory of hedonism about well-being, which will be referred to as Hedonism about Well-Being (with capitals):

All pleasure, and only pleasure, intrinsically contributes positively to well-being and all pain, and only pain, intrinsically contributes negatively to well-being, where pleasure and pain supervene on mental states such that pleasure is the intrinsically good part of mental states that we are consciously aware of and pain is the intrinsically bad part of mental states that we are consciously aware of.

This specific account of hedonism about well-being is the most appropriate for use in this essay because the arguments against hedonism about well-being that are discussed below are used to refute exactly this kind of theory. Therefore, attempting to defend Hedonism about Well-Being against these objections is much more preferable than using an external account of hedonism about well-being, which would only skirt around the objections. Indeed, using Hedonism about Well-Being allows the objections to be tackled head-on, producing a result that will satisfy both traditional hedonists and those who have been arguing against them.

---

<sup>21</sup> Following Crisp (2006c, p103-111) the emphasis on enjoyment (as opposed to pleasure) is to help ward off irrelevant intuitions about pleasure being profligate.

## The Argument from False Pleasures

Hedonism about well-being has had many opponents over the years, the arguments of whom have made it perennially unpopular amongst philosophers (Blake 1926; Crisp 2006a; Feldman 2004; Savery 1934; Silverstein 2000), except for the brief period of British Empiricism (Sumner 1996).<sup>22</sup> This unpopularity, caused by the strength and number of arguments devised to undermine it, has effectively killed hedonism about well-being as a philosophical theory. As with many other deceased entities, hedonism about well-being is still discussed, however, the discussion is often disparaging in tone and focuses mainly on its opposition's latest 'knock-down' argument (Silverstein 2000). Indeed, it seems that hedonism about well-being is often only discussed to allow readers to learn from the mistakes that it supposedly makes (Sumner 1996). This approach of acknowledging hedonism about well-being, refuting it due to the latest objection, and then moving on to more plausible theories is common in texts specifically on well-being (e.g. Darwall 1997; Griffin 1986; Sumner 1996), metaethics (Brink 1989) and normative ethics (e.g. Kagan 1998), and introductory texts on ethics (Bagani & Fosl 2007; Frankena 1973; Furrow 2005; Rachels 2005; Rosen 1993) and political philosophy (Kymlicka 1990). Furthermore, this type of swift rejection of hedonism about well-being is often outright (e.g. Furrow 2005, p112; Griffin 1986, p10) even though some of its staunchest opponents have acknowledged that hedonism about well-being cannot be proven wrong (Moore 1903, p238). There have been several influential arguments used by the anti-hedonistic movement since hedonism about well-being's conception.<sup>23</sup> However, only the argument that is currently most influential will be discussed in any detail in this essay.

---

<sup>22</sup> Indeed, Hedonism could even have been considered popular during this era and was discussed much more positively by the leading philosophers of the time (e.g. Bentham 1789; Hobbes 1650; Hume 1754, 1777; Locke 1689; Mill 1861; c.f. Crisp 2006b, p619).

<sup>23</sup> Early influential arguments came from Plato's Socrates (e.g. *Philebus* 21a) and more recent arguments were most famously discussed by Moore (e.g. 1903, Chap 3).

If a recent text on ethics (or related subject) is going to hastily sweep aside hedonism about well-being, then the Argument from False Pleasures is the brush they are likely to do their sweeping with.<sup>24</sup> And, despite admonitions against dealing with it this way (Crisp 2006b), the same cavalier treatment of hedonism about well-being can also be found in many undergraduate philosophy courses: an explanation of hedonism about well-being, followed quickly by a heralding of the champions of the Argument from False Pleasures, and then on to more plausible theories of well-being. The Argument from False Pleasures attempts to refute hedonism about well-being by arguing that pleasures based on falsity, deception, misinformation, or experiences too far detached from reality, are not as valuable as pleasures based on truth or more directly based on reality. The Argument from False Pleasures can be described as follows:

- P1) Hedonism about well-being states that all pleasure, and only pleasure, intrinsically contributes positively to well-being and that all pain, and only pain, intrinsically contributes negatively to well-being.
  - P2) Pleasure based on truth, or something like it, contributes more positively to well-being than pleasure based on falsity.
  - P3) Therefore, something other than pleasure (truth of some sort) must contribute positively to well-being.
- 
- c) Therefore, hedonism about well-being is false.

The controversial premise of the Argument from False Pleasures is Premise 2. Premise 1 is a common and reasonable definition of hedonism about well-being, as discussed and endorsed above. Indeed, straying from this definition would depart from the core principles that have always underpinned hedonism, resulting in a theory of well-being that may have hedonistic influence but should not be considered purely hedonistic. Premise 3 and the major conclusion, Premise 4, are not contentious because they deductively follow from the first two premises. Premise 2, however, needs evidence of some sort to make it compelling and this evidence is supposedly provided by the current champions of the anti-hedonists: the Experience Machine and the Deceived Businessman thought experiments. These two thought experiments are explained in more detail below.

---

<sup>24</sup> As typified by Bagani & Fosl (2007, p74-76), Brink (1989, p223-224), Darwall (1997, p162, 178), Griffin (1986, p9-10), Kagan (1998, p34-36), Kymlicka (1990, p13-14), and Rachels (2005, p103-104).

The Experience Machine thought experiment, as described by Robert Nozick (1974, p42-45), was the first and is probably the most renowned of the new wave of anti-hedonistic champions; even though, it is not clear that he initially intended it as a refutation of hedonism (Feldman 2004, p109). Indeed, Nozick (1974) employed the Experience Machine as just a cog in his wider argument involving how animals should be treated. However, the function this cog now serves is to cast doubt on the commonplace assumption that nothing else really matters to us “other than how our lives feel from the inside” (1974, p43). Importantly, if Nozick’s argument is correct, then all purely internalist experiential accounts of well-being,<sup>25</sup> including Hedonism about Well-Being, will be false. Regardless of his initial intention, mental images of the Experience Machine, perhaps aided by exposure to *The Matrix* trilogy of films,<sup>26</sup> have converted many would-be undergraduate hedonists into non-believers, further swelling the ranks of those who find Hedonism about Well-Being implausible.

In the Experience Machine thought experiment, Nozick (1974, p42) asks us to imagine a machine built by “superduper neuropsychologists” that can provide us with any experience we desire, such as the pleasure of publishing an award-winning novel. The machine can also provide a comprehensive range of other amazing experiences that we might never have thought of, such as the pleasure of seducing a whole volleyball team at once. Nozick invites us to plug into the Experience Machine, but expects that, even after a two-year test run, we would not accept his invitation (1974, p42). Indeed, when posed with this choice, most people tend to forgo a life of endless pleasures for what they currently have, a presumably less pleasurable but more real life.<sup>27</sup> This choice appears to be based on the “firmly held” negative intuition aroused by considering a life plugged in to the Experience Machine, an intuition that Sobel (2002, p244) thinks strikes “at the heart of hedonism”.

---

<sup>25</sup> Experiential accounts of well-being hold that only experiences can contribute intrinsic value to well-being (as discussed above).

<sup>26</sup> Nozick refers to the machine as a type of tank that you would be floating in while having electrodes plugged into your brain (1974, p42 & 43), a gruesome description given illustration by The Matrix movies.

<sup>27</sup> Personal experience has shown me that about 9 out of 10 undergraduate philosophers do not want to plug into the Experience Machine

Although Nozick does not claim to have all the answers to the question of what, “other than how our lives feel from the inside”, matters to us, he does have a suggestion. He (1974, p43-44) briefly considers being a certain kind of person and achieving certain things as possible reasons for our not wanting to get into the Experience Machine but quickly dismisses them, noting that more specific machines could offer such experiences and we still would not want to plug into them. Nozick then proposes that our wanting to *really* experience the limitless reality of living our own real life, as opposed to having a machine live it for us, is probably what prevents us from plugging in to the Experience Machine (1974, p43-45). If this is true, as many people believe it is, then it shows that something like truth or reality matters to us, not just pleasure. Based on the realisation that truth or reality matters to us, the assumption is then made that truth (or something like it) must contribute positively to well-being, thereby providing evidence for Premise 2 of the Argument from False Pleasures.

The Deceived Businessman thought experiment and its more recent variants have a more traceable evolution than Nozick’s (1974) Experience Machine: from Nagel (1979), to Kagan (1998) and on to other modern variants such as Rachels (2005, p104). Perhaps initially influenced by Nozick’s (1974) Experience Machine thought experiment, Thomas Nagel (1979, p4-5) laid out the foundations for the Deceived Businessman thought experiment and a more formal understanding of the Argument from False Pleasures. Nagel (1979, p4) notes the common remark: “that what you don’t know can’t hurt you” (the underlying principle of all internalist experiential accounts of well-being, such as Hedonism about Well-Being), but then offers reasons to think that the remark is wrong. He gives the example of someone being betrayed and then claims that the reason people are unhappy when they find out about betrayal is because “it is bad to be betrayed”, not because it is bad to find out about it (Nagel 1979, p5). The claim that it is bad to be betrayed implies that deception, lack of truth, or something like it is bad for people whether they experience it or not.

Therefore, if correct, this claim supports Premise 2 of the Argument from False Pleasures by suggesting that the intrinsic ‘badness’ of deception, lack of truth, or something like it negatively contributes to our well-being even when we do not experience it.

The example is painted more vividly (and perhaps most vividly) by Shelly Kagan (1998, p34-36) who also premises his discussion with the “common saying that what you don’t know can’t hurt you”, before going on to give reasons to doubt its wisdom. He asks us to imagine a successful businessman who died thinking that he had achieved everything he wanted: a loving wife and children and the respect of the community (Kagan 1998, p34). However, Kagan (1998, p35) then reveals that the businessman was completely wrong about his assessment of how things had gone: his wife was cheating, his children and the community at large were just using him for their own ends, and his business partner had been stealing from the soon-to-be-bankrupt business. Kagan (1998, p35) asserts that the businessman’s life did not go as well as it could have and points out that if he had really been loved and respected like he thought he was, then his life would have gone better.

Indeed, the intuition that the Deceived Businessman thought experiment elicits is not a pleasant one. Kagan (1998) offers similar reasoning to Nozick (1974) about why we would dislike having a life like the Deceived Businessman’s: a life full of pleasures that are not based on truth or directly on reality. The reason he offers is that the deceived businessman was not *really* living the life that he had wanted to; his preferences were not being satisfied (Kagan 1998, p36). Kagan (1998) concludes that, since we value a life where our preferences are really satisfied over one in which we only think they are, the truth or reality of our experiences affects the value of them. Clearly, this conclusion, if correct, supports Premise 2 of the Argument from False Pleasures because it claims that pleasure (a type of experience) is less valuable if based on falsity or something like it.

Both of these two anti-hedonistic champions, the Experience Machine and the Deceived Businessman thought experiments, evoke powerful intuitions, which are generally thought to provide strong *prima facie* evidence for Premise 2 of the Argument from False Pleasures. Indeed, having identified these two thought experiments as the main reason for hedonism about well-being's current unpopularity, and therefore philosophical death, the remainder of this essay is dedicated to constructing an argument for why they are not the 'knock down' refutations that they are thought to be. The success of this argument will be an important step towards reviving Hedonism about Well-Being. However, before leaping to the defense of Hedonism about Well-Being, the current method of assessing theories of well-being needs to be discussed, as it is in need of re-evaluation. An argument will be presented that the current method of assessing theories of well-being puts too much emphasis on the role of intuitions and that a restricted role for intuitions could help Hedonism about Well-Being stave off the assault from the Experience Machine and the Deceived Businessman thought experiments. The current method of assessing theories of well-being is discussed below. However, it is henceforth referred to as the Old Method of Assessing Theories of Well-Being because one of the aims of this essay is to propose a new method of assessing theories of well-being.

## **The Old Method of Assessing Theories of Well-Being**

It is unusual for moral philosophers, and especially applied ethicists, to be specific about how moral theories should be evaluated, and yet nearly all of them frequently do just that. The justification and evaluation of moral theories is a controversial topic; although, nearly all philosophers seem to approach the matter in a similar way, including many of those who claim to be doing it differently (Kagan 1998, p16). Examples from three authors who have indicated how ethical theories, including theories of well-being, should be evaluated are discussed below and formalised into



the Old Method of Assessing Theories of Well-Being. Special attention is then paid to the role of intuitions in the Old Method of Assessing Theories of Well-Being because Chapter 2 of this essay argues for that role to be restricted.

In the first chapter of his book, *Conduct and Character: Readings in Moral Theory*, Mark Timmons (2006, p8-13) lays out a clear and concise methodology for evaluating moral theories before going on to discuss the major ones. He posits six evaluative criteria and explains that the extent to which moral theories can meet all of the criteria, and how they score on the criteria compared to competing theories, indicates how good a particular moral theory is; how successful it is at fulfilling the practical and theoretical aims that we hold for moral theories (Timmons 2006, p12-13). The criteria Timmons (2006, p9-12) posits are as follows:

- Criterion 1: Consistency
- Criterion 2: Determinacy
- Criterion 3: Applicability
- Criterion 4: Internal Support
- Criterion 5: External Support
- Criterion 6: Explanatory Power

Timmons' (2006, p9)<sup>28</sup> Consistency criterion stipulates that the underlying principles of a theory should not contradict each other or produce results that are inconsistent with each other, such as concluding that a particular action is both right and wrong simultaneously. His Determinacy criterion asserts that moral theories should be able to produce answers (as opposed to drawing blanks) about the majority of moral problems when the relevant facts are available (p9-10). The Applicability criterion requires moral theories to be applicable, and therefore useful, in real life situations (p10). To be applicable, the relevant facts that a theory needs in order to make its moral evaluations should be at least theoretically obtainable. As an example, Timmons (p10) points to the difficulty in ascertaining the likely net consequences for the happiness of all involved, information that hedonistic utilitarianism requires to

---

<sup>28</sup> All further in text citations in this paragraph are to (Timmons 2006).

make moral evaluations.<sup>29</sup> Timmons' (p10-11) Internal Support criterion asks that the evaluations moral theories make are consistent with widely-held "*considered moral beliefs*" (p11, his italics), which are exemplified by beliefs such as "torture is wrong" (p10) and are presumably formed by reflecting upon our moral intuitions.<sup>30</sup>

The External Support criterion stipulates that to the extent to which a moral theory is supported by, or refuted by, non-moral areas of knowledge makes the theory more or less credible (respectively). Timmons (2006, p11) cites Mackie (1977, p203) as noting that ethics overlaps with, and therefore affects and is affected by, "psychology, metaphysics and religion", amongst other disciplines, and suggests that the correctness of a moral theory is partly dependant on a lack of conflict with or, preferably, endorsement from such areas of non-moral enquiry. Timmons' (2006, p12) final criterion, Explanatory Power, requires that the underlying principles of a moral theory must explain the theory and why it is normatively significant in the area that it is used to make moral evaluations. For example, an explanatorily powerful theory of well-being would have underlying principles that clearly explain itself and offer compelling reasons for why the evaluations it makes are relevant to well-being. The clarity of Timmons' evaluative criteria make them useful, although he admits the potential difficulty in using his framework for ranking moral theories because any particular theory might outperform others on some of the criteria while not performing as well on others (2006, p13).

Shelly Kagan (1998, p11-17) also discusses how moral theories should be evaluated, although his methodology is not divided to the extent that Timmons' (2006) is. Kagan (1998, p11) begins his explanation of how he thinks moral theories, including theories of well-being, should be evaluated by noting that defenses of moral theories are not the same as defenses of scientific theories, most of which are based solely on

---

<sup>29</sup> Indeed, measuring, let alone forecasting, happiness is a considerable practical problem for hedonistic utilitarianism; the issue certainly deserves more attention.

<sup>30</sup> More on the relationship between moral intuitions and moral judgments comes later in this section.

empirical evidence. He acknowledges that scientific evidence can help with the factual information needed to make moral evaluations, but avers: “almost everyone agrees” that empirical evidence alone cannot justify fundamental moral principles (Kagan 1998, p12). Kagan (1998, p12-13) goes on to assert that two main types of reasons can be given to believe a moral theory is correct or incorrect: whether the theory produces evaluations that agree with our intuitions and how plausible the rationale of the theory is.

Kagan’s (1998, p13) first evaluative criteria is that a moral theory produces evaluations that fit with our moral intuitions – which he describes as the “immediate reactions” to thought experiments or real life situations. This criterion is similar to Timmons’ (2006) Internal Support criterion, except that Kagan (1998) does not place nearly as much emphasis on the intuitions being reflected upon as Timmons does. Kagan (1998, p13) takes care to emphasise the strength of some moral intuitions about pared-down thought experiments and claims that our inability to disregard our intuitions about these hypothetical scenarios shows how important they are for validating moral theories. Kagan (1998, p13) does go on to admit, though, that our moral intuitions can contradict themselves, and so must be biased sometimes. Indeed, this concession leads him to conclude that we may have to accept a moral theory that clashes with some of our intuitions. However, he maintains that showing how a moral theory fits with the majority of our intuitions is a “significant aspect of defending it” and that, most philosophers would agree (Kagan 1998, p13-14). Indeed, Kagan (1998, p16) believes that even philosophers who claim that they do not justify moral theories in this way usually do so (an example of which is discussed below).

Kagan’s (1998, p14-15) second evaluative criteria is that a moral theory has a “plausible *rationale*” (his italics). Generally, this criterion requires that there is good reason to believe that the theory correctly explains all of the moral features that it aims to account for. For example, the rationale of a moral theory should sensibly

explain things like why a particular principle is morally important and how important it is in any given situation. Kagan (1998) also implies something like Timmons' (2006) criteria of Consistency and Determinacy as being constitutive of a plausible rationale. This is mainly because the more complete a moral theory is, the more plausible it appears. One particularly important area of completeness for Kagan (1998) is that a plausible moral theory's rationale should have a consistent metaethical justification. This is important for Kagan (1998, p15) because he observes that the plausibility we confer on the reasons for believing that a moral theory sensibly explains its principles is considerably affected by our own "conception of the nature of morality". At this point, Kagan notes that our judgements are also affected by our intuitions and are ultimately formed by a reflective equilibrium (discussed below), as we attempt to keep our various intuitions and considered judgements consistent. Indeed, after taking care to separate his two evaluative criteria, Kagan (1998, p15) concludes by claiming that the most plausible moral theory will provide "the best overall fit with our various considered judgments", which allows moral intuitions to play a major role in evaluating moral theories.

Should it be the case, however, that moral intuitions should play such an important role? In his description of the criteria that should be used for evaluating moral theories, James Griffin (1986) chastises the major role intuitions often play, before hypocritically going on to use them himself in a cavalier dismissal of hedonism about well-being. Griffin's (1986) discussion of how to evaluate moral theories is perhaps the most relevant to this essay because it resides at the start of his influential *Well-Being: Its Meaning, Measurement, and Moral Importance*, a book dedicated to establishing a complete account of well-being (which entails evaluating theories of well-being throughout). Griffin (1986, p1) opens the discussion with the observation, similar to the one made by Kagan (1998), that "piecemeal appeal to intuition" is the "most common" method for evaluating moral theories, before turning his hand to the derision of intuition's normative significance. Griffin (1986, p1-2) clearly states that

“intuitions as a class have no probative force” and that even the most “intuitively repugnant” conclusion cannot settle a philosophical debate; it only offers “a reason to start looking for a good argument”. Despite the apparent clarity of Griffin’s (1986) stance on this issue, he goes on to make what can only be described as contradictory statements and he routinely fails to follow his own advice.

Although he thinks their number is small, Griffin (1986, p1-2) heralds certain intuitions as being “solidier than anything else that moral thought is likely to come up with”, such as that “battering babies is wrong”, presumably predicting that moral thought will never be able to plausibly justify such an act in anything like normal circumstances. Griffin may well be right about our inability to plausibly justify such an act, although it appears that we could easily justify *not* performing the act *without* the help of our intuitions on the matter. Regardless, Griffin (1986, p2) remarks that the few moral intuitions of such strength that we have are only useful for weeding out the “wildest of moral theories”. Griffin’s (1986) *stated* position on the use of intuitions for evaluating moral theories can be summed up as follows: intuitions have no normative significance, except for the few cases in which ridiculous moral theories bring about very strong negative intuitions. If this *is* Griffin’s position, then it is surprising that on the same page that he laments the rise of intuitions “so far above their epistemological station”, he also accuses two moderately tame (certainly not wild, at any rate) moral theories: contractarianism and utilitarianism, of being “counter-intuitive” (1986, p2) while providing an example of his Completeness criterion for evaluating moral theories (discussed below).

Furthermore, in the same 412-page book on well-being, Griffin (1986, p9-10) dismisses all internalist experiential accounts of well-being, including hedonism about well-being, with a couple of pages of discussion on and around the notoriously intuition-eliciting<sup>31</sup> Experience Machine thought experiment. Based on the negative

---

<sup>31</sup> See, for example, Sobel (2002, p244).

intuitive response most people feel to the thought of plugging into the Experience Machine, Griffin (1986) claims that, since we do not always desire mental states over other things in life, then mental state, or experiential accounts, of well-being must be wrong. He does admit, however, that by rationally investigating these preferences “with a cold eye”, we might find that they are “dubious”, but he fails to pursue the matter further, relying on a quick and dirty, intuitively supported rejection of hedonism about well-being (Griffin 1986, p9-10). Presumably, Griffin’s (1986, p9) “cold eye” method of evaluating moral theories is composed of the rational application of the two relatively non-intuitive criteria he had mentioned earlier; Completeness and Correctness (which are discussed below).

Griffin’s (1986, p2-3) first criterion for evaluating theories of well-being is Completeness and seems to incorporate three of Timmons’ (2006) criteria: Consistency, Determinacy and Internal Support. For Griffin (1986), the Completeness criterion basically requires that moral theories are able to adhere to their underlying principles when evaluating all cases without contradicting themselves, failing to provide a judgement, or producing results that do not fit with our firmly-held moral intuitions. Griffin (1986, p2-3) rightly notes that the stringency of this criterion will undermine the plausibility of many moral theories that we currently entertain as possibly correct.

The extreme rigidity of this criterion is best shown by the test Griffin (1986, p3) recommends that we use to evaluate how complete a moral theory really is. By spreading the moral theory in question as wide as possible, and especially into areas that it is likely to run into trouble, Griffin (1986, p3) believes that some theories will fail to show their Completeness by producing unintuitive results or abandoning one or more of their principles in order to produce more intuitively satisfying evaluations. Presumably, most of the places that moral theories run into trouble are in unrealistic thought experiments, such as the Experience Machine. Again, we can see that, despite

Griffin's (1986, p1-2) overtly-stated mistrust of intuitions, he recommends that solid intuitions should be used to evaluate moral theories, especially through the use of unrealistic thought experiments. Indeed, Griffin's (1986) contradictory stance provides evidence for Kagan's (1998) claim that even the authors who claim not to rely on intuitions when evaluating moral theories usually do so.

Griffin's (1986, p3) second criterion for evaluating theories of well-being is Correctness and is similar to the metaethical consistency requirement of Kagan's (1986) second criterion: having a plausible rationale. Griffin (1986, p 3) states that his second criterion is more important and even harder to satisfy than the first. He asserts that being believable or at least plausible, as most complete theories will be, is not enough (Griffin 1986, p 3). For Griffin (1986, p 3), the best moral theory is also correct, which means that its metaethical foundations are correct and that those foundations support the moral theory in the right kind of way. He acknowledges the difficulties of assessing compliance with this requirement and calls for more investigation into the relationship between meta- and normative ethics to help with this (Griffin 1986, p 4). However, Griffin (1986 p4) does suggest that the richest possible accounts of prudence ("everything that bears on one's self-interest"), morality, and how the two interact is the ideal starting point for this complex task. In the absence of a more detailed test for Correctness and a reliable methodology for discerning the truth of metaethical theories, Griffin's (1986) second criterion can be assumed to be a sort of metaethical consistency requirement.

Based on the evaluative frameworks described above and observing the evaluative strategies implicit in many authors' assessments of theories of well-being (e.g. Bagani & Fosl 2007; Griffin 1986; Kagan 1998; Kymlicka 1990; Rachels 2005), what follows is an interpretation and explanation of how theories of well-being have generally been assessed in the literature: The Old Method of Assessing Theories of Well-Being:

Theories of well-being should:

- i) Produce intuitively pleasing results in most cases
- ii) Have a compelling rationale, which can provide judgements in all relevant moral cases, while:
  - a) Being internally consistent
  - b) Producing results that fit with pervasive moral judgements (based on reflective equilibrium between our moral intuitions and other judgements)
  - c) Providing good reason to believe the normative significance of the theory
  - d) Being meta-ethically consistent
  - e) Being consistent with science
- iii) Be functional

And the extent to which a theory of well-being achieves these criteria indicates how good a theory it is.

The first two criteria of the Old Method of Assessing Theories of Well-Being get by far the most attention in the literature and the first, being intuitively pleasing, has often taken the role of ‘decisive argument/objection’ in damning evaluations of theories of well-being. Indeed, whether used in isolation of reason, or used under the guise of reflective equilibrium<sup>32</sup> as the main determinant of how compelling a theory’s rationale is, moral intuitions often play a major role in the literature when it comes to evaluating theories of well-being. Furthermore, this is especially the case when the arguments provided for a negative evaluation are not made in great detail (e.g. Bagani & Fosl 2007; Griffin 1986; Kagan 1998; Kymlicka 1990; Rachels 2005). Exactly how our moral intuitions are supposed to help evaluate theories of well-being is discussed below.

As for the compelling rationale criterion, different reasons are often given to demonstrate why a particular theory of well-being does or does not have a compelling rationale, presumably depending on how such tests would help or hinder the particular author’s intended appraisal of the theory. Each test for how rationally compelling a theory of well-being is has been discussed above and so will not be re-explained here. That is, except that Mackie’s (1977) requirement of compliance with Metaphysical and Religious beliefs, which is, for obvious reasons, scrapped in favour of the compliance with current scientific knowledge criterion. The last criterion,

---

<sup>32</sup> Discussed in more detail below.



functionality, closely mirrors 'Timmons' (2006) Applicability criterion. These other factors are rarely used as objections and are, therefore, not discussed further here.

Since the use of moral intuitions is so pervasive in evaluating theories of well-being, and since the extent of this use is what this essay will argue against later, how intuitions are supposed to help evaluate moral principles will now be explained in more detail. Rachels (2003, p10-12) discusses how our moral intuitions can have the same kind of authority that a "competent judge" (p11) might be considered to have in making moral assessments. He acknowledges that the normative significance of our intuitions is reliant on us having an "adequate grasp of their content" (p10), for which Rachels (2003, p11) considers having had relevant experience "may be crucial". Specifically on the topic of intuiting intrinsic value, such as intuitively assessing the intrinsic value of pleasure for well-being, Rachels (2003) refers to various authors and describes several ways that we might be able to acquire an adequate grasp on the content of our moral intuitions. Presumably, Rachels (2003) thinks that this better understanding of our moral intuitions' credentials will reassure us that our intuitions have the moral authority of a competent judge. The methods for understanding the content of our intuitions that Rachels (2003) refers to, however, are mainly through general agreement across subjects and situations and through introspection, and how these types of methods lend authority to our moral intuitions is not exactly clear.

One method for understanding the content of our intuitive evaluations about potential sources of intrinsic value, which Rachels (2003, p11) considers to have "limited application", is simply by observing a potential source of intrinsic value in various subjects and situations and noting the intuitive response. Rachels (2003, p11) cites Lewis (1989, p123) as suggesting that this method is not useful because accurate observations of this kind are difficult when they are of other people's minds and especially in mundane situations, in which potentially-observable reactions are unusual. Rather, Lewis (1989, p122; c.f. Rachels 2003, p11) recommends reflecting on

our own intuitive responses to situations that are as dramatic as possible, while adding and subtracting the potential source of value, so that the difference can be better perceived. When applied to thought experiments too (instead of just experienced situations), this appears to be the approach most generally used in ethics, with the perceived difference identified as the intrinsic value of the object that was added and subtracted in the thought experiments.<sup>33</sup>

Rachels (2003, p11) goes on to cite authors using a more direct method to grasp the content of moral intuitions: through introspection. Audi (1996, p112-114; c.f. Rachels 2003, p11) proposed that introspectively reflecting on evaluative propositions, such as ‘pleasure is good’, creates normatively significant conclusions about the accuracy of such propositions. And, Moore (1903, p91 & 223; c.f. Rachels 2003) claimed that reflecting on potential sources of value in complete isolation is “the only” safe method for determining the intrinsic value of something. Although Rachels (2003) notes that each of these methods is an attempt to add the authority of ‘competent judge’ to our moral intuitions, it is not obvious that any of them really achieve this goal because we cannot so easily understand the content of our intuitions through introspection, a point that is elaborated on in great detail in Chapter 2.

Regardless of the exact method various philosophers endorse for grasping the content of our intuitions, many of them are very confident about the meaning of at least some of their more solid moral intuitions. Indeed, Rachels (2003, p10) hand picked some “fine” philosophers that, like Griffin (1986), were inclined to rely heavily on at least some of their moral intuitions: Ross (1930, p29) appealed to the obvious rightness of promise-keeping to anyone of sufficient maturity, Lewis (1946, p375) considered our intuitive evaluations of good and bad to be correct, Nagel (1986, p146) recommended that after intuiting a response about what is valuable we should then move to find the best explanation for that intuition, rather than questioning it,

---

<sup>33</sup> See, for example, the Deceived Businessman thought experiment (Kagan 1998, p34-36).

and, according to Hare (1973, p148), Rawls (1971, p18, line 9 to p20, line 9) relied on intuitions to support 30 moral propositions in just two pages. Although, Rawls (1971) would claim to have reached these evaluative judgements through a process of reflective equilibrium, in which his reflected-upon moral intuitions would interplay with his beliefs and other intuitions (Daniels 1996), it seems that his intuitions were playing *at least* a major role in those 30 evaluations.

Indeed, even where Rawls (1971, p49; 1974/5, p8) claimed to be using a process of wide reflective equilibrium, in which opposing moral theories are also taken into account (Daniels 2003), to reach moral judgements, his intuitive foundations appear to be doing most of the evaluative work. Some proponents of wide reflective equilibrium have argued that the foundational intuitions involved are not beyond revision (Daniels 1979; 2003). However, it is not clear that our intuitions about what is morally right and wrong change very easily (Myers 2004), as argued in Chapter 2, or that philosophers very frequently remove intuitions from consideration in the face of the reason-based evidence provided by an opponents' moral theory. The important point here is that proponents of wide reflective equilibrium, and other more extreme forms of intuitionism, agree that moral intuitions have non-trivial normative significance for moral judgements.

Rachels' (2003) revealing of the fact that many fine philosophers use intuitions to discern the value of things, however, does not necessarily amount to a good reason for using intuitions for that purpose. Just because moral intuitions are used by fine philosophers does not necessarily make it right to use them, as is clearly shown by the fact that the intuitions held by various fine philosophers, and people generally, can and do contradict each other. Indeed, even if we reflected on and accordingly adjusted our moral intuitions in light of our beliefs and other intuitions about what has value, we would still find ourselves with intuitions that clashed with the reflected-upon intuitions of others. Therefore, as Rachels (2003, p12) concedes, the fact that

people confidently hold moral intuitions that contradict those of their contemporaries means that “many of us confidently believe falsehoods”. With this admission in mind, Rachels (2003, p12) asserts that, while we can never “confirm” the value of something with intuitions, widespread intuitive agreement is enough to be able to argue that “everyone agrees because the truth is obvious upon proper reflection”. Indeed, this claim does corroborate some of those made above about widespread intuitions containing reliable evaluative information. However, even if an intuition is universally held, it could still be unreliable (Liao 2007, p9), as discussed below.

Indeed, there are many examples which show that pervasive intuitions do not necessarily contain reliable evaluative information. History is replete with instances where even widespread and heavily reflected-upon intuitions were wrong. Furthermore, only extensive scientific evidence (that the sun orbits Earth and not vice-versa) or prolonged rational moral argument (that slavery is unfair) could change them. Indeed, because a heavy reliance on pervasive intuitions slows the progress of knowledge, it is not clear why moral philosophers do not reduce their use of intuitive judgements like the thinkers involved with the natural sciences and other branches of philosophy have (Griffin 1986, p2). Furthermore, if, as Kagan (1998) suggests, most moral philosophers are not intuitionists (who rely solely on intuitions) because they consider intuitions to be fallible, then how do they know when any of their particular intuitions are reliable? It seems clear that we do not achieve an adequate grasp on the content of our intuitions and, thereby, are not able to declare our intuitions competent judges solely because lots of intelligent people also hold them. It was also briefly mentioned above that we cannot introspect our intuitions to find out more about them. So what is the content of our intuitions like? How can we find out? And, what does that mean for the Old Method of Assessing Theories of Well-Being? While scientifically explaining the inner workings of our intuitions, Chapter 2 addresses and answers these questions at length.

## Conclusion

First, this chapter disentangled well-being from other accounts of the good life and provided a clear and reasonable definition of hedonism about well-being. Then a specific internalist account of hedonism about well-being (Hedonism about Well-Being) was endorsed. Following this, reasons for why most moral philosophers consider Hedonism about Well-Being to be a dead theory were discussed, thereby setting the scene for the rest of this essay, which is dedicated to taking steps towards reviving the plausibility of Hedonism about Well-Being and which ultimately leads to a suggested overhaul of how theories of well-being are evaluated. The argument that is currently doing the most damage to Hedonism about Well-Being's plausibility, the Argument from False Pleasures, was discussed. This discussion included the introduction of the two champions of the argument, the Experience Machine and the Deceived Businessman thought experiments, both of which elicit strong intuitive responses.

Following this, how theories of well-being are evaluated by moral philosophers was discussed, with reference to the work of Timmons (2006), Kagan (1998) and Griffin (1986). The approach moral philosophers take when evaluating theories of well-being was then formalised into the Old Method of Assessing Theories of Well-Being, which was explained with a focus on the role intuitions play in the methodology. Finally, the role of intuitions, which had been noted throughout this chapter, was discussed in more detail. Special attention was paid to discerning what might give our moral intuitions the authority of a competent judge, including the possibility of reflective equilibrium. However, the reasons to think that our moral intuitions do have such authority that were offered by Rachels (2003) and others were found to be unsatisfactory, prompting the need for an in-depth investigation into exactly how our intuitions work, which is one of the aims of the next chapter.

---

## **Chapter 2:**

# **Towards a Better Understanding of Intuitions and the Role They Should Play in Evaluating Theories of Well-Being**

---

### **Introduction**

This chapter discusses psychological, neuroscientific and philosophical research on intuitions, and moral intuitions in particular, with the goal of providing a clear understanding of how our moral intuitions are formed and what that should mean for their use in evaluating theories of well-being. First, Woodward and Allman's (2007) account of moral intuitive cognition is explained in detail and corroborated by the research of other leaders in the field of cognitive science. Then, this clear understanding of how moral intuitions work is used to assess Woodward and Allman's (2007) recommendations for the appropriate use of moral intuitions. One of Woodward and Allman's (2007) recommendations is then adapted into an argument for restricting the role of intuitions in evaluating theories of well-being. The conclusion of this argument: that intuitions are an unreliable source of evidence for evaluating unrealistic well-being-related thought experiments and, therefore, should not be used if other types of evidence are available, is then discussed and supported. With this argument in mind, the Old Method of Assessing Theories of Well-Being is then found to be in need of at least slight amendment.

The New Method for Assessing Theories of Well-Being is then proposed and explained, including a discussion of how it deals with intuitions more appropriately than the Old Method of Assessing Theories of Well-Being does. This explanation includes a guide on how to apply the New Method for Assessing Theories of Well-Being and is followed by noting its implications for well-being-related thought experiments, such as the Deceived Businessman and Experience Machine, and theories of well-being in general.

## **How Moral Intuitions Really Work**

In a recent article, James Woodward, a philosopher, and John Allman, a neuroscientist, explain what intuitions are and discuss the role they think intuitions have had and should have in ethics (2007). They define intuition as a form of cognition in which many different data sources are simultaneously evaluated and compressed into a simple one-dimensional message that we feel as a visceral sensation or gut feeling (Woodward & Allman 2007, p13). It should be noted that Woodward and Allman's (2007) definition of moral intuition differs from that of some moral philosophers, who consider moral intuition to be a belief based on "careful observation" and "reason" (e.g. Crisp 2002, p71; Lillie 1955; Sidgwick 1907).<sup>34</sup> This description will be rejected in the discussion below and the view of moral intuition supported by many psychologists is argued for: a simple moral evaluation (from very good to very bad) that appears in the conscious mind without any awareness of having deliberated about it (Woodward & Allman 2007; Haidt 2001, p818). Woodward and Allman (2007, p5, p24) argue that a better understanding of how our intuitive cognition works should make us reject some widely-held views about the appropriate uses of intuitions in ethics and guide us to use our intuitions more effectively.

---

<sup>34</sup> That is not to say that our moral intuitions cannot be affected by repeated morally relevant deliberations prior to any particular intuition – just that the process by which the intuitions are created is an unconscious one.

The discussion and examples used by Woodward and Allman (2007) are all focussed on the use of intuitions for evaluating normative moral theories generally, as opposed to evaluating theories of well-being specifically. However, there is no reason that their arguments should carry any less weight when considering theories of well-being. Indeed, both normative theories generally, and theories of well-being specifically, perform an evaluative function (evaluating actions and lives respectively) that comes under the umbrella of intuitive social cognition. This section describes Woodward and Allman's (2007) explanation of intuitions, including a general discussion about how the nature of intuitive cognition affects how they can be usefully applied, and provides external support for their view from other leaders in the field of cognitive science.

Woodward and Allman (2007) provide a neurobiological account of intuitive cognition and contrast it with deliberative cognition, which is also widely used in ethics. They describe an intuition as the visceral sensation that results from a very fast, unconscious and probabilistic processing of many variables in parallel (Woodward & Allman 2007, p13) – a definition that is widely accepted in the cognitive sciences (Lieberman 2000; Myers 2004). Deliberative thought<sup>35</sup> is a much slower cognitive process that consciously uses inductive and deductive reasoning on very limited numbers of variables at a time (Woodward & Allman 2007, p13; Bruner 1960). Despite an abundance of studies showing the shortfalls of intuitive cognition and how deliberation can correct them,<sup>36</sup> both Lieberman (2000, p109) and Sosa (2006, p225) note that it is possible for intuitive cognition to outperform deliberation in some areas. Indeed, various social interactions are likely candidate situations for intuition to have more utility than deliberative thought. For example, when meeting someone new at a party we might not be consciously aware of the constant non-verbal communication (e.g. body language) taking place (Word, Zanna & Cooper

---

<sup>35</sup> Sometimes known as 'analytic' thought (e.g. Bruner 1960).

<sup>36</sup> For example, Lieberman cites the following: (Chaiken, Liberman, & Eagly, 1989; Denes-Raj & Epstein, 1994; Fiske & Neuberg, 1990; Gilbert, 1989; Langer, 1989; c.f. Liberman 2000, p109).



1974) but we often come away feeling that we liked the person even though they supported the wrong football team and did not appreciate the exquisite heavy metal music playing in the background. To understand why this is, we must understand how intuitions are formed, an understanding that Woodward and Allman (2007) believe will lead us to see the types of circumstances in which moral intuitions are likely to be useful and those in which they are not.

Woodward and Allman (2007) plausibly hypothesise that moral intuitions are a subset of social intuitions because of their similar developmental trajectories and their shared neurobiological substrates.<sup>37</sup> Intuitive social cognition has been empirically studied since at least Valentine's (1929) study of men and women's abilities to intuitively judge the character of strangers, providing plenty of evidence for more recent neurobiological analyses to discover how intuitive judgement actually works. Recent findings suggest that, despite our tip-of-the-iceberg conscious mind thinking that it is the cognitive 'be-all-and-end-all' (Myers 2004), the vast majority of our cognition happens without us being aware of it (Bargh & Chartrand 1999).

Unconscious intuitive learning, including intuitive evaluations of people and events, is constantly occurring in our brains, especially when we experience novel stimuli (Myers 2004).<sup>38</sup> When novel stimuli are encountered the brain runs probabilistic inference simulations based on the matches between all of its current stimuli and past experiences (Woodward & Allman 2007). Part of the simulation process involves the reward centre, which produces the message (the visceral sensation or feeling) that we become consciously aware of (Craig 2004; Critchley *et al.* 2004; c.f. Woodward & Allman 2007, p17). Presumably, whether the intuition is felt (consciously noticed), and how strongly it is felt, is based on the predicted results of the unconscious

---

<sup>37</sup> As shown by studies such as: Berthoz *et al.* (2002), de Quervain *et al.* (2004), Rilling *et al.* (2002), Sanfey *et al.* (2003), Shin *et al.* (2000), and Singer *et al.* (2004a, 2004b) (all c.f. Woodward & Allman 2007, p14).

<sup>38</sup> And, Woodward and Allman (2007, p15) cite Critchley *et al.* (2001), Singer *et al.* (2004b), Shin *et al.* (2000) and Berthoz *et al.* (2002) to show the degree of activity in the substrates linked to intuitive social cognition increases with the level of uncertainty in a situation.

simulation. We are effectively always performing unconscious pattern recognition by monitoring our current environment, comparing it to our archive of experiences, and then evaluating the probabilistic forecasts that result. Indeed, the fact that we unconsciously process potential future experiences explains many of the visceral sensations we experience everyday, such as the uneasy sinking feeling we experience as the overly-dramatic music in a cheesy horror movie indicates that a scary shock is just around the corner.

When this knowledge about intuitions is applied to the social context described above, we discover that, while our conscious mind was picking up relatively useless information for friend-evaluating, our intuitive cognition was recording much more useful information and doing it much faster. By tracking the stranger's body position, eye movement, vocal tones, facial expressions and other social cues, and automatically running them through our probabilistic experience-dependant processors, we can very quickly produce a simple but relatively accurate intuition about him or her (Zebrowitz & Collins 1997). The accuracy of this intuition will be based on the amount and type of our previous experiences with strangers and how this particular stranger fits in with those patterns.<sup>39</sup>

Probabilistic models always experience outliers, but the more data they gather, the more accurate their predictions will be (on average and assuming that nothing else dramatic happens that would affect all future experiences). This model of implicit learning has been experimentally proven to be how intuitive cognition works (Lieberman 2000),<sup>40</sup> which, according to Woodward and Allman (2007, p20), gives good reason to think that moral intuitions also work that way.<sup>41</sup> Therefore, we can usually expect any intuition we have in situations that we are very experienced in to

---

<sup>39</sup> Naturally, some people might also be genetically predisposed to be better or worse at certain types of pattern recognition or other processes which may affect the accuracy of their intuitions in most cases.

<sup>40</sup> Lieberman (2000, p110) cites Aslin, Saffran, and Newport (1998), Biederman and Shiffrar (1987), DeGroot (1965), Lunn (1948), and Saffran, Aslin, and Newport (1996).

<sup>41</sup> The link between experience and the accuracy of moral intuitions is discussed in more detail below.

be more trustworthy than our intuitions in completely novel situations. This also explains the intuition had by an experienced fireman, who saved his crew from a potentially fatal fiery fall. Klein (1998) describes the story of the lieutenant fireman who ordered his crew to flee from a kitchen with a small fire in it. Just after they left, the floor collapsed because of a large fire in the basement directly below where they were standing (Klein 1998). When asked what prompted his decision to leave the kitchen before the fire was out, the experienced lieutenant put it down to intuition (Klein 1998). Further questioning then unearthed the actual cues that his intuitive cognition had picked up on; that the small fire wasn't diminishing as quickly as expected and that the house was much hotter than the size of the visible fire could account for (Klein 1998). Had the experienced lieutenant not been there, the other fire-fighters might not have had the same feeling that something was wrong. Or, if they did, they might not have felt it strongly enough to warrant ordering a retreat.

Woodward and Allman (2007), and others,<sup>42</sup> also state that the intuitive/deliberative cognition distinction holds equally between moral intuitions and moral theorising, providing a useful way to think about the different ways that we can arrive at moral conclusions.<sup>43</sup> Therefore, we can expect that our moral intuitions are based on our previous experiences of situations involving moral emotions, moral utterances and any other morally relevant factors, such as harm or fairness. Woodward and Allman (2007) claim that all of the moral data we continuously collect, and unconsciously process, enables us to have insightful moral intuitions, although, only in situations with which we are familiar. They suggest that our intuitive cognition constantly processes our experiences, including morally relevant data, such as the emotional

---

<sup>42</sup> Greene (in press), Haidt (2001) and Haidt & Greene (2002) cited from (Woodward and Allman 2007, p17-18).

<sup>43</sup> It should be noted, however, that Mikhail (2007) has proposed a different account of how moral intuitions work, in which our intuitions are formed by a more deductive, rule-based process. Mikhail's (2007) legalistic account of moral intuition may have some explanatory use but it lacks the neurological evidence needed to strengthen its causal-explanatory power, which may explain why it has not proven popular yet. Furthermore, even if Mikhail's (2007) account of moral intuition proves to be an accurate causal account, the psychologically embedded fundamental moral rules it posits would only affect our moral intuitions and, thereby, our actual moral judgements, how we *should* judge situations would still be up for debate. This is a point that is discussed in more detail in Chapter 4.

states of ourselves and others, and can warn us if it forecasts that we are likely to transgress a moral norm. The warning comes in the form of that all-too-familiar ‘bad feeling’, the visceral sensation often associated with the gut that tells our conscious mind that something is wrong.

When we are asked to consider a thought experiment and we get that bad feeling in our gut, we immediately attribute it to the scenario we are imagining, and rightly so, because that is usually the only new stimuli we have just experienced. It would not make sense for that intuitive response to be caused by some unchanging stimulus that we had already been experiencing for some time. However, as revealed by a better understanding of intuitive cognition, just because an intuition is in response to a thought experiment does not mean that it is in response to the morally important issue ‘isolated’ within it by the experimenter.

The situations described in thought experiments will be matched with the participant’s most similar experiences by her intuitive cognition. Therefore, if when choosing between flicking switches to save people and pushing a fat man to his death to save people, then the pushing option is likely to elicit a more negative intuition. This is because the participant’s most similar experiences are about the times she has pushed someone. Indeed, this is exactly what Greene *et al.* (2001) found during an experiment in which they used neuroimaging equipment to measure participants’ neural activity while they were subjected to problems similar to those just discussed. They discovered that the thought of pushing someone activated an emotional part of the brain that the thought of flicking a switch did not. This finding helps explain why participants seem more reluctant to push someone than to flick a switch to kill them (and save five others). It seems safe to assume that most people have implicitly learnt that pushing people is likely to cause more negative consequences than switch-flicking. The pertinent questions here, however, are: how useful are our intuitions for evaluating moral thought experiments and real life situations, and, assuming that

some are and some are not, how can we discover which of our moral intuitions are the useful ones?

Woodward and Allman (2007) argue that our moral intuitions are probably more useful than we might think in certain situations, while being less useful in others. They claim that the usefulness of a specific intuition in any situation is predominantly based on our amount of experience with that type of situation.<sup>44</sup> If moral intuitions do track morally relevant data in situations, then the probabilistic nature of the process that brings them about should make those intuitions more accurate with increased exposure to relevant types of situations. An exception to this general rule might be if someone continuously acted against their moral intuitions, by continuing to murder people say. This would most likely result in them becoming desensitised to some extent, resulting in a weaker intuitive response to murder over time (Sanborn 2003). Exceptions such as this weaken Woodward and Allman's (2007) thesis about the benefits of experience for moral intuitions. However, this is not the only problem for the thesis that moral intuitions can be normatively significant.

More importantly, Woodward and Allman's (2007) findings imply that we may never be able to know for sure if any particular intuition about a thought experiment is actually of any use. In order to discover the usefulness of our intuitions for this task, we need to be sure that the intuition is in response to the moral issue in question and that it is responding correctly. Discerning whether an intuition is responding to the moral issue that it is supposed to be responding to is mainly dependent on our capacity to introspect or reconstruct our moral intuitions. And, discerning whether the intuition is responding correctly to the moral issue relies mainly on the various biases and restrictions that intuitive cognition is subject to (because of the probabilistic process by which it operates). Woodward and Allman (2007) make

---

<sup>44</sup> Presumably some people and perhaps some kinds of people also have a more reliable intuitive cognition than others because of certain skills or characteristics they have. For example, women are widely thought to have better intuition than men, which (very generally) seems to be the case in certain kinds of situations. Myers (2004 p45-50) and Lieberman (2000, p126) survey the literature on men's and women's intuition.

several claims related to the usefulness of moral intuitions for evaluating thought experiments. They claim that many factors can bias our intuitions, making them erroneous, that effective introspection of moral intuitions is “implausible” (2007, p25), that reconstruction of our intuitions will provide “at best limited insight” (2007, p24) and that, despite all of this, some moral intuitions could contain “useful information” and be “normatively defensible” (2007, p24). Their argument for each claim is explained and evaluated in detail below.

Woodward and Allman’s (2007) assertion that various factors make some of our intuitions erroneous is undoubtedly correct. Indeed, many authors have written on the pitfalls of intuitive cognition, citing screeds of anecdotal and experimental evidence. Even supporters of the uses of intuitive cognition point to anecdotal evidence that they have their limitations. For example, Philip Goldberg has astutely noted that we never hear anyone ask: “give me one good *feeling* why you think John is wrong” (1983, p18).<sup>45</sup> A recent experiment by Valdesolo and DeSteno (2006, p476-477) provides compelling evidence that our mood affects the moral judgements that we make without us being consciously aware of it, revealing the fickle nature of our moral intuitions. Myers (2004, p128) provides a more compelling body of empirical literature evidencing the perils of intuitive cognition, which he summarises as: “intuition’s dozen deadly sins”. Importantly, these ‘deadly sins’ include intuiting relationships where they don’t exist, intuitions being altered by various unrelated framing effects, intuitions being less sensitive to disconfirmations than they should be and our intuitions being highly self-serving. All of these factors could cause an intuition to become useless for moral purposes, as shown below.

First, consider the sin of intuiting a relationship bias that does not actually exist. For example, if our intuition about the morality of an action performed by a black-

---

<sup>45</sup> His italics.

skinned person was based on a racist bias,<sup>46</sup> instead of on the action itself, then the intuition would be useless (Kagan 1998, p14).<sup>47</sup> Regarding framing effects, philosophers' worries about the effects of wording and ordering of thought experiments on moral intuitions (e.g. Fischer & Ravizza 1991; c.f. Petrinovich & O'Neill 1996, p147) appear to be valid. Petrinovich and O'Neill's (1996) study found considerable framing effects from both wording and ordering in simple ethical thought experiments.<sup>48</sup> On intuitive resistance to disconfirmations, Myers (2004, p115-119) reports on several studies that show how experiences that disconfirm our established intuitions do not always weaken them, which explains why some people blame individuals for moral transgressions despite being exposed to evidence that various societal issues are really the cause (Gans 1988; c.f. Sasson 1995, p7). Now, consider self-serving bias: the phenomena that explains why practically everyone (even social psychologists who are well aware of this bias)<sup>49</sup> view themselves as more moral than the average person, despite that being mathematically impossible (Myers 2004, p95). If our intuitions about how moral *we* are are often erroneous, then we have good reason to believe that our intuitions about any moral matter in which we are involved might also be misguided. It is clear that our moral intuitions can be distorted by these biases and probably others as well, providing good reason to believe that some intuitions are very unreliable for evaluating moral situations and thought experiments.

Woodward and Allman (2007) also stress that it is impossible to introspect our intuitions – to gain any information about what is causing them by merely 'peering' into them. The fact that our intuitions do not come 'labelled' in any sense (even when we 'look closer') goes against the folk conception and the view of some philosophers

---

<sup>46</sup> "White Americans tend to conflate criminality with blackness" (Sasson 1995, p 6), despite the fact that it is impossible for 'blackness' to cause crimes.

<sup>47</sup> See Dasgupta *et al.* (2000) on the extent of race-related bias.

<sup>48</sup> It should be noted that Petrinovich and O'Neill (1996) also stress that framing effects can vary greatly depending on the context, an area that they think requires more research before any across the board generalisations about framing affects can be made.

<sup>49</sup> (van Lange, Taris & Vonk 1997)

on intuitions, despite being considered canonical by psychologists.<sup>50</sup> Various experiments, such as one performed by Lewicki (1986), reveal why psychologists are decided on the issue. Lewicki (1986) showed several photographs of people to participants, while informing them about the character of each. Lewicki then showed the participants further photographs and asked them to describe the characters of the people depicted. Despite the participants not knowing that Lewicki had correlated all of the types of characters with the hair length in each photograph, the participants generally identified the correct character type fairly successfully through implicit learning. However, when asked to reflect on their intuitive responses, the participants erroneously claimed that non-relevant features, such as the portrayed people's eyes, were the cause of their judgement. If intuitions were introspectible, then the participants in Lewicki's study would have answered along the lines of: 'I know it doesn't make sense, but that person's hair makes me feel like they're nasty!'

Similar results can be found with respect to moral intuitions. Jonathan Haidt (2001) reports on an experiment that he conducted with his colleagues<sup>51</sup> in which participants were asked if they felt incest to be morally permissible if birth control was used and nobody was harmed. Most participants immediately said that incest is wrong, even in those circumstances. However, when pushed to justify their moral judgement, most participants could not satisfactorily do so (Haidt 2001, p814). These participants had a definite moral intuition about the hypothetical incest case but they could not say what caused it. This is a predicament Haidt and colleagues termed 'moral dumbfounding' (Haidt 2001, p817), which could not occur if our moral intuitions are in fact accessible to introspection.<sup>52</sup> Since our moral intuitions (just like

---

<sup>50</sup> For example: Bastick (1982), Clinchy (1975), Griffin and Ross (1991), Hagafors and Brehmer (1983), Smith and DeCoster (1999) (all cited from Lieberman 2000, p110).

<sup>51</sup> Documented in an unpublished study: (Haidt, Bjorklund & Murphy 2000; c.f. Haidt 2001).

<sup>52</sup> Presumably, some participants provided justifications for their moral judgement against incest, such as 'God deems it so', which they perceived as entirely plausible. However, the purpose of Haidt's (2001) article is to provide good reason to believe that those justifications are nothing more than post hoc confabulations that were not the cause of the moral intuition. Therefore, while we may attempt "careful observation" of our moral intuitions as Sidgwick recommends (c.f. Crisp 2002), the result of such



all of our intuitions) do not, themselves, provide any information about what caused them, we cannot discover through introspection whether they are reacting to a specific moral stimuli or a morally-irrelevant bias, such as those discussed above.

Due to the various biases of intuitive cognition and the fact that we cannot introspect our intuitions for more information, Woodward and Allman (2007, p24) are also sceptical about the possibility of reliably reconstructing our moral intuitions. In Klein's (1998) fire-fighter example, the experienced lieutenant's negative intuition about the situation can be reconstructed to show, from the ground up, its likely causes. The abnormally warm room and unusually resistant fire were plausible causes of the experienced lieutenant's intuition to flee the kitchen. If the warm room and the resistant fire amount to a complete and rational explanation of the intuition, and in the absence of any rationally compelling competing hypotheses, then perhaps this particular reconstruction tells us that the intuition plausibly reflected the unconscious processing of those cues. This would indicate that the experienced lieutenant's intuition probably tracked the relevant features of the situation and interpreted them correctly, making that intuition very useful! The fact that the cues in this case were physical realities that could be reliably corroborated after the fact, allows us to be fairly confident about this reconstruction, although, because there is no direct measurable link from the cues to the intuition, we can still never be totally sure.<sup>53</sup>

However, when attempting to find the possible cues with which to reconstruct *moral* intuitions, we often have no physical realities that we know to be relevant, as in Haidt's (2001) incest case. In fact, we have to propose possible moral justifications (moral principles, such as 'incest is always wrong')<sup>54</sup> in an attempt to explain the intuitions, which is problematic because normative ethicists use those moral

---

introspection should not be dubbed 'moral intuition'. Rather, moral confabulation seems to better fit the bill.

<sup>53</sup> This point is explained in more detail in the paragraph directly below.

<sup>54</sup> The, perhaps, more apparent moral principle of 'it is wrong to procreate in a manner that is more likely to create a deformed child' is ruled out in Haidt's case by stipulations in the thought experiment.

intuitions to evaluate moral principles.<sup>55</sup> Any attempt to reconstruct a particular case will involve assuming that the individual concerned has a particular implicitly learned moral heuristic. And, that is a big assumption! Just because a moral principle could explain a moral intuition in one situation does not mean that other moral principles could not also explain that intuition while simultaneously disconfirming a moral intuition about another situation that the first moral principle endorses. Such a result, two examples of which are given in Chapter 3, casts considerable doubt on the existence of specific moral heuristics. Therefore, attempts at one-off reconstruction of moral intuitions are subject to a vicious chicken and egg problem, which makes it very difficult to reconstruct a moral intuition without heavily biasing the outcome. Indeed, we can rightly ask: do we actually have unconscious but firmly believed moral principles that cause our moral intuitions or do we just judge that various moral principles happen to cohere with our moral intuitions in specific cases? In order to answer this question, not only would a wide range of moral situations need to be investigated, each situation would have to also be subject to multiple minor variations to see what kind of moral heuristic or principle, if any, was being consistently and unconsciously applied. This kind of investigation would be a large project but one that has nevertheless been begun by moral psychologists.

Naturally, it is possible that moral intuitions are in fact caused by one or more implicitly learned heuristics that might be very similar in form to moral principles. Studies by moral psychologists attempting to discern the causal mechanisms of moral intuitions are starting to be published (e.g. Cushman, Young & Hauser 2006; Haidt & Joseph 2007; Hauser *et al.* 2007). However, the results are only statistical inferences (which, in individual cases, are limited to providing *likely* causes) and are based on some *a priori* assumptions about the moral and non-moral differences in abstract thought experiments. Essentially, the authors of these studies, just like many moral

---

<sup>55</sup> As discussed earlier, in Chapter 1.

philosophers,<sup>56</sup> seem to assume that their sparse thought experiments (which are supposedly only differentiated by the moral issue in question) are engaged with *literally*. Unfortunately, any normally functioning person will engage in some sort of intuitive cognising when they imagine a thought experiment, which is likely to result in them experiencing an intuition about it that was probably not the result of having followed all of the stipulations in the thought experiment entirely literally.

Recall the discussion from above, which explained why most experimental participants had the intuition that pushing a fat man to kill him and save five others was morally wrong while they deemed flicking a switch to kill him and save five others as at least morally permissible. It seems that the participants' unconscious processing of the scenario involved probabilistic inferences about other likely consequences of pushing the man to his death. Consequences such as the usual responses of others to such purposefully-harmful actions that the participants have experienced in the past. Unbeknownst to the participants, their intuitive cognition probably filled in the 'meat' of this 'bare-bones' thought experiment, making it more realistic but producing an intuition based on more than just its literal components (which had been specifically chosen to isolate the moral issue). It is this facet of our intuitive cognition that makes it the most difficult to reconstruct moral intuitions because even if we experimentally isolate the moral issue, we cannot prevent our unconscious mind from processing the thought experiments in the context of screeds of other trivially related experiences.<sup>57</sup>

It seems like the only way to avoid this subconscious adding of irrelevant information would be to do extensive real-time neurological and psychological testing on

---

<sup>56</sup> For example, any thought experiment mentioning a 'trolley bus' and "Judith Thomson's (1971) case of spores that become attached to furniture and grow into people, Michael Tooley's (1972) example of pills that turn kittens into human babies, and so on" (c.f. Woodward & Allman 2007, p24).

<sup>57</sup> Indeed, I wonder whether 'context-free' moral dilemmas, such as 'is it better for 3 people to die than 4?', elicit no intuitive response at all, leaving deliberative cognition to produce a purely reason-based moral judgement about the benefits of 4 lives over 3.

participants that were performing immoral and morally suspect tasks. However, this avenue of research is unlikely to get the chance to help because of the existing moral beliefs of university human ethics committees. Furthermore, recall the benefits experience provides for intuitive cognition, discussed above. Even if having lots of experience with a particular moral issue does not in fact make our moral intuitions any more reliable, having no experience with a moral issue or anything even vaguely similar will certainly make the intuition much less reliable. If any supposedly morally important feature of a thought experiment was completely novel to a participant, then any intuitive response of that participant may not have taken that feature into account, due to an inability to match it to any prior experience. Alternatively, the participants' intuitive cognition could have incorrectly coded the feature by matching it to a similar, but morally irrelevant, past experience. A good example of this occurrence is discussed in Chapter 3 during the analysis of the intuitive evidence supposedly provided by the Deceived Businessman thought experiment. All things considered then, even if we attempted to control for identifiable biases, without the benefits of introspection or dramatic (and theoretically unlikely) advances in neuroscience, we will never be sure that our attempt to reconstruct a moral intuition has been successful.

So, it seems that there are plenty of good reasons to conclude, as Woodward and Allman (2007) do, that intuitions are not accessible via direct introspection, that they are very unlikely to be able to be reconstructed in a reliable way, and that they are susceptible to various biases that would make them inappropriate to use when evaluating unrealistic thought experiments. However, Woodward and Allman (2007, p24) also claimed that moral intuitions caused by experiments can be useful and, therefore, "normatively defensible". They correctly point out that, despite all of the problems with intuitions discussed above, any moral intuition *could* be responding to the morally relevant feature (and nothing else) in a thought experiment. Furthermore, Woodward and Allman (2007) point out the widely-reported powers of intuitive

cognition and particularly social intuitions<sup>58</sup> in certain situations. Indeed, they claim that moral intuitions could usefully reflect morally relevant information about a situation and that they probably do in certain conditions.

In the event of these conditions (complexity and relevant experience – discussed in more detail below) pertaining, Woodward and Allman (2007) believe that moral intuitions can be usefully used in combination with reason to evaluate thought experiments or various other situations. Indeed, they aver it would be wasteful to totally disregard intuitions about the right kind of moral situations because of the often-hidden powers of intuitive moral cognition, particularly the power to process many variables simultaneously and consider relevant issues that had not necessarily been disclosed to the conscious mind. Hence, Woodward and Allman (2007, p27) state that the normative significance of moral intuitions will have to be identified on a case-by-case basis and is entirely dependant on the presence of the relevant conditions. Based on the implications of these findings, Woodward and Allman (2007) provide some general advice for when we should and should not use intuitions to help evaluate moral situations and thought experiments, correlating with how appropriate they consider the intuitions to be in various types of cases. Their advice is discussed and expanded on below.

## **Recommendations for the Use of Intuitions in Evaluating Theories of Well-Being**

This section discusses and then expands upon the implications of Woodward and Allman's (2007) findings about intuitions as they see them. First, Woodward and Allman's (2007) warnings of when not to use intuitions in moral theorising are discussed. The warnings are followed by Woodward and Allman's (2007) recommendations for appropriate use of moral intuitions. These recommendations

---

<sup>58</sup> As confirmed by a wealth of studies presented by Myers (2004, chap 2).

are then incorporated into a formal argument for reducing the role of intuitive evidence from unrealistic thought experiments in evaluating theories of well-being.

Woodward and Allman (2007) stress two main ways in which they consider intuitions to be misused by moral philosophers: relying solely on moral intuitions to evaluate thought experiments and using moral intuitions to help evaluate unrealistic thought experiments. Woodward and Allman (2007) advise that relying solely on moral intuitions to evaluate thought experiments is a bad idea because it is likely to produce a result heavily marred by the many potential biases intuitions are subject to. This recommendation seems fair. However, because such an approach is rare when evaluating theories of well-being, it will not be discussed further in this essay.

Woodward and Allman (2007) also advise against the use of intuitions to help evaluate unrealistic thought experiments because intuitive cognition generally processes them unreliably. Ethics courses and articles are riddled with extraordinary thought experiments, some so unrealistic and bizarre that one wonders how they were ever concocted in the first place. Woodward and Allman (2007, p24) cite Judith Thomson's (1971) scenario of spores attaching to furniture and becoming people and Michael Tooley's (1972) example of pills that can transform kittens into human babies as quintessential examples of thought experiments that are so unlikely that they verge on the impossible. Woodward and Allman's (2007) problem with extraordinary scenarios like these is that no one has ever experienced anything like the relevant parts of them and so our intuitive cognition has very little to go on when processing them. As discussed above, any intuitive response that we draw from them is likely to be based on the few supposedly morally irrelevant features that we are actually familiar with, such as the nuisance of mould on furniture.

Thought experiments can also be unrealistic in the way that they ask us to imagine implausible features of situations with which we may be much more familiar. For

example, Bernard Williams's (1973, p221-222) Jim in the Jungle thought experiment asks us to imagine that the evil Pedro and his troops will kill twenty villagers unless Jim the hapless explorer kills one of them. The implausible features of this thought experiment are that it expects us to: completely rule out any alternative to the two options provided by Pedro, to believe Pedro (the crazy bad guy) is totally true to his word, and to disregard the non-immediate future implications of our choice.<sup>59</sup> The only way most of us can relate to this scenario is vicariously, through the actions of characters like John Rambo. And, let's face it, unlike wimpy Jim, Rambo would have disarmed Pedro, incapacitated a dozen of his troops and gotten the villagers to safety (all while incurring only several minor bullet wounds). As discussed above, our intuitive cognition cannot always follow word-for-word the stipulated rules of thought experiments, such as disallowing heroics. Our intuitive cognition unconsciously matches what it can from the thought experiment to our past experiences noting any correlations that arise, regardless of whether or not those experiences lay outside of the scope of the unrealistic stipulations of the scenario.

Woodward and Allman's (2007) well-reasoned admonition of the use of intuitions to evaluate unrealistic thought experiments is an important one because not only does it bring the findings of many moral philosophers into doubt, it also challenges the methodology currently used by psychologists who are trying to unlock the secrets of moral intuition. Hauser and colleagues (2007, p3) followed the lead of a procession of researchers<sup>60</sup> who stick to "artificial dilemmas as opposed to real world cases", citing reasons that the discussion above refutes. They believe that artificial cases eliminate bias because participants will have "no familiarity with or personal attachment to" the details of the case (Hauser *et al.* 2007, p4). While they are right to assert that certain group-based biases can be eliminated by having only anonymous characters in

---

<sup>59</sup> Interestingly, Williams (1973, p221) is aware of the difficulties of using thought experiments, such as those discussed in the body text, but he considers sufficiently-detailed scenarios to be able to at least point to potential problems.

<sup>60</sup> (O'Neill & Petrinovich 1998; Petrinovich, O'Neill & Jorgensen 1993; Mikhail *et al.* 1998; Greene *et al.* 2001; Greene *et al.* 2004; c.f. Hauser *et al.* 2007, p3).

thought experiments (Hauser *et al.* 2007, p4), this could also be achieved in realistic scenarios through better sampling techniques. The real problem here is participants' lack of familiarity with the general details of the case, which stretches the capabilities of their intuitive cognition to the point where the probabilistic processing will draw in irrelevant but nevertheless correlated experiences.

Furthermore, Hauser and colleagues (2007, p4) claim that the artificiality of the scenarios allows them to change one of the few extant features to isolate morally important issues. This claim is based on a poor understanding of intuitive cognition. Recall the discussion above, which explained how our intuitive cognition implicitly adds 'meat' to 'bare bones' thought experiments based on our past experiences. So, while Hauser and the other experimenters might think they are isolating particular features, they might actually be tapping in to a new set of details that the participants' intuitive processes now deem relevant. If faced with this admonition, these authors would likely demand to know in what situations Woodward and Allman would propose that moral intuitions are appropriate to use.<sup>61</sup>

Woodward and Allman's (2007) recommendations for when moral intuitions should be used are based on the relevant experience of the participant, and the complexity of the scenario, in question. Based on the accuracy advantages that implicit learning gleans from relevant experience, Woodward and Allman (2007) conclude that the moral intuitions of people with similar experience to a hypothesised scenario will be much more useful for evaluating it than the intuitions of other, less-experienced, people. Just as Klein's (1998) experienced fire-fighter had a useful intuition that his less-experienced colleagues did not, Woodward and Allman (2007) suggest that people with first-hand experience of torture would have more useful intuitions about it than the top generals and politicians who have only read the odd report about it.

---

<sup>61</sup> Other advice might be to use neuroimaging and self-reports on participants in real-life situations that were easy to replicate with high fidelity. This is a technique similar to one already attempted by Montague *et al.* (2002) and is recommended by Casebeer and Churchland (2003) and hinted at in Casebeer (2003).



Interestingly, this claim of Woodward and Allman (2007) implies that there could be ‘moral experts’, with lots of experience in specific moral issues, whose intuitions could be particularly normatively significant in their specialised area. It is not clear that this particular conclusion of Woodward and Allman’s (2007) is correct, especially in situations where desensitisation to immoral acts due to repeated exposure to them might occur (Sanborn 2003). Indeed, more research on this issue would be fruitful.

Woodward and Allman (2007) also aver that the more complex the scenario under evaluation is, the more useful our intuitive responses will be. They base this claim on the distinctively different capabilities of intuitive moral cognition and moral reasoning, or deliberation, discussed above. Moral reasoning is a deliberative cognitive process that can only process a few variables at a time. This makes it incredibly slow at processing large amounts of data because it has to find a way to amalgamate and order the variables to be able to take them all into account at once (Lieberman 2000). Intuitive moral cognition, on the other hand, has much more ‘brain power’ at its disposal (Myers 2004), allowing it to process large quantities of data in parallel and integrate them into a simple felt judgement very quickly (Dijksterhuis *et al.* 2006). Woodward and Allman (2007) point out that reality’s natural combination of complexity and being commonly experienced makes it the prime candidate situation for our moral intuitions to have normative significance. Presumably, Woodward and Allman would also confer normative significance to the intuitions relevantly experienced participants have about realistic thought experiments. If Woodward and Allman’s (2007) conclusions about more and less appropriate uses of intuitions when evaluating thought experiments are correct, then the Old Method of Assessing Theories of Well-Being (discussed in Chapter 1), seems in need of at least minor amendment.

The most pertinent conclusion from Woodward and Allman’s (2007) research for evaluating Hedonism about Well-Being is that intuitions about unrealistic thought

experiments are likely to be unreliable; a conclusion that will now be adapted to show that the Old Method of Assessing Theories of Well-Being is definitely in need of at least minor amendment. The following argument draws heavily on the discussion above, and provides well-being specific examples, to make a very strong case in support of the conclusion that intuitions about unrealistic well-being-related thought experiments are unreliable and should only be used when no other reliable form of evidence is available.<sup>62</sup>

- P1) Intuitions about unrealistic well-being-related thought experiments are often misleading.
  - P2) There is no reliable method for knowing when intuitions about unrealistic well-being-related thought experiments are or are not misleading.
  - P3) If some form of evidence is often misleading and there is no reliable method for knowing when that form of evidence is or is not misleading, then that form of evidence is unreliable and should only be used when no other reliable form of evidence is available.
- 
- c) Therefore, intuitions about unrealistic well-being-related thought experiments are unreliable and should only be used when no other reliable form of evidence is available.

In support of Premise 1, the considerable evidence that intuitions can be misleading because of the influence of one or more unconscious bias that was discussed above also applies to intuitions about unrealistic well-being-related thought experiments. These biases can make our intuitive judgements of unrealistic thought experiments misleading in several ways. Our intuition may not be in response to the morally-relevant aspect of the thought experiment; it might be in response to something not stipulated, or it might even be in response to something specifically stipulated against in the thought experiment. Perhaps most commonly, though, our intuition will be in response to the morally-relevant aspect of the thought experiment *and* other aspects that are considered irrelevant, but which might have even more impact on the intuition than the relevant aspect. Furthermore, even when our intuitions are in response to the morally-relevant aspect of an unrealistic thought experiment, and

---

<sup>62</sup> While the argument is worded to specifically relate to theories of well-being, it could easily be amended to reach a similar conclusion for moral theorising generally.

only or at least mainly that aspect, they still might be responding in a misleading way because of any number of these duplicitous implicitly learned biases.

Despite the great care exercised by some moral philosophers in creating their thought experiments, the intuitions they elicit could be in response to some other supposedly irrelevant aspect of the scenario, as opposed to the particular aspect that they have attempted to isolate. These types of unreliable responses can occur because of our unconscious processing of irrelevant or non-existent aspects of thought experiments; the adding-meat-to-bare-bones feature of intuitive cognition discussed above. Presumably, Woodward and Allman (2007) would want to add at this point that having more experience relevant to the particular thought experiment would help to decrease the incidence of these types of bias. And, while it could be true that a relevantly experienced participant would have less misleading intuitions, it really depends on how realistic the scenario is and any *a priori* assumptions that are made regarding what is relevant.

Although it may be the case that our intuitions are not in response to the morally-relevant features of unrealistic thought experiments, a more likely occurrence would be that our intuitions are in response to the morally-relevant features *and* other features as well. The main problem that arises from this is that the non-relevant features might have more emotional salience than, and therefore override, the relevant ones. The importance of emotions in moral judgement has been stressed by Haidt (2001), who has clearly demonstrated that our moral judgements are very much affected by our implicitly learned and emotionally mediated social attitudes.<sup>63</sup> Indeed, using neuroimaging technology Greene *et al.* (2004) found that moral judgements made about emotionally salient thought experiments were quite different from those that did not engage emotional processing. The emotional salience problem is exasperated by the way that intuitive cognition processes thought experiments.

---

<sup>63</sup> Although, there is still debate about whether the emotional aspect of moral judgements is a hindrance or not (Sinnot-Armstrong 2006; Woodward & Allman 2007).

Intuitive cognition's adding in of subjectively-experienced features to get a better handle on the bare bones of sterile and unrealistic thought experiments can lead to the intruding features drawing attention away from the morally-relevant ones because they can be emotionally distracting. One example of how our intuitive cognition adds information in this way is by attributing various implicitly learned type characteristics to tokens that we unconsciously associate with that type, despite not being able to consciously perceive any signifiers with which to categorise it (Castelli *et al.* 2004).

Furthermore, any intuition based on contemplation of a single thought experiment (as opposed to comparing multiple thought experiments) could be subject to various emotion-affecting biases that make our intuitions more positive or negative than they would otherwise be. Indeed, if participants had been exposed to a negative mood-affecting stimulus that was too subtle to be consciously noticed, then that stimulus can make any evaluative intuition that they have more negative than it would have otherwise been (Myers 2004, p26-27). For example, graduate students' evaluation of their own work was more negative after being unconsciously exposed to a picture of their supervisor making an angry face (Murphy & Zajonc 1993) and participants rated people as nicer after being unconsciously exposed to pictures of kittens and nastier after being unconsciously exposed to pictures of dead people (Krosnick *et al.* 1992) or the word 'hostile' (Higgins, Rholes & Jones 1977). Since our evaluative intuitions are so easily primed by uncontrollable factors and our intuitions can be swayed by intruding emotionally salient past experiences, even if our intuitions are partially in response to the morally relevant aspect of a thought experiment, they may well still be misleading.

It must also be considered possible that our intuitions are responding to the morally-relevant aspect of the unrealistic thought experiment *and only* that aspect, or at least *mainly* that aspect. However, any such intuition may still be misleading because of how some implicitly learned and morally irrelevant biases affect the way the morally

relevant factor could be unconsciously processed. Although they do not wield neuroscientific evidence, prominent ethicists, such as Bentham (1789), Singer (1974) and Unger (1996),<sup>64</sup> stress that our moral intuitions about particular scenarios are consistently self-serving or based on obsolete social practises that had once conferred some reproductive or survival benefits onto our ancestors (c.f. Woodward & Allman 2007, p3).

Two biases of particular concern for evaluative intuitions about well-being are familiarity and self-serving biases, as they tend to make us evaluate the kinds of lives we know and have experienced more positively than they might deserve. Myers (2004, p39-40) summarises the research on familiarity bias (a type of status quo bias), noting that repeated exposure to numerous stimuli, from Chinese characters to people, made participants' evaluations of them more positive.<sup>65</sup> Myers (2004, p40) astutely considers the familiarity bias to have a good evolutionary explanation in that familiar things are safe (they have allowed us to survive and perhaps reproduce so far) and what we are not familiar with could be dangerous and impair our evolutionary fitness. Applying familiarity bias to thought experiments about well-being, it is clear that any lives that are markedly different from ours, such as Socrates' Oyster Example<sup>66</sup> (in which readers are required to imagine that they are an oyster),<sup>67</sup> will have an intuitive disadvantage when compared to a life more like our own.

The well-documented implicit self-serving bias has also been implicated in many studies that report our various evaluative biases (Greenwald & Banaji, 1995; Steele, 1988; c.f. Jones *et al.* 2002). For example, we are much more inclined to prefer partners (Pelham *et.* 2001), professions, and places to live (Pelham, Mirenberg, & Jones 2002) that have the same first letter of their name as we do, although we are

---

<sup>64</sup> It should be noted that Unger (1996) thinks that he can avoid this problem by using further thought experiments to expose these biases. However, it is not clear exactly how a positive result could ever arise from this approach.

<sup>65</sup> Myers (2004 p39-40, 257) cites two comprehensive reviews of the literature (Bornstein 1989; 1999).

<sup>66</sup> (c.f. Plato in *Philebus*, ii. p353).

<sup>67</sup> This example is explained in more detail in the next section.

not consciously aware of this preference.<sup>68</sup> This preference for things with similar names to ourselves has been dubbed “implicit egotism” (Pelham, Mirenberg, & Jones 2002) because it is thought to stem from an evolutionarily advantageous positivity about the self that leads to, amongst other things, more successful interactions with others and better health (Buss 2000; Myers 1992). This phenomenon explains why Philip Petit is enjoying his dream job as a professor of philosophy and political theory at Princeton<sup>69</sup> and Paul and Patricia Churchland are happily married philosophers at the University of California at San Diego.<sup>70</sup>

The implication of this bias for thought experiments about well-being is that we are likely to have biased intuitions about at least the names of any characters, places or professions that they contain and, unfortunately, substituting numbers would not help because a similar bias exists for the numbers of our birthdays (Kitayama & Karasawa 1997)! Indeed Garner’s (2005) study showed that participants reported increased ability to relate to and liking of characters in scenarios that had similar names to themselves. Therefore, any Olivias or Olivers pondering Socrates’ Oyster Example (c.f. Plato in *Philebus*, ii. p353), might be less resistant to the oyster’s life than Johns and Julies. Consider also Mill’s (1861) assertion that opera contributes to one’s well-being in a more valuable way than pushpin and Aristotle’s claim that philosophising is the pinnacle of human flourishing.<sup>71</sup> Both of these philosophers appear to have been suffering from a heavy bout of self-serving bias when they made these self-affirming claims. While the evidence above makes it clear that many biases can make our intuitions misleading, the problem is considerably compounded because we cannot know for sure whether or how much they are misleading, as discussed below.

---

<sup>68</sup> It should be noted that this name letter bias is most likely over and above any mere familiarity bias (Nuttin 1987; Hoorens & Nuttin 1993; Pelham, Mirenberg & Jones 2002) and is more likely to be related to identity and the self-serving bias (McGuire & McGuire 1981).

<sup>69</sup> Information on Philip Petit retrieved from his Princeton homepage on 5/02/2008 (<http://www.princeton.edu/~ppettit/>).

<sup>70</sup> Information on the Churchlands from a New Yorker article by Larissa MacFarquhar, accessed on 5/02/2008 and available from: (<http://www.scribd.com/doc/3113/Paul-and-Patricia-Churchland-Profile>)

<sup>71</sup> *Nicomachean Ethics*.

In defense of Premise 2, as discussed above, intuitions cannot be introspected or reliably reconstructed, which means that there is no reliable method for knowing when intuitions about unrealistic well-being-related thought experiments have been misled by one or more of the biases mentioned above. Even in realistic scenarios, about which Woodward and Allman (2007) maintain that our relevantly-experienced moral intuitions could have normative significance, we cannot know if our intuitions are misleading because of their being influenced by irrelevant biases. For instance, if I were to compare my life with that of a friend, I might consider my life to be better for me than his life is for him. However, despite this being a very realistic case and one that I have ample relevant experience of, I could never know if my intuitive evaluation was made misleading because of familiarity or self-serving biases. Introspecting my intuition would reveal nothing and attempting to reconstruct it would be unfruitful. For example, attempting a reconstruction, I could suggest to myself that my higher relative evaluation of my own life was based on my friend being recently divorced. Upon considering my friend's life and then considering the concept 'divorce' one after the other, I might find that my moderately negative intuitive response was the same to each. Unfortunately, however, this correlation does not necessarily imply that it was my friend's divorce that was the main cause of my intuition about his life. Furthermore, this correlation might be wholly inappropriate because my friend might consider his divorce as amongst the most well-being-enhancing things that had ever happened to him.

Well-being is a subjective phenomenon such that, when we attempt a reconstruction from the outside, it is difficult not to import various biases and assumptions that corrupt the process. Even when experimenters attempt to isolate various features of scenarios in multiple control groups, the participants' intuitive cognition will most likely add implicitly learned attitudes from their subjective experiences to the scenario, making the isolation of certain aspects very difficult. Furthermore, even reconstruction of our intuitions about *our own* well-being requires that our conscious

mind can reliably remember, interpret and forecast our happiness, which is simply not the case, as discussed at length by several leading cognitive scientists (Gilbert 2006, Haidt 2006; Loewenstein & Schkade 1999). Indeed, the dogged pursuit of wealth by the majority of society, despite its minimal impact on their happiness (Hamilton 2003; Layard 2005) is amongst the best evidence to believe Gilbert's (2006) claim that we are not nearly as good as we think we are at predicting what will make us happy. The evidence discussed earlier and directly above, clearly shows that we cannot introspect or reliably reconstruct our intuitions, which is a considerable obstacle to our intuitions being useful for evaluating unrealistic well-being-related thought experiments.

While premise 3 of the argument (repeated below for convenience) might seem innocuous, Woodward and Allman (2007) appear to disagree. They admit that moral intuitions can be misleading and that there is no reliable method for knowing if they are misleading or not. However, Woodward and Allman (2007) claim that the moral intuitions of relevantly-experienced people about realistic scenarios can be normatively significant to use *with* any other relevant form of evidence because they might be accessing information that would otherwise be unavailable. When Woodward and Allman (2007) provide more detail about how such an unreliable phenomenon as moral intuition can have normative significance, they point to intuitive cognition's sensitivity to the multiplicity of potential short and long-term consequences, something our more-cumbersome deliberative cognition cannot easily imitate.

- P3) If some form of evidence is often misleading and there is no reliable method for knowing when that form of evidence is or is not misleading, then that form of evidence is unreliable and should only be used when no other reliable form of evidence is available.

Woodward and Allman (2007) are right about intuitive cognition having some design features that are better suited to processing very realistic and complex scenarios than standard deliberative cognition. However, intuitive cognition's advantages can be



adequately emulated by the deliberative cognition of a moral philosopher taking full advantage of the resources available to her: by reasoning with others about potential consequences, writing down all of the relevant information and perhaps even running it through a computer programme. A moral philosopher's ability to take the time to research and reason about a well-being-related thought experiment means that their intuition on the scenario is unlikely to be able to offer any privileged information that is relevant to the evaluation. Our intuitive cognition may always be faster than our deliberative efforts, but it is unlikely to be as reliable because of the many implicit biases intuitive cognition inadvertently incorporates. Indeed, just as a good researcher would not rely on the unverifiable testimony of a disreputable source (for fear of misinformation) if there was anything else to go on, a good moral philosopher should not rely on her immediate intuition about a thought experiment if she has the time to produce a well-reasoned judgement about it. Therefore, while it might be wise to at least partially rely on intuitions in current real life situations, the ample time and other resources available to moral philosophers makes their evaluations of unrealistic well-being-related thought experiments an entirely different matter, one where moral intuitions are unreliable and certainly not the only form of evidence available.

The conclusion of the argument follows deductively from the premises above. Naturally, the claim the conclusion makes (that intuitions about unrealistic well-being-related thought experiments are unreliable and should only be used when no other reliable form of evidence is available) is only noteworthy if there are other forms of evidence that are reliable and relevant to well-being-related thought experiments. I have mentioned above and reiterate here that reason can make reliable evaluative judgements in unrealistic well-being-related thought experiments. Although this assertion is explained and supported in more detail below, the role of reason in moral judgements needs to be defended from the possible accusation of self-delusion. Haidt (2001) provides good reason to believe that our moral judgements are often made by our intuitive cognition and then, when challenged to explain the

judgement, our conscious deliberative mind confabulates a reason. It could mistakenly be inferred from this finding that any attempt to use reason to evaluate a thought experiment would really just involve an intuitive judgement and a fabricated justification for it. Fortunately, moral philosophers are experienced in the use of reason and are subjected to various forms of peer review that endeavour to ensure that their use of reason is consistent.

For example, if a hedonist was using Hurka's (1993) Perfectionism to justify her evaluation of a thought experiment in an article, her peers would spot her inconsistent use of reason and she would be very lucky to be invited to 'revise and resubmit'. Even if moral philosophers' reasons for finding particular types of evidence, or moral principles, more reasonable than others might be *initially* formed by a process of reflective equilibrium,<sup>72</sup> which takes intuition into account at an earlier stage, once philosophers begin to present an argument they must apply reason consistently or have their justification ignored. Those reasoned justifications of their evaluations of thought experiments are then able to be analysed for consistency, and compared with other competing justifications. Since the end result of reason is so easily subjectible to these checks and balances, we can be more confident that the justifications offered by most moral philosophers are not mere moral intuitions cloaked in confabulation, but predominantly the product of deliberative cognition. With the weight of the argument supported by the defences of its premises above and its purpose now also vindicated, its implications for evaluating theories of well-being are addressed below.

---

<sup>72</sup> As discussed in Chapter 1, reflective equilibrium is the endpoint of a process in which intuitions about cases and reasoned principles adjust in light of each other to form a judgement (Rawls 1971; c.f. Woodward & Allman 2007, p3).

## The New Method for Assessing Theories of Well-Being

Based on the above argument for intuitive evidence from unrealistic thought experiments being considered unreliable, the Old Method of Assessing Theories of Well-Being is clearly in need of amendment. The New Method of Assessing Theories of Well-Being, which follows, takes a better understanding of intuitive cognition into account by restricting the role that intuitive evidence can play in evaluating theories of well-being.

Theories of well-being should:

- i) Produce intuitively pleasing results in *realistic* cases
- ii) Have a compelling rationale, which can provide judgments in all relevant moral cases, while:
  - a) Being internally consistent
  - b) Producing results in *realistic* cases that fit with pervasive moral judgements (based on reflective equilibrium between our moral intuitions and other judgements)
  - c) Providing good reason to believe the normative significance of the theory
  - d) Being meta-ethically consistent
  - e) Being consistent with science
- iii) Be functional

And the extent to which a theory of well-being achieves these criteria indicates how good that theory is.

Although the New Method of Assessing Theories of Well-Being only swaps two occurrences of a single word of the previous definition, the changes this new definition demands would revolutionise ethics as we know it. This is because, while unrealistic thought experiments have been ubiquitous in moral philosophy, their humble realistic counterparts are much harder to find. Indeed, consider the most influential thought experiments posed against Hedonism in the past, for example; Socrates' Oyster thought experiment (c.f. Plato in *Philebus*, ii. p353) is utterly unrealistic.

In the Oyster thought experiment, Plato's Socrates asks Protarchus to imagine a life without much pleasure but full of the higher cognitive processes, such as knowledge, forethought and consciousness and compare it to a life that is the opposite. Socrates describes this opposite life as having perfect pleasure but the mental life of an oyster.

Any negative intuition about the pleasant life of an oyster is clearly not based on an understanding of what it would actually be like to be an oyster, since it is impossible for us to imagine what it would be like to be unconscious. Instead, our intuition would be based on imagining *ourselves* being stuck in one particular state (albeit a nice one, akin to being slightly drunk lying in a warm bath) for eternity, which would soon become torturous (not pleasurable) for beings with our conscious cognitive capacities.

And what, then, of the thought experiments that are currently thought to refute Hedonism: the Deceived Businessman and the Experience Machine? Both of these thought experiments are clearly unrealistic. Do they hold any normative significance for Hedonism about Well-Being? Before delving deeply into these questions, a brief guide will be provided as to how the New Method of Assessing Theories of Well-Being should be applied and how arguments against established theories of well-being should be dealt with.

To evaluate a theory of well-being using the New Method of Assessing Theories of Well-Being, the theory must be tested against the criteria mentioned above. If a theory of well-being can provide credible evidence to believe that it produces intuitively pleasing results in realistic cases, is rationally compelling, and is functional, then, according to the New Method of Assessing Theories of Well-Being, it is plausible. Furthermore, fulfilling the criteria past the point of plausibility (towards, say, being very convincing) is what determines if a theory is better than other plausible theories.

With how to apply the New Method of Assessing Theories of Well-Being in mind, it is pertinent to outline how to deal with objections to a theory once its plausibility has been established, since arguments against it will doubtless arise at some point. A

framework for dealing with objections to theories of well-being is discussed below and outlined as follows:

When confronted with an objection to a theory of well-being:

- a. Assess how rationally compelling the argument (for the objection) is by checking for:
  - i. Internal consistency
  - ii. Relevance
  - iii. Scientific consistency
  - iv. Evidence (good reasons to believe it):
    1. Reason-based evidence
    2. Intuitive evidence
- b. Reject the objection if any of i, ii, iii, or iv can be sufficiently undermined
- c. If the objection is rejected, then the theory of well-being maintains its plausible status. However, if the objection cannot be rejected, then the theory of well-being loses its plausible status. The more un-rejected objections against a theory, the worse that theory is.

Any objection against a theory of well-being should be assessed on how rationally compelling the argument it presents is. First of all, the argument should be internally consistent, that is, it must not contradict itself in anyway. The proposed argument against the established theory of well-being should also be relevant; it should be made clear exactly how the argument objects to the theory. Any objection to a plausible theory of well-being also needs to be at least consistent with current science and not based on un-falsifiable premises. If the objection is based on an implausible belief about science, then that objection will have no normative significance and not be able to make the theory of well-being in question implausible. Furthermore, if the objection is based on a premise that is un-falsifiable by science, such as 'the word of God is truth', then the objection will also fail to make the theory of well-being in question implausible. This is because un-falsifiable evidence has no identifiable 'weight', or normative significance, and cannot, therefore, play the role of disconfirming evidence. Most arguments against theories of well-being tend to be able to comply with these three criteria fairly easily. However, fulfilling the final criterion will most often be a matter of substantial debate.

When providing good reasons for believing an argument that objects to a plausible theory of well-being, the evidence that can be provided will either be reason-based or

intuitive. Reason-based evidence is often still tinged with intuitions about fairly fundamental aspects of the relevant theories involved. However, such evidence remains predominantly reason-based and generally allows for rational debate along logical lines. Because of this, reason-based evidence is preferential to intuitive evidence, which is subject to the various biases discussed above and does not easily lend itself to rational debate. Furthermore, unrealistic thought experiments are even more likely to produce intuitions that have no normative significance and should only be relied on in inescapable instances, such as when considering potential states of being that have not yet been experienced. However, even in such instances of attempting to assess relatively unknown states, our intuitions should play as small an evaluative role as possible. Indeed, based on the findings above, intuitions about unrealistic thought experiments alone should never play a more important role than that of a warning sign in assessing theories of well-being. It will be argued in the next chapter that this is the role that our intuitive responses to the Deceived Businessman and the Experience Machine thought experiments should play in assessing Hedonism about Well-Being.

And, how should warning signs be treated? Warning signs call for further investigation. Acknowledged but still unresolved warning signs should tend to cast some (but not conclusive) doubt on the plausibility of the theory of well-being that gave rise to them. To begin to erase this doubt, proponents of a theory of well-being need just to provide a plausible but morally-irrelevant causal explanation for the intuition (Sobel 2002, p244). By casting doubt on the relevance of the intuition to the theory, the intuition is quickly heading towards losing its warning sign status. Reason-based refutations of the normative significance of such intuitions are the most effective way to rule them out of consideration when assessing the theory of well-being in question. However, the intuitions aroused by certain counter-thought experiments (even unrealistic ones) can also undermine intuitions that are claimed as evidence against a theory of well-being. It may seem hypocritical to allow the use of

intuitive evidence from unrealistic counter-thought experiments after arguing against the normative significance of intuitions about unrealistic thought experiments. However, if an unrealistic counter-thought experiment manages to produce the opposite intuition to that of the original thought experiment, then not only is that intuition's credibility brought into doubt, but intuitions about all unrealistic thought experiments become a much more questionable source of evidence.

To prevent this cancelling of their warning sign, opponents of the theory under scrutiny should attempt to provide morally-relevant reasons that justify the existence of the intuitive response that they are attempting to use as evidence. It would be next to impossible to prove that the reasons they provide *cause* the intuition, but the reasons themselves can be analysed and discussed to assess how rationally compelling they are. However, if no reason can be provided that could support the verdict of the intuition, then it can be assumed that the intuition is not in response to anything relevant.

Naturally, such a post-intuitive justification is often given, and it is these justifications that proponents of besieged theories of well-being, such as Hedonism about Well-Being, need to respond to. The best response that a proponent of Hedonism about Well-Being could hope for is a dual response that can refute his opponents' reasoned justifications *and* their intuitive responses. Theoretically, refuting the reasoned justification should be sufficient, especially when a theory is only aiming for plausibility. However, a dual approach, which also encourages opponents to question their initial intuition, will be more persuasive. And, being more persuasive is essential when attempting to revive a philosophically dead theory by refuting thought experiments, such as the Deceived Businessman and the Experience Machine, which are widely thought to be conclusive objections. Indeed, such a complete refutation should be able to convince even those who still believe that moral intuitions about unrealistic thought experiments have some normative significance. With these

guidelines in mind, the contribution that the Deceived Businessman and the Experience Machine thought experiments could provide to the evaluation of Hedonism about Well-Being is discussed in the next chapter and the other implications of the New Method of Assessing Theories of Well-Being are summarised below.

The New Method of Assessing Theories of Well-Being is a fairly radical departure from the Old Method of Assessing Theories of Well-Being. The obvious difference between the views is the restriction of intuitive evidence from unrealistic thought experiments. In the Old Method of Assessing Theories of Well-Being, intuitions from all sources were welcomed as evidence, but this was because moral philosophers were so confident that their intuitions (which they believed always reflected their wealth of life and ethical experience) must be correct (e.g. Lillie 1955, p17-20). This view overstates the benefits and understates the weaknesses of intuitive evidence from unrealistic thought experiments, showing an understandable lack of knowledge about the workings of intuitive cognition. In the New Method of Assessing Theories of Well-Being, moral intuitions about unrealistic thought experiments are demoted to the role of warning signs. If widely applied, the New Method of Assessing Theories of Well-Being will change many of the current views on theories of well-being, as shown by its application to Hedonism about Well-Being in the next chapter. Furthermore, any changes to current views on well-being will have flow on effects in moral philosophy generally because of well-beings' fundamental role in normative ethics (Kagan 1998).<sup>73</sup> Finally, it should also be noted that the New Method of Assessing Theories of Well-Being could easily be adapted so that it could be used to assess normative moral theories, thereby causing more implicative ripples in the wider area of moral philosophy.

---

<sup>73</sup> The role of well-being is certainly fundamental in standard consequentialist positions at any rate (Sobel 1994).



## Conclusion

This chapter has provided a detailed account of moral intuitive cognition and discussed and expanded on the role that Woodward and Allman (2007) recommend for intuitions in ethical theorising. First, Woodward and Allman's (2007) description of moral intuitive cognition as a fast, unconscious probabilistic process that could compute large amounts of data in parallel and produce an intuition (a simple visceral sensation) was explained in detail and corroborated by other leaders in the field of cognitive science. Then, this greater understanding of moral intuitions was used to assess and mostly agree with Woodward and Allman's (2007) recommendations for appropriate use of moral intuitions in ethics. One conclusion of theirs in particular was formulated into an argument that made it specifically relevant for evaluating theories of well-being. The argument concluded that intuitions about unrealistic well-being-related thought experiments are unreliable and should only be used when no other reliable form of evidence is available.

Based on this argument for intuitive responses to unrealistic thought experiments playing a considerably restricted role in evaluating theories of well-being, the New Method of Assessing Theories of Well-Being was proposed. This new method was then discussed, including an explanation of how to deal with objections to a theory of well-being that had been established as plausible. Particular attention was paid to dealing with intuitive evidence against theories of well-being because that issue takes centre stage in Chapter 3, as the normative significance of the Deceived Businessman and the Experience Machine thought experiments for Hedonism about Well-Being is challenged.

---

## Chapter 3:

# Refuting the Argument from False Pleasures

---

### Introduction

With the New Method of Assessing Theories of Well-Being sufficiently explained and supported in Chapter 2, the current chapter uses it to assess Hedonism about Well-Being. A complete assessment is not possible within the space of this essay, so specific attention is focussed on the most popular argument against Hedonism about Well-Being: the Argument from False Pleasures. Indeed, after Hedonism about Well-Being is briefly established as plausible, the rest of this chapter is dedicated to assessing the normative significance that the Deceived Businessman and the Experience Machine thought experiments supposedly hold for it. Each is analysed in turn, revealing that, even when interpreted as charitably as possible, neither proves to have the normative significance that the proponents of the Argument from False Pleasures suppose they do.

The power of the Deceived Businessman thought experiment is found to rest purely on a strong but misguided intuition that should not have been consulted in the first place. It is also argued that any attempt to rework the Deceived Businessman thought experiment would not avoid this criticism of irrelevance. The strength of the Experience Machine thought experiment is also found to rest on misguided

intuitions, although less obviously so. The Experience Machine is then reinterpreted into what must be its strongest reincarnation to date: the Pleasure Machine. Even when given this new look, however, the Experience Machine fails to provide any good reasons to think that Hedonism about Well-Being is implausible. Based on the irrelevance of the Deceived Businessman and Experience Machine thought experiments, it is concluded that the Argument from False Pleasures lacks the evidence it requires to refute Hedonism about Well-Being.

Importantly, the process used to evaluate the evidence that the Deceived Businessman and the Experience Machine thought experiments are supposed to provide is so thorough that it is compatible with the New Method of Assessing Theories of Well-Being *and* approaches that would place much more importance on intuitive evidence, such as a reflective equilibrium-type method. This fact makes the conclusions of this chapter significant even if the arguments of Chapter 2 are not considered persuasive and the New Method of Assessing Theories of Well-Being is not accepted.

## **Evaluating Hedonism about Well-Being**

A thorough application of the New Method of Assessing Theories of Well-Being to Hedonism about Well-Being, and all of the objections against it, would be a considerable task and one that would be too lengthy to accommodate in this essay. Due to this space limitation, only a very brief positive case will be made for the plausibility of Hedonism about Well-Being and only one argument against its plausibility will be dealt with. However, the argument against Hedonism about Well-Being that will be refuted in this chapter, the Argument from False Pleasures, is of paramount importance because it currently convinces most moral philosophers that

Hedonism about Well-Being is an implausible theory.<sup>74</sup> And, although not all of the conceivable evidence that could be used to support the Argument from False Pleasures can be discussed in this essay, the main evidence: the Deceived Businessman and Experience Machine thought experiments, will be. Because of the important role these two thought experiments play in Hedonism about Well-Being being considered implausible, this chapter will argue against their normative significance. And, if it can be shown that the Deceived Businessman and Experience Machine thought experiments do not provide good evidence to support the Argument from False Pleasures, then it will be assumed that the Argument from False Pleasures fails to provide a good reason to believe that Hedonism about Well-Being is not a plausible theory.

Many authors, including those who explicitly argue against Hedonism about Well-Being, have stated that it appears plausible upon first inspection (as, for example, discussed between Plato's Socrates and Protagoras).<sup>75</sup> Hedonism about Well-Being will now be evaluated using the New Method of Assessing Theories of Well-Being (repeated below for convenience) to show that it is at least *prima facie* plausible, thereby providing reason for Hedonism about Well-Being's opponents to provide the arguments against it that they already have.

Theories of well-being should:

- i) Produce intuitively pleasing results in *realistic* cases
- ii) Have a compelling rationale, which can provide judgments in all relevant moral cases, while:
  - a) Being internally consistent
  - b) Producing results in *realistic* cases that fit with pervasive moral judgements (based on reflective equilibrium between our moral intuitions and other judgements)
  - c) Providing good reason to believe the normative significance of the theory
  - d) Being meta-ethically consistent
  - e) Being consistent with science
- iii) Be functional

And the extent to which a theory of well-being achieves these criteria indicates how good that theory is.

---

<sup>74</sup> For example, Bagani & Fosl (2007, p74-76), Brink (1989, p223-224), Darwall (1997, p162, 178), Griffin (1986, p9-10), Kagan (1998, p34-36), Kymlicka (1990, p13-14), and Rachels (2005, p103-104).

<sup>75</sup> See Nussbaum (2001, chapter 8) for discussion.

Hedonism about Well-Being plausibly provides intuitive results when evaluating well-being in realistic cases by reducing all instrumentally well-being affecting experiences to the pleasurable and painful parts of mental states. For example, when we hear of the lives of others who suffer much pain, we often pity them for their poor well-being. Furthermore, if we later hear that their life is no longer full of pain, but full of joy instead, then we are happy for them about their improved well-being. Consider also the case of a deserving student feeling very pleased after hearing that she was awarded a distinction for her thesis, but then becoming very upset when she discovers that there was an administrative mistake and that her thesis had not even been marked yet. The joy that the student first felt and the negative mental states that she was later experiencing are clearly good indicators of her well-being during those times. Examples of Hedonism about Well-Being producing unintuitive results are also fairly easy to come by. However, these examples tend to be more unrealistic, making them much less relevant in using the New Method of Assessing Theories of Well-Being to evaluate Hedonism about Well-Being.

Hedonism about Well-Being also plausibly satisfies the New Method of Assessing Theories of Well-Being's criteria for having a compelling rationale. As captured by the definition discussed in the introduction to this essay, Hedonism about Well-Being, states that all pleasure, and only pleasure, intrinsically contributes positively to well-being and the opposite for pain. The rationale for this theory is that our well-being is wholly and only dependant on the momentary and enduring mental states that we are aware of, that these mental states are always either neutral, painful, or pleasurable to some degree, and that the pleasurable mental states make our lives intrinsically better for us while the painful ones make them intrinsically worse.

Hedonism about Well-Being's rationale is internally consistent because it does not contradict itself. Regardless of whether the mental states we are aware of are split into potentially simultaneous and overlapping specific feelings, or discrete moments of

total feeling, these mental states are *on the whole* exclusively positive, negative or neutral, thereby not allowing any contradiction. This rationale is also clearly relevant to Hedonism about Well-Being, as the New Method of Assessing Theories of Well-Being requires, and provides a plausible explanation for what the theory means and how it applies to well-being. Hedonism about Well-Being's rationale states that pleasurable and painful mental states (which we are aware of) intrinsically make our lives better or worse for us (respectively), thereby providing a plausible explanation of how it can be used to theorise about well-being.

Implicit in Hedonism about Well-Being's rationale is the more detailed explanation supporting pleasure and pain being the *only* things that intrinsically contribute to well-being. This explanation argues that many things instrumentally increase or decrease our well-being, but these effects are best measured by the pleasurable and painful feelings we have (that we are aware of), which intrinsically contribute to our well-being. Hedonism about Well-Being's rationale is also determinately adequate because for any real or hypothetical case, for which the relevant factual information is available, Hedonism about Well-Being can provide a clear answer. For example, Hedonism about Well-Being's rationale claims that anything that occurs, but does not cause an individual to be aware of a pleasurable or painful feeling, does not affect that individual's well-being. Hedonism about Well-Being's rationale also claims that anything that occurs, and *does* cause an individual to subjectively experience a pleasurable or painful feeling, *does* affect that individual's well-being. Furthermore, Hedonism about Well-Being's rationale claims that an individual's well-being can be evaluated solely by reference to that individual's experiences of pleasure and pain. Using this explanatory system, and given the relevant factual data about pleasure and pain, Hedonism about Well-Being can provide an evaluation of the well-being of any individual and whether certain changes to their life would improve or worsen their well-being.

Furthermore, despite numerous historical allegations that there is no such unifying pleasurable component to all of our experiences of positive feelings,<sup>76</sup> current science provides no evidence to support these claims. While we obviously experience different pleasurable feelings, there is no neurobiological evidence for there being different types of pleasure in the brain. Furthermore, neuroscientists have found that several types of supposedly different ‘pleasures’ are mediated by the same neural circuitry.<sup>77</sup> Lieberman and Eisenberger (2006) also argue that the experience of social and physical pains is at least somewhat mediated by the same basic neural circuitry. Finally, when the fundamental components of Hedonism about Well-Being’s rationale are considered by themselves, and not via unrealistic thought experiments, they are at least plausibly intuitive. For example, the fundamental hedonistic claim that the subjective experience of pleasure is intrinsically good for the individual experiencing it must be intuitively plausible. If it were not, many authors would not bother to posit arguments against it (e.g. Kagan 1998, p30; Rachels 2005, p103). Although drastically brief, this discussion of Hedonism about Well-Being’s rationale should be enough to show that Hedonism about Well-Being is at least *prima facie* theoretically plausible.

According to the New Method of Assessing Theories of Well-Being, a good all-round theory of well-being is also functional. Therefore, what follows is a brief account of how Hedonism about Well-Being could be useful for policy-making. To be useful for policy-making, a theory of well-being needs to be able to be at least approximately measured. Bentham (1789) and Mill (1861) struggled with the obvious practical difficulties of measuring of pleasure, but technology has dramatically improved since then. Recently, some cognitive scientists are measuring what they think might be

---

<sup>76</sup> For example: Anscombe (1957), Brentano (1969), Rawls (1971), and Sidgwick (1907).

<sup>77</sup> See: Berridge (2003; 2007), Depue and Morrone-Strupinsky (2005), Kringelbach (2005), and Pecina, Smith, and Berridge (2006).

pleasure with neuroimaging<sup>78</sup> and attempts are being made to find correlations between what we self-report about pleasure and pain and what is happening in our brains.<sup>79</sup> These results have led some economists and philosophers to think that pleasure or happiness might be able to be measured accurately enough to be useful for public policy.<sup>80</sup> Indeed, some economists have already used this kind of approach for real policy decisions (e.g. van Praag & Baarsma 2005). It is far from clear that pleasure or happiness will ever be able to be measured very accurately. However, it is more than plausible that it could be measured accurately *enough*, making Hedonism about Well-Being a potentially useful theory for both personal and public policy-making.<sup>81</sup>

The brief discussion above shows that Hedonism about Well-Being plausibly produces intuitive results in realistic cases, has a compelling rationale, and is potentially functional. These results mean that, according to the New Method of Assessing Theories of Well-Being, Hedonism about Well-Being is at least a *prima facie* plausible theory. With Hedonism about Well-Being's plausibility established, the remainder of this chapter is dedicated to the refutation of the strongest argument against it: Argument from False Pleasures, which is repeated here for convenience:

- P1) Hedonism states that all pleasure, and only pleasure, intrinsically contributes positively to well-being and that all pain, and only pain, intrinsically contributes negatively to well-being.
  - P2) Pleasure based on truth, or something like it, contributes more positively to well-being than pleasure based on falsity.
  - P3) Therefore, something other than pleasure (truth of some sort) must contribute positively to well-being.
- 
- c) Therefore, Hedonism is false.

<sup>78</sup> See: Berridge (2003; 2007), Davidson (1992; 2000), Davidson, Jackson, and Kalin (2000), Depue and Morrone-Strupinsky (2005), Pecina, Smith, and Berridge (2006), Robinson and Berridge (2000; 2001), Tindell *et al.* (2006), and Wheeler and Carelli (2006).

<sup>79</sup> See: Coghill, McHaffie, and Yen (2003), Derbyshire *et al.* (2004), Koyama *et al.* (2005), and Urry *et al.* (2004).

<sup>80</sup> For example: Kahneman (2000), Kahneman and Sugden (2005), Kahneman, Wakker, and Sarin (1997), Layard (2005), Veenhoven (2004).

<sup>81</sup> See: Dolan and White (2007) for an interesting review of the three types of well-being discussed in this essay and how they might be used in policy-making, in which mental-state accounts (such as Hedonism) play a central role.



And, following the New Method of Assessing Theories of Well-Being discussed in Chapter 2, arguments against a theory of well-being that has been established as plausible can be assessed as follows:

When confronted with an objection to a theory of well-being:

- a. Assess how rationally compelling the argument (for the objection) is by checking for:
  - i. Internal consistency
  - ii. Relevance
  - iii. Scientific consistency
  - iv. Evidence (good reasons to believe it):
    1. Reason-based evidence
    2. Intuitive evidence
- b. Reject the objection if any of i, ii, iii, or iv can be sufficiently undermined
- c. If the objection is rejected, then the theory of well-being maintains its plausible status. However, if the objection cannot be rejected, then the theory of well-being *loses* its plausible status. The more un-rejected objections against a theory, the worse that theory is.

The Argument from False Pleasures will not be refuted on the grounds of internal consistency, relevance, or scientific consistency because it is a valid argument that would undermine the plausibility of Hedonism about Well-Being if its premises were true. As mentioned in Chapter 1, Premise 2 is the contentious premise and, therefore, proponents of the Argument from False Pleasures need to provide evidence to support it. The reason-based evidence for Premise 2 is limited to the statement itself; ‘Pleasure based on truth, or something like it, contributes more positively to well-being than pleasure based on falsity’. However, the intuitive evidence for this premise is generally considered to be very strong (e.g. Griffin 1986; Kagan 1998; Nozick 1974). As discussed in Chapter 1, the intuitive evidence for Premise 2 of the Argument from False Pleasures is usually elicited by consideration of either or both of the Deceived Businessman and the Experience Machine thought experiments. For this reason, the remaining sections of this chapter will constitute a detailed refutation of the ‘intuitive evidence’ generally aroused by consideration of these two thought experiments.

## Deflating the Deceived Businessman

The Deceived Businessman thought experiment is simpler than the Experience Machine thought experiment and will, therefore, be tackled first. The Deceived Businessman thought experiment will be briefly explained again, including a discussion of what intuitive and reason-based evidence it supposedly provides for the Argument from False Pleasures. The intuitive evidence elicited by the Deceived Businessman thought experiment is found to be *prima facie* strong. However, no compelling reason-based evidence is found to support those intuitions. Drawing on the findings of Chapter 2, reasons are then provided why intuitions about unrealistic thought experiments, and specifically the intuitions aroused by the Deceived Businessman thought experiment, are unreliable. Intuitive evidence, in the form of counter-thought experiments, is then offered to weaken any stubborn belief that the moral intuitions elicited by the Deceived Businessman thought experiment are in fact based on the intrinsic badness of deception or untruthfulness. Part of Hedonism about Well-Being's rationale is then used as a reason-based argument against the intuitions brought out by the Deceived Businessman thought experiment being relevant to Hedonism about Well-Being. It will be argued that proponents of the Deceived Businessman thought experiment lack a good response to this challenge, having to rely on the already-shaky intuitions it arouses.

One final reason-based argument against the relevance of the intuitions elicited by the Deceived Businessman thought experiment is then posited, which provides a compelling alternate causal explanation for the intuitions. With the relevance of the Deceived Businessman thought experiment in assessing the plausibility of Hedonism about Well-Being looking entirely negligible, reworking the thought experiment to help it overcome some of its problems is then considered. It is discovered, however, that the Deceived Businessman thought experiment cannot be reworked to avoid the criticisms against it without losing its initial intuitive force. Following the guidelines for dealing with objections to the plausibility of theories of well-being discussed in

Chapter 2, the arguments discussed in this section amount to a clear refutation of the relevance of the Deceived Businessman thought experiment in assessing the plausibility of Hedonism about Well-Being.

As discussed in Chapter 1, the Deceived Businessman thought experiment is described most clearly by Shelly Kagan (1998, p34-36). Kagan (1998, p34) begins his discussion with the common saying: “what you don’t know can’t hurt you”, before going on to give reasons to doubt its wisdom. He asks us to imagine a successful businessman who died thinking that he had achieved everything he wanted: a loving wife, adoring children, and the respect of the community (Kagan 1998, p34). However, Kagan (1998, p35) then reveals that the businessman was completely wrong about his assessment of how things had gone: his wife was cheating, his children and the community at large were just using him for their own ends, and his business partner had been stealing from the business, which will soon be bankrupt.

Kagan (1998, p35) asserts that the businessman’s life did not go as well as it could have and points out that if he had really been loved and respected like he thought he was, then his life would have gone better. This point requires the reader to consider the life of the Deceived Businessman, then the Non-Deceived Businessman and then to compare them. Indeed, the intuition that consideration of the Deceived Businessman’s life elicits is not a particularly pleasant one. Kagan (1998) suggests that the reason for this unpleasant intuition is that the Deceived Businessman was not *really* living the life that he had wanted to: his preferences were not being satisfied (Kagan 1998, p36). Kagan (1998) concludes that, since we value a life where our preferences are really satisfied over one in which we only think they are, the truth or reality of our experiences affects the value of them. Clearly, this conclusion, if correct, supports the controversial premise from the Argument from False Pleasures because it claims that the subjective experience of pleasure is less valuable if based on falsity or something like it.

To be painstakingly clear about the evidence that the Deceived Businessman thought experiment is supposed to provide for the Argument from False Pleasures, the intuitive and reason-based evidence will be discussed in more detail here. It is clear that the intuition raised by consideration of the Deceived Businessman's life is a negative one and that the one elicited by consideration of the Non-Deceived Businessman's life is not. Since Kagan (1998, p13) believes that generally agreed upon, and fairly un-reflected upon, moral intuitions that arise from consideration of abstract thought experiments have normative significance, he also believes that our intuitive responses to the Deceived and Non-Deceived Businessmen's lives are compelling evidence. Kagan (1998, p36) also suggests that the reason for our intuitions in this case is that the Deceived Businessman had a bad life and certainly a worse life than if he had not been deceived.

Kagan (1998, p36) goes on to provide the following reason-based evidence for the common intuition that the Deceived Businessman's life would have been better for him if he were not deceived: individual's lives go better for them if they *really* get what they want, as opposed to just *experiencing* getting what they really want in their lives. This reason-based evidence appears to provide a *prima facie* good explanation for the intuitions raised by the Deceived Businessman thought experiment. However, recall that Kagan (1998, p34) began this argument by stating that he would provide reasons to believe that the "common saying that what you don't know can't hurt you", is false. If we interpret this common saying as part of Hedonism about Well-Being's rationale, then it can be stated as: states of the world only affect an individual's well-being if it leads to them having a positive or negative subjective experience. Now, compare this with Kagan's (1998, p36) reason why the Deceived Businessman's life could have gone better. Both propositions are *prima facie* plausible and yet they contradict each other. It should also be noted that Kagan's (1998, p36) proposition is a fundamental part of the rationale of desire-fulfilment accounts of well-being: that

getting what we desire intrinsically contributes positively to our well-being. Despite both of these propositions being plausible, because they contradict each other and each of them is vital for the plausibility of their respective theory of well-being, then at least one of the propositions must be false and, therefore, at least one of the theories must be wrong. To resolve this apparent stalemate, the evidence for and against each proposition must be considered.

The evidence used to defend the desire-fulfilment theorists' proposition that individuals' lives go better for them if they *really* get what they want, as opposed to just *experiencing* getting what they really want, and also to attack the hedonistic experiential proposition, is ultimately circular and rests solely on intuitions. The intuitions elicited by the Deceived Businessman thought experiment lead to the general preference for the Non-Deceived Businessman's life over the Deceived Businessman's life. This preference *itself* is not compelling evidence to believe that having our preferences fulfilled improves our well-being because it begs the question *why* we might desire things in the first place. Indeed, whether or not that preference is reliable is the question that we are trying to answer. Therefore, to have the force to contradict the hedonistic proposition, the proposed reason for this preference: that getting what we desire intrinsically contributes positively to our well-being, still requires evidence. And, the only evidence that the Deceived Businessman thought experiment provides for the desire-fulfilment theorists' proposition is the intuition it arouses.<sup>82</sup>

Similarly, the evidence that supports the hedonistic proposition that only states of the world, which an individual subjectively experiences at some point can affect their well-being, is also intuitive. Furthermore, the attack the hedonists can make against the desire-fulfilment theorists: exactly how does something that we do not

---

<sup>82</sup> As will be discussed, the intuition is matched with an inference to the best explanation (getting what we desire (regardless of if we are aware of it) intrinsically contributes positively to our well-being) which is refuted below.

subjectively experience affect our well-being?, can be answered by their proposition: that getting what we desire (regardless of if we are aware of it) intrinsically contributes positively to our well-being. Therefore, how compelling the answer given by the desire-fulfilment theorists is, in this case, rests entirely on how compelling the evidence for their claim (the intuitions aroused by the Deceived Businessman thought experiment) is. If these intuitions are found to be compelling evidence to believe the desire-fulfilment theorists' proposition, and disbelieve the hedonistic one, then the Argument from False Pleasures (with evidence from the Deceived Businessman thought experiment) will undermine Hedonism about Well-Being's plausibility. For this reason, the rest of this section is dedicated to providing reason-based and intuitive evidence against the normative significance of the intuitions elicited by the Deceived Businessman thought experiment for evaluating Hedonism about Well-Being.

As discussed at length in Chapter 2, our intuitions about unrealistic well-being-related thought experiments are often misleading and, because there is no way to know whether or not they are misleading, unreliable. Indeed, because we cannot introspect or reliably reconstruct our moral intuitions, there is no way to know what our intuitions about the Deceived Businessman's life are in response to and if they are responding in a reasonable way. Therefore, it is not clear what exactly, if anything, the intuitions aroused by the Deceived Businessman thought experiment are providing evidence for. Furthermore, because, as discussed in Chapter 2, the Deceived Businessman thought experiment is drastically unrealistic, the intuitions it elicits are even less likely to be reliable. In fact, one aspect of the thought experiment is so unrealistic that, because of the way our intuitive cognition works, it *alone* is enough to prevent the intuitions the Deceived Businessman thought experiment arouses from being relevant for assessing Hedonism about Well-Being. That aspect is the ruling out of the Deceived Businessman being caused to suffer any pain from the deceptions that riddled his life – *after we have heard about them*.

As mentioned in Chapter 2, expecting our intuitive cognition to receive the information about all of the deceit involved in the Deceived Businessman's life and then to compute that he never finds out or suffers from it in any way is asking far too much from a probabilistic process that cannot always follow verbal rules. The resulting intuition in this case will be based on our past experiences of deceit, including how it generally made life worse for the person being deceived and everyone involved regardless of when, if, and how they found out about it. Even though Kagan (1998, p34) clearly stipulated that the Deceived Businessman was considering his life on his death bed (and so would have no chance of finding out about the deception) that is not the way that our intuitive cognition processes the scenario. Since most of us have no experience of being on our death beds, then our intuitive cognition will not be able to take that 'no possible future' stipulation about the Deceived Businessman's current state into account. Since the Deceived Businessman thought experiment asks us to accept such aspects of the scenario in a way that our cognitive intuition is not capable of doing, the intuitions that arise in response to it are very likely to be misleading.<sup>83</sup>

Furthermore, because we cannot introspect or reliably reconstruct our intuitions, there is no way to rule out the fact that the intuitions elicited by the Deceived Businessman thought experiment are probably unreliable. Therefore, the intuitions aroused by consideration of the Deceived Businessman thought experiment are not reliable enough to play a normatively significant role in assessing Hedonism about Well-Being. However, as per the guide for applying the New Method of Assessing Theories of Well-Being, discussed in Chapter 2, the intuitions can be viewed as a warning sign for further investigation. To allay this warning sign, intuitive and more reason-based evidence against the relevance of the intuitions elicited by the Deceived Businessman thought experiment follow.

---

<sup>83</sup> The unrealistic nature of the Deceived Businessman thought experiment is discussed in more detail below.

Although the normative significance of intuitive evidence from unrealistic thought experiments has been dramatically down-graded in the New Method of Assessing Theories of Well-Being, some such evidence will be provided here for two reasons. Firstly, by providing thought experiments that elicit intuitions which appear to contradict previously held intuitions, the case for the reliability of intuitions has further doubt cast upon it. Importantly, this doubt of the reliability of intuitions provides further motivation for accepting the New Method of Assessing Theories of Well-Being. Secondly, the thought experiments below are designed to cast doubt on the normative significance of the specific intuitions aroused by the Deceived Businessman thought experiment to show that those particular intuitions are unreliable. This specific doubt thereby further undermines the so-called evidence that the Deceived Businessman thought experiment is thought to provide for the Argument from False Pleasures. For example, if a thought experiment can arouse equally good or bad intuitive evaluations of experiences based on falsity and truth, then we can ask why the falsity of the pleasure does not matter in this case but it does in the Deceived Businessman thought experiment. Indeed, this is the question that the reason-based evidence below provides a compelling answer for. Before answering that question, however, three thought experiments will be discussed: first one designed to show that false pleasures can contribute more to well-being than the truth, then one designed to show that false pleasures and pleasures based on truth can be equally valuable, and then one designed to show that false pains are equally as bad as real pains.

The following thought experiment is based on the character Amelie, from the 2001 movie of the same name,<sup>84</sup> and is designed to show that, intuitively, false pleasures contribute more to well-being than the truth does in at least some instances. Imagine that Amelie decides to become a ‘do-gooder’ and dedicates her life to promoting the well-being of others. Amelie’s landlady is a widower whose only source of happiness

---

<sup>84</sup> See <<http://movies.yahoo.com/movie/1807565371/info>> (accessed on 4/03/2008) for more information on the movie *Amelie*.



appears to be from reminiscing about the amazing romance that she had shared with her husband by rereading the many exquisite love letters that he had written to her before he died 10 years ago in a tragic accident while on a business trip abroad. Amelie was going to visit the far off place where her landlady's husband died and so her landlady asked if Amelie could find out anything about her husband's death while she was there. Naturally, Amelie, the do-gooder, agreed. However, Amelie's extensive investigations revealed good reason to believe that her landlady's husband was having an affair while on the business trip. Amelie considered telling her landlady the truth but decided against it because she knew that it would take away the only thing of value to the old woman: her cherished memories about her loving husband, thereby devastating her life. Instead Amelie, a master forger who had had exposure to the original love letters, wrote a wonderfully tender final letter of unyielding love and devotion to the landlady from her husband, dated just before the accident. Amelie knew the letter would bring great joy to her landlady and knew that she would never find out the truth. She was right.

When imaging this thought experiment, it is hard not to cringe at the thought of Amelie telling her landlady the sordid truth and watching her spiral into depression – our intuition about telling the landlady the truth is certainly not a pleasant one. Furthermore, the intuition that most of us have about Amelie giving her landlady a letter that will make her ecstatically happy is a positive one; certainly more positive than the intuition about telling the truth in that situation. Therefore, Amelie had, and most of us have, 'the intuition' that false pleasures are better for Amelie's landlady's well-being than the truth in this scenario. It should be clear that this thought experiment does not directly argue against the claim that pleasures based on truth contribute more to well-being than false pleasures. Rather, it is meant to raise doubts about the strength of the connection between truth and well-being, especially when compared to the strength of the connection between pleasure and well-being.

The following example, henceforth referred to as the Deceived Anthropologist thought experiment, aims to provide more direct intuitive evidence that false pleasures can be equally as good as pleasures based on truth or reality. Imagine an anthropologist looking back at her life from her death bed. She holds herself to have had an excellent life. She had a hugely rewarding career as the much-esteemed head of the International Anthropologists Society. And, it was there that she met and married her loyal husband, with whom she raised three lovely children. Her remarkable academic career was entirely based on her ground-breaking work explaining the demise of an ancient culture, the Lokians. Indeed, all of the joys of her career: the high position, the world travel, and the large and attentive audiences, and indeed all of the joys of her whole life, resulted from her well-received account of the Lokians' demise. This anthropologist certainly had a life full of positive well-being.

Now imagine that the great pleasures of the anthropologist's life were all the result of an ancient deception. The Lokians were a mischievous culture who lived on an isolated island, which they steadily deforested, resulting in the wind and rain removing all of their remaining food sources. They knew that they had to relocate quickly if they were to survive. But, before they left, they perfectly planted many fake clues to make any future visitors think that the Lokians had all perished on the island as a result of some bizarre accident. Unfortunately, a few days into the Lokians mass emigration a huge storm scuttled all of their ships. The remains of the Lokians and their escape vessels sank to the bottom of the ocean where they were completely devoured by giant deep-sea spiders.<sup>85</sup> Since the only evidence left of the Lokians' demise was the fake evidence they planted on their isolated island, *no one would ever discover the real truth of the matter*. Indeed, even if someone knew the truth of this matter, they would be ridiculed if they ever asserted it because of the complete lack of evidence for it and the considerable evidence against it.

---

<sup>85</sup> For more information on Giant deep-sea Sea Spiders, which devour sessile organisms through their long proboscis, see <<http://www.deepseaphotography.com/data.php?id=29400>> (accessed on 4/03/2008).

In her youth, the anthropologist fell for the Lokians' trick, producing several studies which took the false evidence planted by the Lokians at face value. Despite being based on (completely undiscoverable) falsity, these studies went on to be published and led to the rest of her most-enjoyable personal and work lives. Now imagine a life exactly the same in all respects, except that the anthropologist in this case was not tricked by any mischief on the part of the Lokians, since those ancient tricksters were wiped out before they had the chance to play any practical jokes. So, both anthropologists experienced very rewarding personal and work lives but one of them owes all of the great pleasures in their life to an undiscoverable deception that she had fallen for.

Did the life of the Deceived Anthropologist have any less well-being than the Non-Deceived Anthropologist? Intuitively, the Deceived Anthropologist had a life full of well-being and so too did the Non-Deceived Anthropologist.<sup>86</sup> Indeed, this seems correct *even though* all of the good things in the Deceived Anthropologist's life were a result of deception. Therefore, this case provides intuitive evidence that whether pleasures are based on truth or on deception seems to have no bearing on the extent to which those pleasures contribute to well-being. As it stands, then, the Deceived Anthropologist thought experiment provides intuitive evidence that pleasure can contribute to well-being *regardless* of its basis in truth or falsity. Furthermore, when the

---

<sup>86</sup> At this point, Kagan might respond with a reason-based, rather than intuitive, response to the thought experiment. He might claim that the Deceived Anthropologist really desires not to be deceived in the way she is, that she really desired to know the truth about the demise of the Lokians. If this is true, which we can assume it to be so, the desire-fulfilment account of well-being would hold the Deceived Anthropologist's life as worse than the Non-Deceived Anthropologist's life. However, we could imagine that if the Deceived Anthropologist's desire to really know what happened to the Lokians was fulfilled when she first encountered the Lokians' isolated island, that her life would have been very different and presumably worse for her. Since there was no evidence at all to support the real account of the Lokians' demise, any attempt to publish such a tall story would be ridiculed. This would most likely have resulted in her never becoming an anthropologist (let alone the head of the International Anthropologists Society), never meeting her loyal husband and perhaps never having three loving children. Considering the moment where the Deceived Anthropologist first thinks about the demise of the Lokians, it seems that getting what she really wanted in that moment would be bad for her well-being, while being deceived bodes very well for her future life. This result provides intuitive evidence that appears to undermine Kagan's potential response and the notion that getting what you *really* want is good for you, which, since Kagan puts normative stock in intuitive responses, should be enough to prevent him from raising such a response.

intuitions aroused by the Deceived Anthropologist thought experiment are compared with the intuitions produced by contemplation of Kagan's (1998) Deceived Businessman thought experiment, it is clear that they contradict each other. This contradiction gives even further reason to believe that intuitions elicited by unrealistic thought experiments are unreliable, since only one of the two can be correct, thereby casting doubt on the so-called intuitive evidence aroused by the Deceived Businessman thought experiment. Indeed, such a contradiction calls for closer analysis because, if these two intuitions contradict because the one aroused by the Deceived Businessman thought experiment is shown to be more likely to be misleading, then one of the two major pieces of so-called evidence for the controversial premise of the Argument from False Pleasures should be deemed too unreliable to use for that purpose.

The best method for discovering which intuition is most appropriate in this case is to analyse the differences between the scenarios in an attempt to find any discrepancies in the 'facts' that might affect the resultant intuitions. While the Deceived Anthropologist thought experiment seems to be quite unrealistic (a whole culture of practical jokers!), it is no more so than Kagan's (1998) Deceived Businessman thought experiment, which asks that we imagine that someone is detested by his family and colleagues without that having any negative effects on their behaviour towards him. However, there are discrepancies between the two scenarios regarding various issues with the deceptions in the protagonists' lives, including: the amount of deceptions, the distance between the deceivers and the deceived, and the plausibility of the complete concealment of the deceptions. These differences are discussed below.

It could be argued that because the Deceived Anthropologist thought experiment contains only one deception and because that deception happens in the distant past, attention is sufficiently drawn away from the deception, making it difficult for our

intuitive cognition to ‘pick up on it.’ However, it is clearly stated in the Deceived Anthropologist thought experiment that all of the great pleasures in her life were a result of the deception, providing good reason to think that the author has no intention of ‘hiding’ the deception. Both thought experiments make all of their protagonist’s great pleasures to be the result of deception. Having said this, it is true that the deception seems slightly removed from the Deceived Anthropologist while it is all around the Deceived Businessman.<sup>87</sup> In fact, the very reason for putting some distance between the Deceived Anthropologist and her deception was to make two important facts of the thought experiment more plausible.<sup>88</sup>

Firstly, in the Deceived Anthropologist thought experiment it is entirely plausible that she would never know of the deception that led to her many pleasures. Secondly, it is also abundantly clear that even if the deception were never revealed, the Deceived Anthropologist would not have a worse life because of the way her deceivers would treat her. This stands in stark contrast to the Deceived Businessman thought experiment, in which it seems impossible to believe that the Deceived Businessman did not know about the odious opinions his loved and liked ones harboured of him and that, even if he did not know about it, that their malice for him would not be reflected in worse experiences for him. This inherent intuitive imaginative resistance that the Deceived Businessman thought experiment elicits far outweighs any that the Deceived Anthropologist thought experiment does.<sup>89</sup> This shows that the intuitions aroused by the Deceived Anthropologist thought experiment are more likely to be

---

<sup>87</sup> It has been suggested (e.g. by Unger 1996) that we have an intuitive social distance bias that influences our unconscious moral evaluations, which might help to explain why the deception in the Deceived Anthropologist thought experiment is considered to have less effect on the well-being of the Deceived Anthropologist. However, it would not explain why the deception is not considered important at all. Furthermore, the extent to which social distance bias is an influencing factor in these two thought experiments would provide further evidence that our intuitions about moral issues can be, as Unger (1996) suggests, irrational and, therefore unreliable.

<sup>88</sup> Note that these facts are implied in the Deceived Anthropologist thought experiment and stipulated in the Deceived Businessman thought experiment. However, this makes little difference because while the facts are obvious in the Deceived Anthropologist thought experiment without needing to be explicitly stated; they are not obviously accepted in the Deceived Businessman thought experiment *despite* being stipulated!

<sup>89</sup> The problem of intuitive imaginative resistance to the Deceived Businessman thought experiment is taken up in more detail below.

appropriate than those that arise in response to consideration of the Deceived Businessman thought experiment. Therefore, it is more likely that the common intuitive response to the Deceived Businessman thought experiment: that pleasure based on truth contributes more positively to well-being than pleasure based on falsity, is probably wrong.

This final piece of intuitive evidence is designed to show that pains based on falsity are equally as bad as pains based on truth, in order to further evidence the same kind of claim for pleasures. Imagine that you are suffering physical and psychological pain based on falsity, such as feeling pain from a phantom limb and being upset about the (faked) death of the person you love stalking. The falsity that underlies these pains seems to have no bearing whatsoever on how bad they are for the well-being of the person suffering them. Concluding that false pains can be just as bad as real pains should lead us to ask why this should be any different for pleasures. Hedonism about Well-Being informs us that there is no important difference and that the experience of pleasure contributes positively to well-being just like the experience of pain contributes negatively to well-being. That is, experiences of pleasure and pain intrinsically contribute to well-being *regardless* of what caused them, at least in the cases where we cannot know of or be otherwise affected by those causes.

Consideration of these three counter-thought experiments elicits intuitive evidence that pleasures based on truth are just as good for our well-being as pleasures based on falsity (that we never experience). This evidence achieves two of the goals of this essay. First, the fact that the intuitions elicited contradict the intuitions aroused by the Deceived Businessman thought experiment undermines the reliability of intuitive evidence from unrealistic thought experiments generally. Secondly, these three counter-thought experiments all undermine the so-called intuitive evidence from the Deceived Businessman thought experiment, which purports to support Premise 2 from the Argument from False Pleasures: that ‘pleasure based on truth, or something

like it, contributes more positively to well-being than pleasure based on falsity'. Indeed, the weight of these three counter-thought experiments provides a compelling reason to hear both reason-based evidence against the intuitions raised by the Deceived Businessman thought experiment and potential alternate causal explanations for those intuitions being misleading in this case.

There is a strong reason-based argument that the lives of the deceived and the non-deceived businessmen went just as well for each of them. In Kagan's (1998, p34) words, the argument is: "what you don't know can't hurt you", which mirrors part of Hedonism about Well-Being's fundamental rationale: that states of the world only affect an individual's well-being if it leads to them experiencing a positive or negative mental state. The challenge posed by this argument is: how it is possible that the deception in the Deceived Businessman's life affected his well-being in any way if he never experienced anything bad whatsoever because of it?

To meet this challenge, proponents of the Deceived Businessman thought experiment must provide good reason to believe that something which is not experienced in any way can affect a person's well-being. The only recourse for proponents of the Deceived Businessman thought experiment at this stage is to posit a fundamental rationale of desire-fulfilment accounts of well-being and attempt to support it with intuitive evidence. They would respond by claiming that a person getting what they *really* want makes their life go better, *regardless* of if they know about or experience it in any way. And, they would support this with the intuitions aroused by the Deceived Businessman thought experiment. Again, the debate seems to reach the same stalemate of fundamental rationales discussed above. However, the intuitive support for the rationale that a person getting what they *really* want makes their life go better, *regardless* of if they know about or experience it in any way has been comprehensively scrutinised in this chapter so far. Furthermore, the aim of this essay is to take steps towards reviving Hedonism about Well-Being *to the level of plausibility*,

which means that a stalemate with a theory of well-being that is considered plausible would not be such a bad result. However, the case against the intuitive evidence provided by the Deceived Businessman thought experiment can be continued with further reason-based arguments demonstrating that it is irrelevant in assessing Hedonism about Well-Being.

The final attack on the relevance of the intuitive evidence elicited by the Deceived Businessman thought experiment in evaluating Hedonism about Well-Being, and why this strike is fatal, will now be explained. Intuitions in general and the specific intuition raised by the Deceived Businessman thought experiment have already been brought into disrepute by the preceding arguments. Now, an alternative causal explanation of the intuition aroused by contemplation of the Deceived Businessman thought experiment will be posited and shown to be more plausible than the causal explanation offered by Kagan (1998).

Kagan's (1998) argument clearly implies that the negative intuition that arises upon contemplation of the Deceived Businessman's life is caused by his pleasures seeming to contribute less to his well-being because they are based on falsity. The argument runs as follows: since the falsity is the only difference between the Deceived and the Non-Deceived Businessman's pleasures in the thought experiment, then falsity must be the cause of the negative intuition that the Deceived Businessman's life elicits. However, this argument overlooks the possibility that the thought experiment might not be processed by our intuitive cognition exactly as it is written. As mentioned above, the probabilistic processing that creates our intuitions makes predictions based on how our prior experiences allow us to unconsciously interpret the stimuli we are currently experiencing. This feature of our intuitive cognition makes it virtually impossible to process all aspects of the Deceived Businessman thought experiment exactly as it is stipulated. In particular, our intuitive cognition will fail to accurately process the stipulations that the Deceived Businessman is on his deathbed and that



he could be constantly deceived without that affecting how others would behave towards him. These two factors of the Deceived Businessman thought experiment will be explained in more detail below and will be used to show why it is much more plausible that the negative intuition elicited by contemplation of the Deceived Businessman's life is caused by the fear of finding out about the deceptions and of suffering because of the deceptions *even if* we never found out about them.

First, our intuitive cognition will fail to accurately process the stipulation that we are on our deathbed because almost all of us (currently living people) have never experienced the feeling that we are just about to die. Since our intuitive cognition cannot process from this particular vantage point, then we are likely to evaluate the predicament of the Deceived Businessman as if he had a future of an at least short, if not indeterminable, length. From this vantage point, the possibility that the Deceived Businessman would never discover the deceptions of all those around him seems extremely unlikely. All of our experience of mass deception, which is unlikely to be nearly as mass as in the thought experiment, tells us that the deceived will find out. If, as I have argued, our intuitive cognition cannot take account of the stipulation that the Deceived Businessman will never find out about the deception all around him, then it seems plausible to think that the negative intuition is more likely to be caused by our intuitive forecast that the deceptions will be discovered and that negative experiences will result.

Secondly, regardless of if our intuitive cognition can properly process the idea that the Deceived Businessman would never find out about all of this deception, it is still practically impossible for our intuitive cognition to process the stipulation that none of the mass deception results in negative experiences for the Deceived Businessman. Indeed, it is particularly difficult to consciously imagine being deceived and yet never

experiencing this deception.<sup>90</sup> Typically, when people loathe us, we can tell. Furthermore, our experiences of interactions with people who detest us are usually much less pleasant than those with people who love and respect us. For these reasons, our probabilistic intuitive cognition would not be able to compute the businessmen having equally pleasurable experiences when one businessman's experiences are with a group of people who abhor him and the other's are with a group of people who admire him. This means that the intuitions that arise in response to the Deceived Businessman thought experiment cannot follow a vital stipulation of the scenario. If we cannot intuitively process the stipulation that the Deceived Businessman never experiences anything different to the Non-Deceived Businessman, despite being deceived by everyone around him, then we cannot intuitively 'isolate' the unknown and un-experienced deception like we are supposed to. The result of this is that it seems much more plausible to think that the negative intuition that the Deceived Businessman thought experiment elicits is caused by our experiences of people having negative mental states when they are deceived.

Therefore, while proponents of the Argument from False Pleasures would assert that the cause of the negative intuition that arises from contemplation of the Deceived Businessman's life is that false pleasures are worse than pleasures based on truth, an alternative causal explanation is more plausible. Because of the way our intuitive cognition processes information it is more plausible to think that the negative intuition arises from our past encounters of having negative experiences when we are

---

<sup>90</sup> Consider all of times you have experienced being deceived (which necessitates that you found out about them). Now consider having the chance to repeat those experiences. Presumably, the thought of reliving some of your worst experiences arouses a negative intuition. Now consider all of the times that you have been deceived but never experienced in anyway (which necessitates that you did *not* find out about them). Now consider having the chance to repeat those experiences. Presumably, the thought of reliving these moments would be the same as reliving any particular moment in your life (in which you were not experiencing being deceived). How would you say each of these two types of events have affected your well-being? Obviously, your experiences of deception were bad, but assessing the impact of the latter type of deception is much more difficult. First of all, you cannot know that any unknown and un-experienced deceptions have ever happened to you. Secondly, even if they had happened to you, they have no phenomenology whatsoever, which means that you would not know what it was like when they did happen. For these reasons it seems hard to imagine how unknown and un-experienced deceptions can negatively contribute to our well-being.

deceived (regardless of if we know about the deception) and our intuitive forecast that the deceptions will be discovered and more negative experiences will result. Since this more plausible alternative causal explanation can ground all of the negativity in mental states, it can be explained within a hedonistic framework and, therefore, does not constitute an objection to Hedonism about Well-Being. Indeed, this discussion has shown that the negative intuition elicited by the Deceived Businessman thought experiment is irrelevant in supporting the premise that pleasures based on falsity do not contribute as positively to well-being as pleasures based on truth. The corollary of this is that the intuition aroused by the Deceived Businessman thought experiment, as presented by Kagan (1998), is too unreliable to support the Argument from False Pleasures and so is irrelevant for assessing Hedonism about Well-Being.

Is it possible to adjust the Deceived Businessman thought experiment in any way so that it can avoid this last and most telling objection to its relevance for assessing Hedonism about Well-Being? It seems not. Since the Deceived Businessman thought experiment makes it clear to the reader that the Deceived Businessman is deceived, it is impossible for our intuitive cognition to forget that fact. The only possible way to avoid this problem of intuitive imaginative resistance would be to restate the thought experiment without any mention of the deception and ask which life seems better. If we did this, though, our intuitions about the two lives would be exactly the same, since each businessman's life would be described as the same *from their perspective*. This is important because this essay concerns the concept of well-being: the good life *for the one living it*.

Proponents of the Argument from False Pleasures might protest at this point that this essay is denying other theories of well-being the chance to compete in the Deceived Businessman thought experiment by limiting it to what is experienced by its protagonists. However, this is not the case. This essay is only insisting that, given internalism about pleasure and *for the purposes of gathering intuitive evidence only*, this is how

the thought experiment must run if it is not to arouse inappropriate intuitions caused by the natural limitations of intuitive cognition. If the Deceived Businessman thought experiment were to be analysed by reason alone, then it could remain in its original form. When reason alone is used to assess the Deceived Businessman thought experiment, the fundamental rationales of hedonism and desire-fulfilment clash head-on, which as mentioned above results in something of a stalemate. The implications of such a stalemate are that the Deceived Businessman thought experiment offers no decisive evidence that pleasures based on falsity contribute any less positively to well-being than pleasures based on truth and, therefore, that the Argument from False Pleasures must be supported by stronger evidence if it is to render Hedonism about Well-Being implausible.

## **Extinguishing the Experience Machine**

Those who argue that Hedonism about Well-Being is a fundamentally flawed theory most frequently use The Experience Machine thought experiment as evidence. Therefore, this thought experiment is the most important thorn to remove from Hedonism about Well-Being's side. In this section, the Experience Machine thought experiment is briefly explained again, including a discussion of what intuitive and reason-based evidence it supposedly provides for the Argument from False Pleasures. It is argued that if the Experience Machine thought experiment provides any evidence at all in support of the Argument from False Pleasures, then that evidence is purely intuitive. This intuitive evidence (the intuition that pleasures based on reality contribute more to our well-being than those which are not) is then shown to be unreliable and irrelevant for the purposes of supporting the Argument from False Pleasures.

First, it is explained how the Experience Machine thought experiment is argued to support the Argument from False Pleasures. Then, intuitive and reason-based

evidence is used to show that it is highly improbable that our negative intuition about getting into the Experience Machine is caused by the fact that pleasures based on reality contribute more to well-being than those which are not and to provide further reason to mistrust our intuitive responses to unrealistic thought experiments generally. An attempt to adjust the Experience Machine thought experiment to avoid this problem is then attempted on behalf of the proponents of the Argument from False Pleasures. And, although the result of this adjusting: the Pleasure Machine thought experiment, avoids some of the problems plaguing the Experience Machine thought experiment, it too is ultimately found to have no normative significance for evaluating Hedonism about Well-Being. These points lead to the conclusion of this section: that the only potential evidence that the Experience Machine thought experiment might provide in support of the Argument from False Pleasures has no normative significance for evaluating of Hedonism about Well-Being and, therefore, should be discounted from any such assessment.

In the Experience Machine thought experiment, Nozick (1974, p42) asks us to imagine a machine built by “superduper neuropsychologists” that can provide us with any experience we desire, such as the pleasure of publishing an award-winning novel. The machine can also provide a comprehensive range of other amazing experiences that we might never have thought of, such as the pleasure of having all of your meals cooked to perfection by the world’s best chefs. Nozick invites us to plug into the Experience Machine, but expects that, even after a two-year test run, we will not accept his invitation (1974, p42). Indeed, when posed with this choice, most people tend to forgo a life of endless pleasure in the Experience Machine for what they currently have, a presumably less pleasurable but more real life. This choice appears to be based on the “firmly held” negative intuitions aroused by considering a life plugged in to the Experience Machine (Sobel 2002, p244). Indeed, Nozick devised the Experience Machine thought experiment to prove that more than just how our

“experiences feel from the inside” matters to us (Nozick 1974, p42).<sup>91</sup> In standard form, the argument would look something like this:

- P1) The Experience Machine could give us any experiences, including all of the best experiences.
  - P2) If all that matters to us is how our experiences feel from the inside, then we would want to get into the Experience Machine.
  - P3) People do not want to get into the Experience Machine.
- 
- c) Therefore, more than just how our experiences feel from the inside matters to us.<sup>92</sup>

Although Nozick did not claim to have all the answers to the question of what, “other than how our lives feel from the inside”, is important to us, he did have some suggestions. Nozick (1974, p43-44) briefly considered being a certain kind of person and achieving certain things as possible reasons for our not wanting to get into the Experience Machine. However, he quickly dismissed these possibilities, noting that more specific machines could offer such experiences and we would still not want to plug into them.<sup>93</sup> Nozick then proposed that our wanting to *really* experience the limitless reality of living our own real life, as opposed to having a machine live it for us, is what prevents us from plugging in to the Experience Machine (1974, p43-45). Furthermore, it is clear from his writing<sup>94</sup> that Nozick believed that more than just how our experiences feel from the inside matters to us *because* a life in which we experience reality is better than a life in which our experiences seem real but are not.<sup>95</sup>

---

<sup>91</sup> Nozick wanted to make this point so that he could extend it to animals (1974).

<sup>92</sup> This argument is valid and could easily be made sound by softening P3 and the conclusion to acknowledge that not all people choose not to get into the Experience Machine. Nevertheless, the majority of people do not want to get into the experience machine and that majority is probably vast enough to ignore the few dissenters.

<sup>93</sup> Nozick (1974, p43) also states that getting in to the Experience Machine is kind of like committing suicide. However, getting in to the Experience Machine would only be like committing suicide in that you would no longer get to experience reality. Indeed, it is markedly different in some ways. For example, compare not being conscious at all to consciously experiencing the most amazing pleasures that life has to offer (that seem real but are not). Clearly, death is much worse than being in the Experience Machine.

<sup>94</sup> See for example: “[In the Experience Machine] there is no *actual* contact with any deeper reality, though the experience of it can be simulated” (Nozick 1974, p43). And: “What is most disturbing about [Experience Machines] is their living of our lives for us” (Nozick 1974, p44). And: “what we desire is to live (an active verb) ourselves, in contact with reality” – his brackets (Nozick 1974, p45).

<sup>95</sup> Note that I am not surreptitiously slipping in the word ‘experience’ in my paraphrase of his position. Indeed, Nozick does seem to have carefully avoided using the word ‘experience’ when explaining why being “in contact with reality” matters to us (Nozick 1974, p45). However, being in contact with reality but not experiencing it is clearly not what Nozick was getting at. Imagine a man with no sense organs. He is living a real life and is in contact with reality but not in a meaningful way because he does not experience it

It is imperative to note here that Nozick provided the importance to us of real experiences as the reason *for* our intuitive response to the Experience Machine thought experiment. However, he provided no direct evidence for *why* experiences based on reality *should* matter more to us than those that are not.

It appears that Nozick (1974), like many other writers on this topic, assumed that it is self-evident that experiences based on reality matter more to us than those that are not. Indeed, all Nozick has done in his passage about the Experience Machine thought experiment is to state the proposition and hope that readers agree with him about its self-evidence. If this approach provides any evidence at all for the proposition that experiences based on reality matter more to us than those that are not, then that evidence is purely intuitive.

Proponents of the Argument from False Pleasures take a similar approach when they refer to the Experience Machine thought experiment in order to provide evidence that pleasurable experiences based on reality matter more to us than those which are not. By replacing ‘experiences’ with ‘pleasurable experiences’, advocates of the Argument from False Pleasures can make the Experience Machine thought experiment more specific to their cause against hedonists without losing anything important. (Nothing important is lost in this transition because hedonists assert that the only valuable experiences are pleasurable ones anyway). Sobel (2002, p244) discusses such an argument for why the Experience Machine thought experiment shows us that Hedonism about Well-Being is “wildly implausible”,<sup>96</sup> which is paraphrased below. It should be noted that Sobel focuses more on real *achievements* (2002, p244)<sup>97</sup>, as opposed to real *experiences* as the reason for why we have a negative

---

at all. When we are in contact with something in any kind of meaningful way, then presumably we will experience it.

<sup>96</sup> Sobel (2002, p240) classifies the type of hedonism discussed in this essay as “quantitative hedonism”, following Frankena’s (1963, p68-69) terminology for distinguishing between Bentham’s and Mill’s versions of hedonism.

<sup>97</sup> There will not be a lengthy case against the value of real achievements (above the pleasure they provide) in this essay because it only aims to defend hedonism about well-being against the Argument from False

intuition when we consider getting into the Experience Machine thought experiment. This difference is irrelevant though, because Sobel (2002) believes that our negative intuition about getting into the Experience Machine is sufficient to make it bad for our well-being. Sobel would be a staunch supporter of P6,<sup>98</sup> the most important premise of the argument below, the premise that this essay goes on to argue is false.

- P1) A normal life would have some pleasurable experiences and they would be based on reality.
  - P2) A life in the Experience Machine would give us a life full of pleasurable experiences but they would not be based on reality.
  - P3) If we prefer a normal life over a life in the Experience Machine, then pleasurable experiences based on reality matter more to us than those which are not.
  - P4) We prefer a normal life over a life in the Experience Machine.
  - P5) Therefore, pleasurable experiences based on reality matter more to us than those which are not (from P1-P4).
  - P6) If X matters more to us than Y, then X contributes more positively to well-being than Y.
- 
- c) Therefore, pleasurable experiences based on reality contribute more positively to well-being than those which are not (from P5-P6).

The above argument is supposed to show how the Experience Machine thought experiment provides reason to believe that pleasurable experiences based on reality contribute more positively to well-being than those which are not. However, it will now be argued that Premise 6 of the argument is false. Just because X *does* matter to us more than Y, does not entail that X *should* matter to us more than Y or that X contributes more positively to our well-being than Y. This fact remains true, even if X matters more to *everyone* than Y. Indeed, universal agreement in intuition does not

---

Pleasures. However, the objection might be able to be rejected quite quickly. Achievements come in many forms. An academic would probably consider becoming a professor an achievement, Hitler would probably have considered annihilating millions of Jews an achievement and a pebble collector would probably consider her 1,000+ pebble collection as an achievement. Consider working hard to attain an unknowable but great achievement. You have no idea what you will have achieved at the end of your toil but you know it will be great. For example, you might have written the best novel of all time or you might have managed to orchestrate the slaughter of millions of innocent people. If you would happily work hard for an unknowable but great achievement, then we have reason to think that achievement matters to us and, therefore, might have some value. However, we do not like the idea of working towards an unknowable but great achievement because the outcome might not please us. Therefore, while certain achievements may be valuable because of *what they achieve*, they are not valuable *because they are achievements*.

<sup>98</sup> As shown by the quote: “our intuition that the person hooked up to electrodes is not experiencing the most value a person could experience is firmly held and strikes at the heart of hedonism.” (Sobel 2002, p244).



guarantee reliability of that intuition (Liao 2007, p9). The reasons for why X matters to us are very important here. There are many potential reasons for X mattering more to us than Y. However, some of these reasons will have no relevance for well-being, which would make the fact that X matters to us have no significance to any evaluation of well-being we might make. For example, I might judge that having another Long Island Iced Tea is a really good idea and it might matter to me a lot that I get another one. Indeed, I might be very angry if I was refused another one. But this in no way entails that it would positively contribute to my well-being to actually have another Long Island Iced Tea.

Faced with such counter-examples, the only means to make Premise 6 true is to assume that a desire-fulfilment account of well-being is true. Making this assumption would allow for everything that matters to us being something that we desire to be a certain way, which directly links all the things that matter to us to our well-being. This strategy will not work, however, because assuming an opposing account of well-being to be true does not allow Hedonism about Well-Being a fair trial. Proponents of other views about well-being need to show why Hedonism about Well-Being is implausible *without* using their own fundamental rationales as self-evident premises. So, Premise 6 is false, but we still need to understand what the reasons are for it mattering that pleasurable experiences are based on reality (in the case of the Experience Machine thought experiment) and whether or not those particular reasons are relevant to evaluating Hedonism about Well-Being.

All that the proponents of the Argument from False Pleasures can say at this stage is that the reason pleasurable experiences based on reality matter more to us than those which are not is simply because pleasurable experiences based on reality contribute more positively to our well-being than those which are not. Furthermore, the only evidence available for this claim in the Experience Machine thought experiment is just the supposed intuition that our negative intuitive response to the idea of getting

into the Experience Machine is caused by the fact that pleasures based on reality contribute more positively to well-being than those which are not.<sup>99</sup> It should be clear that the Experience Machine thought experiment fails to provide substantial reason-based support for the Argument from False Pleasures, and that intuitive support is all it has to offer. Consequently, the remainder of this section is dedicated to arguing that our intuitive response to getting into the Experience Machine is not caused by the so-called fact that pleasures based on reality contribute more positively to well-being than those which are not. The success of the forthcoming arguments shows that *even if* it were a fact that pleasures based on reality contribute more positively to well-being than those which are not, the Experience Machine thought experiment certainly provides no evidence to reliably support such a conclusion.

While evaluating the supposed intuition that our negative intuitive response to the idea of getting into the Experience Machine is caused by the fact that pleasures based on reality contribute more positively to well-being than those which are not, intuitive and reason-based criticisms will be made. These intuitive and reason-based criticisms should be sufficient to convince even the most intransigent fans of intuitive evidence, such as Sobel (2002), that intuitive evidence from thought experiments is unreliable and that the specific intuition supposedly elicited by the Experience Machine thought experiment has no normative significance to the Argument from False Pleasures. First, the reason-based conclusions of Chapter 2 are applied to the Experience Machine thought experiment to show that intuitive responses to thought experiments are unreliable generally and indeed specifically in this case. Then, intuitive evidence in the form of the Trip to Reality thought experiment is provided to show that imagining pleasurable experiences that are not based on reality does not necessarily

---

<sup>99</sup> A problem with the argument used by the proponents of the Argument from False Pleasures here is that they are claiming a *causal* relationship between the reality of our experiences and our intuitive response to getting into the Experience Machine despite it being clear that there is no evidenced deductive or inductive argument offered to show any relationship at all. Indeed, all of the power of the argument used by the proponents of the Argument from False Pleasures is a correlation between our intuitions about getting into the Experience Machine and our intuitions about having false experiences. Nevertheless, evidence will be provided to show that there is no such causal relationship between our intuitions about the Experience Machine and the reality of our experiences, so proving that the two are merely correlated.

cause us to have negative intuitions. Following this, the 'Trip to Reality' thought experiment is compared to the Experience Machine thought experiment to provide evidence to believe that the negative intuitive response we have to getting into an experience machine is based on several other factors, all of which are completely irrelevant to the 'Argument from False Pleasures' attack on Hedonism about Well-Being.

The reason-based arguments of Chapter 2 showed that intuitions about thought experiments are not very reliable for evaluating theories of well-being because we cannot find out exactly what caused them. It was argued that it is impossible to discover the cause of our intuitions through introspection and that we should not be confident about any attempt to reconstruct the cause of the intuition. It was also argued that our intuitions about unrealistic thought experiments are especially likely to be unreliable because our intuitive cognition uses probabilistic processing based on similarities between what we are imagining and what we have experienced. And, since we do not always have appropriate experiences to be matched with unrealistic scenarios, the intuitive response is likely to be based on irrelevant factors and, therefore, it will probably be unreliable. The Experience Machine thought experiment is unrealistic in two ways: it is overly demanding of our intuitive cognition and it involves situations that we have had no experience of.

The Experience Machine thought experiment is unrealistically over demanding in that it requires our intuitive cognition to be able to properly take into account a stipulation that it simply cannot. It stipulates that while we are in the Experience Machine, we do not realise that our pleasurable experiences are not based on reality *after* the stipulation that our experiences in the Experience Machine are all fake (Nozick 1974, p43). Since our intuitive cognition processes the available data probabilistically and not deliberately, it cannot follow some verbal rules, such as forgetting something

that it is already aware of (under normal circumstances).<sup>100</sup> The correct way to gauge our intuitive response to what a life in the Experience Machine would actually be like for the one inside it would be quite different to Nozick's (1974) thought experiment. Instead, the thought experiment should make no mention whatsoever about machines or fake experiences, while providing lots of descriptions of pleasurable experiences.

The Experience Machine thought experiment is also unrealistic in that it is based on a fantastical contraption the likes of which no one has ever had any direct experience of.<sup>101</sup> Indeed, humankind may never experience anything like a perfectly working Experience Machine because it may not even be possible. Importantly, because we have no experience with anything like the Experience Machine, our intuitive cognition is likely to have difficulty evaluating what being in the Experience Machine would actually be like. Indeed, any intuitive evaluation of being inside the machine is likely to be based on movies and real life experiences that are unlikely to do the machine justice.

Since intuitions about thought experiments are generally unreliable and especially unreliable when the thought experiment is unrealistic, our intuitive response to the Experience Machine thought experiment should be considered so unreliable that it constitutes no normatively significant evidence. However, strong intuitions, like our response to the Experience Machine thought experiment, constitute a warning sign that should be investigated further. This conclusion is somewhat opposed to Sobel's (2002, p244) view that our intuitive response to the Experience Machine thought experiment is normatively significant evidence that puts the onus onto the hedonists to explain away the intuition by undermining its credibility. Indeed, Sobel (2002,

---

<sup>100</sup> A relevant abnormal circumstance that might help our intuitive cognition to forget some information might be the timely intervention of copious amounts of alcohol.

<sup>101</sup> Even if you were thinking of the last news report you have seen about virtual reality, you are probably still no where near having an idea of what being inside the Experience Machine would actually be like phenomenologically.

p244-245) believes that any attempt to explain away this intuition is doomed to failure. Nevertheless, the remainder of this section investigates the intuition further and in doing so will successfully rise to Sobel's (2002) challenge.

If it is true that our negative intuitive response to getting into the Experience Machine is caused by the fact that pleasures based on reality contribute more positively to well-being than those which are not, then we would expect that anyone in the Experience Machine would want to go back to reality and stay there if they could. As the following thought experiment, the Trip to Reality, shows, however, our negative intuitive response to the Experience Machine thought experiment is clearly not caused by the difference in how real the pleasurable experiences are because in this thought experiment we are much less inclined to choose reality over a life in an experience machine.

Imagine that your life is much like it is for you right now, except that it contains much more pleasure and less pain. To help yourself do this, imagine all of the most pleasurable experiences you have ever had and try to multiply them. For example, perhaps your job is challenging but also very interesting and rewarding. And, imagine that your outrageously good-looking spouse is a qualified masseuse who loves nothing more than giving you a daily after-work massage while one of your talented children prepares you a delicious dinner. You still have ups and downs, but everything generally feels much more like those amazingly pleasurable moments. Imagine that you leave your wonderful and loving family for a weekend to attend a conference on the Experience Machine thought experiment. While you are there, someone informs you that you are actually in an experience machine. She offers you a red and a blue pill. She explains that taking the blue pill will take you back to reality and taking the red pill will bring you back into the machine and totally wipe any memories of having being in reality. Being a curious philosopher you swallow the blue pill (something that most people probably would not do).

It turns out that reality is fairly similar to the world you have been experiencing inside the machine. The main difference, however, is that you do not feel as happy and pleased as you used to. You experience similar things in reality, it is just that none of them feel as good or are as enjoyable as before. You also often find yourself with strange aches and pains that you never used to experience, such as toothache. Exploring reality, you notice that lots of people are in experience machines. Those who are not in a machine claim to prefer not to be in an experience machine because they just do not like the idea of it. While exploring, you find out that all of your immediate family members, except for your father, are in experience machines. You spend time with him but you cannot help but feel like your interactions with him in reality pale in comparison to the amazing relationship you used to have with him. In fact you fight with your father in reality and it feels terrible. You consider trying to make friends with other real people but you can imagine that any friendship you can make will not be nearly as enjoyable as those you have in the machine. In fact, even though you know that they are not actually real, you miss your friends and family in the machine terribly. You have now experienced enough of reality to know that a life there would be brutish and full of suffering, especially when compared with your life in the machine, in which every day brought new pleasant experiences. It is time to make the choice. Will you take the red pill so that you can go back to your pleasurable life with no idea that it is not in fact real? Or will you throw the red pill away and try to make the best life you can in the harsh surrounds of reality?

In the Trip to Reality thought experiment, the thought of getting into an experience machine does not elicit the same intuition that exactly the same act does in the Experience Machine thought experiment. Indeed, when considering the Trip to Reality thought experiment, dramatically more people choose a life in an experience machine than when considering the Experience Machine thought experiment.<sup>102</sup> This essay goes on to argue that this is mainly because in the Trip to Reality thought

---

<sup>102</sup> At least in my experience of presenting each scenario at seminars.

experiment it is easier to see how much better our lives would be for us in an experience machine.

There are some important differences between the Trip to Reality and the Experience Machine thought experiments that can explain why they elicit such diverse intuitions. Significantly, however, they both involve exactly the same decision about getting into an experience machine: choosing a more pleasurable life over a more real life. This is important because it means that it would be incredibly unlikely that how *real* the pleasurable experiences in each life will be has anything to do with our intuitive responses to either scenario. Indeed, in light of this evidence, anyone claiming that our negative intuition about getting into the Experience Machine is caused by the difference in how real the pleasures would be in each life needs to provide an explanation for the different intuition raised by the Trip to Reality thought experiment. This result casts more doubt on the reliability of our intuitions about unrealistic thought experiments generally because it shows again how intuitions can be so easily misguided. This result also gives very good reason to believe that our intuitive response to the Experience Machine thought experiment in particular is unreliable because we no longer have a good candidate causal explanation for it. However, several other factors can adequately explain the different intuitive responses we have to getting into an experience machine in each of the thought experiments. These factors are discussed below.

One important difference between the Trip to Reality and the Experience Machine thought experiments is that the Experience Machine thought experiment taps into a group of overlapping psychological biases: risk or loss aversion and endowment, status quo or familiarity bias.<sup>103</sup> These biases are importantly linked by our valuing gains only about half as much as we value avoiding equivalent losses of things we already have or know – the status quo (Gilovich, Griffin & Kahneman 2002;

---

<sup>103</sup> Depending on who has written about these biases and when they were writing, some of them actually refer to the same phenomenon.

Kahneman & Tversky 2000; Kahneman, Knetsch & Thaler 1991; Tversky & Kahneman 1991). In the Experience Machine thought experiment the choice to get into the machine involves giving up something valuable that you have and are familiar with (your current life) for something that is supposedly more valuable but fairly unknown to you (a life in the machine). It is scary to consider swapping our whole life for another one if we are unsure what the new life will be like. This fear of the unknown that could be affecting our intuitions in the Experience Machine thought experiment is removed in the Trip to Reality thought experiment because being in the machine is linked much more directly to our current life and experiences. Indeed, the life in an experience machine that we imagine when considering the Trip to Reality thought experiment is much less scary than the life we imagine when considering the Experience Machine thought experiment. Even though Nozick (1974, p42) explains that we will experience great things in the Experience Machine, he describes it in a particularly unsettling way: as “floating in a tank, with electrodes attached to your brain.” However, if our intuitive cognition is picking up on risks in the Experience Machine thought experiment but not in the Trip to Reality thought experiment, is there a good reason for this?

In Sumner’s (1996, p95) discussion of the Experience Machine thought experiment, he raises the idea that we might be afraid that the machine will malfunction or be sceptical that a machine could really do what Nozick (1974, p42) claimed that it can. Silverstein (2000, p284) thinks this fear can be avoided by being “diligent in the use of our imagination” and exercising “a bit of mental dexterity”. However, I will argue that a better understanding of how our intuitive cognition works shows that Sumner’s (1996, p95) worries are valid. If we consider the Experience Machine thought experiment with just our deliberative cognition, then it is not so hard to accept the conceptual possibility of a perfectly reliable and functional experience machine. However, it was shown above that the Experience Machine thought experiment only



provides intuitive evidence to support the Argument from False Pleasures and intuitive evidence comes from intuitive cognition.

With the fact that our intuitive cognition is a probabilistic process that draws on our current and past experiences in mind, consider the amount of machines that you have experienced never having a fault and the amount that you have experienced having some fault. The overwhelming majority of machines that you have had experience of have had some functional deficit or have been in need of some sort of repair at some time during your lifetime. Now consider that the Experience Machine would have to be an extremely complicated contraption, involving a powerful computer. How many computers have you experienced that have never crashed? It is no wonder that our intuitive cognition cannot accept the stipulation that the machine will work perfectly. Where there is a risk to our whole life involved, it makes sense to be intuitively cautious. Now notice that, in the Trip to Reality thought experiment, the life in the machine is made to seem like the best parts of the life you are currently living. This encourages our intuitive cognition to infer that a life in an experience machine would not be effected by any machine failure because it is basing its projections on the life we have already experienced, which has never been effected by a machine failure in that way. It should be clear then, that a better understanding of how our intuitive cognition works reveals that the risk of machine failure or underperformance probably does have an effect on our negative intuitive response to getting into the Experience Machine in Nozick's (1974) thought experiment.<sup>104</sup>

The way our intuitive cognition works also prevents us from being able to fully appreciate what life in an experience machine would really be like. This is especially

---

<sup>104</sup> It is also possible that our intuitive aversion to getting into the Experience Machine is mainly because it is different and unknown (even when there are no obvious sources of risk). Imagine if the thought experiment asked instead for readers to choose between their current life and a much more pleasurable life in another (culturally similar) part of the (real) world. The thought experiment could also stipulate, like Nozick (1974, p43) does, that your loved ones can go with you. It seems like any intuitive resistance to choose the new, more pleasurable life would be mainly based on a status quo bias that overvalues the familiar and undervalues the unknown.

the case in Nozick's (1974) Experience Machine thought experiment but also affects our intuitive response to the Trip to Reality thought experiment. As discussed at length above, because our intuitive cognition is grounded in our experiences, intuitions about aspects of unrealistic thought experiments that we have never experienced are presumably created by probabilistically processing any experiences that we have had that are most similar to those un-experienced aspects. In his description of the Experience Machine, Nozick (1974, p42) does not go out of his way to describe the true extent of the pleasurable experiences that an experience machine could offer. He lists writing a novel, "making a friend" and "reading an interesting book" as examples of the amazing opportunities available to us in the Experience Machine (Nozick 1974, p42). More expressive language is used in the Trip to Reality thought experiment, in which readers are urged to imagine their most pleasurable experiences and to multiply them. The problem remains, however, that it is very hard to imagine what our most pleasurable experiences multiplied by ten, or even two, would be like. Our intuitive cognition simply cannot do justice to the experiences that an experience machine could offer and this causes our intuitive response to getting into an experience machine in either scenario less positive than it should be according to the stipulations of both thought experiments. Our intuitive inability to appreciate the amount of pleasure on offer from a life in the Experience Machine, just like our intuitive fear of machine failure, could explain our negative intuitive response to Nozick's (1974) Experience Machine thought experiment.

Finally, the way that our intuitive cognition works also means that Nozick's (1974, p43) attempts to unrealistically rule out moral obligation as a contributing factor to our intuitive response to the Experience Machine thought experiment will not work. A good reason for not wanting to get into an experience machine is that, in doing so, you would violate your moral obligations to others. Nozick (1974, p43) attempts to rule this out by saying that any significant others "can also plug in" to an experience machine so that you do not have to feel bad for leaving them behind. Putting it like

that seems to affect our intuitive response because it puts us in the position of having to encourage the people we have moral obligations towards to get into an experience machine just because we want to plug in and do not want to feel bad about it for the few minutes before we do. Treating other people as a mere means like this certainly seems intuitively immoral to us. So, it seems that Nozick's (1974) attempt to do away with interfering moral intuitions has failed in the Experience Machine thought experiment and that our moral obligations to others probably do help cause the negative intuition we have about getting in to the Experience Machine.

This effect that moral obligations could have on our intuitive response could be diminished by stipulating that everyone you know has already decided to get in an experience machine and that you can get in at the same time as them if you want. This stipulation might be difficult for our intuitive cognition to deal with, which is why it would only diminish and not eliminate this bias on our intuitive response. It should also be pointed out that any moral obligations which affect our decision to get into an experience machine should be eliminated if possible because we are assessing whether we can use the resulting intuitions as evidence for evaluating theories about subjective well-being, not moral theories generally. Indeed, it is perfectly compatible with Hedonism about Well-Being for some one to choose not to get into an experience machine for moral reasons. For instance, a hedonist could believe that a life in an experience machine is much better than a real one but they may choose to remain in reality so that they can perform some moral task. Acting rationally is often thought of as acting prudentially, or so as to achieve well-being. However, we also want to allow for non-prudential but morally good acts as rational in some cases (Gert 2004), such as foregoing a life of amazing pleasure so that we can help others.

The influence of moral responsibility is somewhat limited in the Trip to Reality thought experiment because moral obligations are limited to fake people in your machine life and one grumpy family member (who you do not feel particularly

attached to) in reality. This difference between the Experience Machine and the Trip to Reality thought experiments also helps to explain the different intuitive responses they elicit to the thought of getting in to an experience machine.

When all of these important contextual differences between the Experience Machine and the Trip to Reality thought experiments are considered together, there is good reason to think that they were the cause of our differing intuitive responses to getting into an experience machine in each scenario. Even if some other contextual differences between the two thought experiments were partly responsible for causing the differing intuitions, only an inference to the worst explanation would conclude that the reality of the pleasurable experiences is doing any work in either scenario. Whether or not the pleasure is based on reality clearly has no effect on our intuitive response to the Trip to Reality thought experiment, which means that there is also good reason to think that it has no effect on our intuitive response to the Experience Machine thought experiment.

If proponents of the argument from False Pleasures wish to continue using the Experience Machine thought experiment as evidence to support the claim that pleasures based on truth or reality are better than those which are not, then they need to provide a good reason for why reality affects our intuitions about Experience Machine thought experiment but not the Trip to Reality thought experiment. Indeed, considering the aforementioned other important differences between the two thought experiments, which are much more likely to cause different intuitive responses, it seems very doubtful that any such reason exists. Therefore, unless there is some way that the Experience Machine thought experiment can be tweaked to avoid these biases, the intuitions it provides should be considered unreliable and irrelevant in supporting the Argument from False Pleasures and, therefore, to hold no normative significance for evaluating Hedonism about Well-Being.

If the Experience Machine thought experiment were going to be tweaked in an attempt to offer normatively significant evidence to the Argument from false Pleasures, it should be made more specific and attempt to remove all potential intuitive biases. The following attempt to tweak the Experience Machine thought experiment: the Pleasure Machine thought experiment, stipulates a more specific and less intimidating machine and reduces the influence of some of the contextual factors on the intuitive response it elicits.

The Pleasure Machine is like a large comfortable hat that triangulates electromagnetic waves to stimulate various areas of the brain. When you have a pleasure machine on, its functions prevent your senses from sending information to your conscious mind like they normally do. Furthermore, while blocking your external senses, it stimulates the pleasure centre of the brain continuously. The result is total sense deprivation, you cannot see, hear, taste, smell or feel anything, except for an overwhelming sensation of pleasure. Your body will be well looked after so that it remains in perfect health while you have a pleasure machine on. Pleasure machines have been successfully used by thousands of people for more than one hundred years and you have already thoroughly enjoyed every second of several extended 'pleasure holidays' with a pleasure machine on. You have also just heard that all of the people that you know have decided to spend the rest of their life with a pleasure machine on. Would you put on a pleasure machine for the rest of your life so that you could experience constant pleasure? Or would you prefer to remain in reality and experience real pleasures and pains?

Like the Experience Machine thought experiment, the Pleasure Machine thought experiment elicits a negative intuitive response to the thought of spending the rest of our lives in a machine, although, for most people, a weaker one. However, because of its specificity and avoidance of some irrelevant biases, the Pleasure Machine thought experiment has more chance of providing evidence to support the Argument from

False Pleasures. Unfortunately for the proponents of the Argument from False Pleasures, however, the Pleasure Machine thought experiment still only provides intuitive evidence and, as argued several times above, intuitive evidence is unreliable and should only be used as a warning sign. All that a hedonist has to say is: even if she has a negative intuition about it her reason tells her that choosing to live the rest of her life with a pleasure machine on would be much better for her well-being than remaining in reality. If Hedonism about Well-Being is considered plausible before this thought experiment's normative significance is investigated, then an unsupported intuition should not be considered evidence enough to undermine that plausibility. Regardless, the intuition that the Pleasure Machine thought experiment elicits still fails to be compelling because of the problem raised by considering the Trip to Reality thought experiment, in which how real our pleasures were had no effect on our intuition in that case.

If proponents of the Argument from False Pleasures wished to use the Pleasure Machine thought experiment as evidence, then they would have to provide a good reason for why reality matters in the case of the Pleasure Machine thought experiment but not in the Trip to Reality thought experiment. Indeed, it seems unlikely that there is such a reason, especially when it is considered that the negative intuitive response to putting on a pleasure machine for the rest of our lives is more likely to be caused by our inability to imagine what pleasure without other sensations would be like. Since we have no experience of extreme pleasure without obvious cause or stimuli to associate it with, our intuitive cognition cannot accurately evaluate what having a pleasure machine on would really be like.<sup>105</sup> Concerns about boredom and the pleasure machine not being able to make us continuously feel pleasure may also make our intuition about getting in to it more negative. However, that only reinforces the point that our intuitive cognition cannot properly evaluate a life with a pleasure machine on. For these reasons, tweaking the Experience Machine thought

---

<sup>105</sup> Indeed, I cannot help but imagine the oscillating colourful patterns of my Windows Media Player graphical display when attempting to imagine pleasure decoupled from all other sensations.

experiment into the Pleasure Machine thought experiment could make it more specific to the needs of the proponents of the Argument from False Pleasures. However, there would be little point, since both thought experiments still fail to produce any evidence that has normative significance for Hedonism about Well-Being.

This section has argued that, at least in terms of the Argument from False Pleasures, the Experience Machine thought experiment provides no evidence that is reliable or relevant for evaluating Hedonism about Well-Being. The argument began by showing that if the Experience Machine thought experiment provides any evidence at all in support of the Argument from False Pleasures, then that evidence is intuitive. The supposed evidence was identified as the intuition that pleasures based on reality contribute more to our well-being than those which are not. The exposing of this supposed evidence began by discussing the reasons to think that intuitions about unrealistic thought experiments are unreliable generally and especially in this case. This was followed by an explanation of the trip to Reality thought experiment, which produced exactly the opposite intuition and, thereby, had the effect of undermining the reliability of intuitions from unrealistic thought experiments generally and specifically the intuition that pleasures based on reality contribute more positively to well-being than those which are not.

Subsequently, several differences between the two thought experiments were discussed and it was argued that these differences were much more likely to be causing the differing intuitions in each scenario. These differences included intuitive fear of machine failure, our intuitive cognition not being able to properly evaluate a life in an experience machine and our moral obligations to others. Brief consideration was then given to how the proponents of the Argument from False Pleasures might alter the Experience Machine thought experiment in an attempt to overcome the problems raised in this section. The Pleasure Machine thought experiment was

suggested as an improvement on the Experience Machine thought experiment but it too failed to avoid all of the criticisms raised in this section and so proved ultimately unsuccessful in providing any normatively significant evidence to support the Argument from False Pleasures.

## Conclusion

This final chapter used the insights from Chapter 2 and the resulting New Method of Assessing Theories of Well-Being to explain how supposed intuitive evidence from thought experiments should be assessed. It was argued that intuitions should just act as warning signs, especially when they are elicited by unrealistic thought experiments and when there is reason-based evidence to use instead. A brief case was then made for the plausibility of Hedonism about Well-Being, before it was shown that Hedonism about Well-Being remains at least plausible in the face of the reason-based evidence that the Deceived Businessman and the Experience Machine thought experiments are widely considered to provide.

The intuitive evidence that each thought experiment is supposed to provide was then clearly articulated and challenged with both intuitive and reason-based arguments. Because the intuitive evidence raised by each of these thought experiments was taken seriously, the conclusions made about them will be significant both for readers who accept the New Method of Assessing theories of Well-Being and, importantly, those who do not. Even proponents of reflective equilibrium, who might value our intuitive responses to unrealistic thought experiments more highly than Chapter 2 argued that they should be, must heed the conclusions of this chapter because of the thorough way it refuted the intuitive evidence.

First, the negative intuition aroused by considering the Deceived Businessman's life: that false pleasures do not contribute as positively to well-being as pleasures based on truth, was shown to be both unreliable and irrelevant in assessing Hedonism about



Well-Being. Then, the negative intuition aroused by considering getting into the Experience machine, that pleasures based on reality contribute more positively to well-being than those which are not, was also shown to be both unreliable and irrelevant to assessing Hedonism about Well-Being. For both the Deceived Businessman and the Experience Machine thought experiments, attempts were made to adjust them to avoid some of the intuitive biases that they trigger. However, these attempts failed to avoid the major problems with each thought experiment, resulting in the strong conclusion that neither the Deceived Businessman nor the Experience Machine thought experiments provide any reliable support for the Argument from False Pleasures.

Since this chapter has shown that neither the Deceived Businessman nor the Experience Machine thought experiments provide any support for the Argument from False Pleasures, regardless of exactly how highly intuitions are valued as a source evidence for assessing theories of well-being, the Argument from False Pleasures becomes much less persuasive. As discussed earlier, the Deceived Businessman and the Experience Machine thought experiments are the most used sources of evidence to support the controversial premise of the Argument from False Pleasures. However, now that the evidence that they supposedly provide has been shown to hold no normative significance for evaluating Hedonism about Well-Being, the Argument from False Pleasures sorely lacks evidence to support the premise: that pleasure based on truth, or something like it, contributes more positively to well-being than pleasure based on falsity. Proponents of the Argument from False Pleasures, therefore, need to provide some other evidence to support their controversial premise if they desire to see their argument revived.

Furthermore, any new evidence that they do concoct will need to escape the criticisms brought against the Deceived Businessman and the Experience Machine thought experiments in this essay. It is not clear that this will be possible, especially if

the New Method of Assessing Theories of Well-being is accepted. Regardless, at this stage the Argument from False Pleasures is left with no good evidence to support its controversial premise and, therefore, provides no good reason to believe its conclusion: that Hedonism about Well-Being is false. This result is a very positive one for Hedonism about Well-Being, which has long been considered implausible because of the Argument from False Pleasures. Although the Argument from False Pleasures has often been used as a quick and easy way to dismiss Hedonism about Well-Being, there are other popular arguments against its plausibility. However, perhaps with the main objection to Hedonism about Well-Being refuted, more effort will be directed towards a complete revival of its plausibility by refuting the other arguments against it. Indeed, with the main argument against Hedonism about Well-Being refuted, now is the perfect time for a full-scale revival of Hedonism about Well-Being and perhaps hedonism generally.

---

# Conclusion

---

## Argument Summary

This essay took positive steps towards reviving Hedonism about Well-Being, a theory that has been perennially unpopular (Blake 1926; Crisp 2006a; Feldman 2004; Savery 1934; Silverstein 2000).<sup>106</sup> The most important step taken by this essay, to refute the Argument from False Pleasures, was vital for reviving Hedonism about Well-Being because it was the argument most frequently used to reject its plausibility. Another important step taken in this essay was to criticise the methods currently used to evaluate theories of well-being and to propose an alternative. It was argued that too much importance had been given to intuitive evidence from unrealistic thought experiments, despite their dubious epistemic credentials. For this reason, the alternate method proposed, the New Method of Assessing Theories of Well-Being, restricted the role of intuitive evidence from unrealistic thought experiments in evaluating theories of well-being. This step was important for strengthening the main argument of this essay against the normative significance of the Argument from False Pleasures in evaluating Hedonism about Well-Being.

The refutation of the Argument from False Pleasures was based on ruling out the normative significance of the so-called evidence provided for it by Nozick's (1974)

---

<sup>106</sup> That is, except for the brief period of British Empiricism, when the likes of Bentham (1789) and Mill (1861) were giving detailed accounts of hedonism (Sumner 1996).

Experience Machine and Kagan's (1998) Deceived Businessman thought experiments. The evidence supposedly provided by these two unrealistic thought experiments was argued to be mainly intuitive. And, while this essay averred that intuitive responses to unrealistic thought experiments do not constitute reliable evidence for evaluating theories of well-being, this intuitive evidence was nevertheless analysed in great detail. Then, bolstered by a better understanding of intuitive cognition, the intuitive evidence from the Experience Machine and Deceived Businessman thought experiments was challenged with both intuitive and reason-based evidence. The challenges showed that the intuitions elicited by each thought experiment were unreliable and irrelevant in supporting the Argument from False Pleasures. The result of this discussion was the conclusion that the Argument from False Pleasures lacks sufficient evidence to pose any challenge to Hedonism about Well-Being's plausibility. A more detailed, chapter-by-chapter, account of how this conclusion was reached and the other implications of this essay follow.

## Chapter Outline and Implications

In Chapter 1 of this essay, the scene was set by defining well-being as the prudentially good life, or the life that is good for the one living it. To set the target for revival, a specific internalist definition of hedonism about well-being was then established (referred to as Hedonism about Well-Being), which was defined as follows:

All pleasure, and only pleasure, intrinsically contributes positively to well-being and all pain, and only pain, intrinsically contributes negatively to well-being, where pleasure and pain supervene on mental states such that pleasure is the intrinsically good part of mental states that we are consciously aware of and pain is the intrinsically bad part of mental states that we are consciously aware of.

Chapter 1 also explained the Argument from False Pleasures and the two thought experiments that are used to provide evidence for it: the Experience Machine and Deceived Businessman. The *prima facie* strength of the Argument from False Pleasures was established based on the force of the evidence provided by the Experience

Machine and Deceived Businessman thought experiments. Following this, how theories of well-being are evaluated by moral philosophers was discussed, with reference to the work of Timmons (2006), Kagan (1998) and Griffin (1986). The approach moral philosophers take when evaluating theories of well-being was then formalised into the Old Method of Assessing Theories of Well-Being, which is not particularly different to a reflective equilibrium approach. Specific attention was also focused on the role intuitions play in the Old Method of Assessing Theories of Well-Being, especially in cases where Hedonism about Well-Being was being evaluated. Finally, the role of intuitions, which had been noted throughout the chapter, was discussed in more detail. Rachels' (2003) attempt to discern what might give our moral intuitions the authority of a competent judge was discussed. However, the various reasons to think that our moral intuitions do have such authority that Rachels (2003) offered were found to be unsatisfactory. This prompted the need for an in-depth investigation into exactly how our intuitions work in order to illuminate how they might have normative significance, which was then discussed in Chapter 2.

Chapter 2 provided a detailed account of moral intuitive cognition and discussed the role that Woodward and Allman (2007) recommend for intuitions in ethical theorising. First, Woodward and Allman's (2007) definition of moral intuitive cognition as a fast, unconscious probabilistic process that can compute large amounts of data in parallel and produce an intuition, a simple visceral sensation, was explained in detail and corroborated by other leaders in the field of cognitive science. Then, this greater understanding of moral intuitions was used to assess and mostly agree with Woodward and Allman's (2007) recommendations for appropriate use of moral intuitions in ethics. In particular, their recommendation that intuitive evidence from unrealistic thought experiments should not be considered reliable was agreed with and adapted into an argument for restricting the use of intuitive evidence from unrealistic thought experiments in evaluating theories of well-being.

The argument's conclusion was that intuitions are an unreliable source of evidence for evaluating unrealistic well-being-related thought experiments and, therefore, that they should not be used if other types of evidence are available. The argument was based on the premises that intuitions are often affected by biases that can make them unreliable and that the nature of moral intuitions means that we can never really know whether any particular moral intuition is biased, especially moral intuitions about unrealistic well-being-related thought experiments. This argument provided the basis for proposing the New Method of Assessing Theories of Well-Being. The New Method of Assessing Theories of Well-Being was then briefly explained, including a guide on how to apply it. Importantly, this guide included a description of how to deal with objections to theories that have already been established as plausible. The guide provided a particular focus on dealing with intuitive evidence from unrealistic thought experiments, in which it was argued that intuitive evidence from unrealistic thought experiments should just act as warning signs, especially when there is reason-based evidence to use instead.

Lastly, the implications of the New Method of Assessing Theories of Well-Being were discussed, pointing to the considerable effect it would have if applied to the current objections to theories of well-being and their unrealistic thought experiment-based intuitive evidence. Furthermore, there are many other implications of Woodward and Allman's (2007) conclusions about the normative significance of various types of thought experiments, especially in normative and applied ethical debates. Most importantly, anyone wishing to use unrealistic thought experiments and analogies to argue their point should consider if their use of those scenarios is generating normatively significant moral judgements or just moral intuitions based on irrelevant biases.

Chapter 3 used the insights from Chapter 2 and the resulting New Method of Assessing Theories of Well-Being to evaluate the plausibility of Hedonism about

Well-Being in light of the Argument from False Pleasures against it. First, a brief case was made for the plausibility of Hedonism about Well-Being. Then the evidence that the Deceived Businessman and the Experience Machine thought experiments are widely considered to provide was assessed according to the guide from Chapter 2. The intuitive evidence that each of the Deceived Businessman and the Experience Machine thought experiments is supposed to provide was clearly articulated and challenged with both intuitive and reason-based arguments. Because the intuitive evidence raised by each of these thought experiments was taken seriously, the conclusions made about them are significant for readers who accept the New Method of Assessing theories of Well-Being and, importantly, those who do not. Even proponents of reflective equilibrium, who might value our intuitive responses to unrealistic thought experiments more highly than Chapter 2 of this essay has argued that they should be, must heed the conclusions of Chapter 3 because of the thorough way the intuitive evidence was refuted.

First, the negative intuition aroused by considering the Deceived Businessman's life: that false pleasures do not contribute as positively to well-being as pleasures based on truth, was shown to be both unreliable and irrelevant to assessing Hedonism about Well-Being. This was mainly argued on the basis that it is at least very difficult for our intuitive cognition to abide by the Deceived Businessman's stipulations that we would never find out about the deception or experience any negative effects because of it.

Then, the negative intuition aroused by considering getting into the Experience Machine: that pleasures based on reality contribute more positively to well-being than those which are not, was also shown to be both unreliable and irrelevant to assessing Hedonism about Well-Being. This was mainly argued on the basis that the intuitive response to getting into the Experience Machine reacted to irrelevant biases, such as risk aversion and imaginative resistance (not being able to fully appreciate the life in

the Experience Machine). And, also because our intuitive responses to getting into the Experience Machine could easily be thrown off by our specific combination of moral emotions and obligations (which were supposed to have been ruled out by the stipulations of the thought experiment).

For both the Deceived Businessman and the Experience Machine thought experiments, attempts were made to adjust them to avoid some of the intuitive biases that they trigger. However, these attempts failed to avoid the major problems with each thought experiment, resulting in the strong conclusion that neither the Deceived Businessman nor the Experience Machine thought experiments provide any reliable support for the Argument from False Pleasures' attack on Hedonism about Well-Being's plausibility.

Since Chapter 3 showed that neither the Deceived Businessman nor the Experience Machine thought experiments provide any support for the Argument from False Pleasures, regardless of exactly how highly intuitions are valued as a source evidence for assessing theories of well-being, the Argument from False Pleasures must be seen as much less persuasive than it has been. As discussed earlier, the Deceived Businessman and the Experience Machine thought experiments are the most used sources of evidence to support the controversial premise of the Argument from False Pleasures. However, now that the evidence that they supposedly provide has been shown to hold no normative significance for evaluating Hedonism about Well-Being, the Argument from False Pleasures sorely lacks evidence to support the premise: that pleasure based on truth, or something like it, contributes more positively to well-being than pleasure based on falsity. Proponents of the Argument from False Pleasures, therefore, need to provide some other evidence to support their controversial premise if they desire to see their argument revived.



Furthermore, any new evidence that proponents of the Argument from False Pleasures do concoct will need to escape the criticisms brought against the Deceived Businessman and the Experience Machine thought experiments in this essay. It is not clear that this will be possible, especially if the New Method of Assessing Theories of Well-being is accepted. Regardless, at least for now the Argument from False Pleasures is left with no good evidence to support its controversial premise and, therefore, provides no good reason to believe its conclusion: that Hedonism about Well-Being is false. This result is a very positive one for Hedonism about Well-Being, which has for so long been considered implausible because of the Argument from False Pleasures.

The result of these arguments was to conclude that the Argument from False Pleasures lacks sufficient evidence to pose any challenge to Hedonism about Well-Being's plausibility. There are, of course, many other arguments against the plausibility of Hedonism about Well-Being. However, with the most frequently used argument against it rejected, it is hoped that more thought will be applied to refuting the other arguments against Hedonism about Well-Being. This increased attention, and continuing scientific developments, may eventually lead to a full revival of Hedonism about Well-Being's plausibility, which would allow it to regain equal plausibility with the other popular views about well-being. Indeed, if other attempts to make Hedonism about Well-Being more plausible are based on internalist experiential accounts, then it may turn out that Bentham's (1789) hedonism was in fact the most plausible account of well-being all along.

---

## References

---

Arneson, R. (1999). Human Flourishing Versus Desire Satisfaction, *Social Philosophy and Policy*, 16(1): 1-48.

Angner, E. (2005). *Subjective Measures of Well-Being: A Philosophical Examination*, Doctoral Dissertation, University of Pittsburgh. Available from: <<http://etd.library.pitt.edu/ETD/available/etd-08192005-130324/unrestricted/AngnerDissertation2005.pdf>>. Accessed on 21/12/2008.

Angner, E. (forthcoming). Subjective Measures of Well-Being: Philosophical Perspectives, in *Handbook on the Philosophical Foundations of Economics as a Science*, H. Kincaid & D. Ross (eds.), Oxford: Oxford University Press.

Angner, E. (in press). The Philosophical Foundations of Subjective Measures of Well-Being, in *Happiness and Capabilities*, L. Bruni & P. Porta (eds.), Oxford: Oxford University Press.

Anscombe, G.E.M. (1957). *Intention*, London: Basil Blackwell, 1958.

Aristotle (unknown). *Nicomachean Ethics*, Indianapolis & New York: Bobbs-Merrill, 1962.

Aslin, R.N., Saffran, J.R., & Newport, E.L. (1998). Computation of Conditional Probability Statistics by 8-Month-Old Infants, *Psychological Science*, 9: 321-324.

Audi, R. (1996). Intuitionism, Pluralism, and the Foundations of Ethics, in *Moral Knowledge?*, W. Sinnott-Armstrong & M. Timmons (eds.), Oxford: Oxford University Press, reprinted in Audi (1997).

Audi, R. (1997). *Moral Knowledge and Ethical Character*, Oxford: Oxford University Press.

Badcock, C. (2000). *Evolutionary Psychology: A Critical Introduction*, Cambridge: Polity Press.

Bagani, J. & Fosl, P.S. (2007). *The Ethics Toolkit: A Compendium of Ethical Concepts and Methods*, Wiley-Blackwell.

Bargh, J.A. & Chartrand, T.L. (1999). The Unbearable Automaticity of Being, *American Psychologist*, 54: 462-479.

Bastick, T. (1982). *Intuition: How We Think and Act*, New York: Wiley.

Bentham, J. (1789). *An Introduction to the Principles of Morals and Legislation*, Latest edition: Adamant Media Corporation, 2005.

Berridge, K.C. (2003). Pleasures of the Brain, *Brain and Cognition*, 52: 106-128.

Berridge, K.C. (2007). The Debate over Dopamine's Role in Reward: The Case for Incentive Salience, *Psychopharmacology*, 191: 391-431.

Berthoz, S., Armony, J.L., Blair, R.J.R., Dolan, R.J. (2002). An fMRI Study of Intentional and Unintentional (Embarrassing) Violations of Social Norms, *Brain*, 125: 1696-1708.

Biederman, I., & Shiffar, M.M. (1987). Sexing Day-Old Chicks: A Case Study and Expert Systems Analysis of a Difficult Perceptual-Learning Task, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13: 640-645.

Blackburn, S. (2005). Hedonism, in *Oxford Dictionary of Philosophy*, S. Blackburn. (ed.), Oxford University Press.

Blake, R.M. (1926). Why Not Hedonism? A Protest, *International Journal of Ethics*, 37(1): 1-18.

Bornstein, R.F. (1989). Exposure and Affect: Overview and Meta-Analysis of Research, 1968-1987, *Psychological Bulletin*, 106: 265-89.

Bornstein, R.F. (1999). Source Amnesia, Misattribution, and the Power of Unconscious Perceptions and Memories, *Psychoanalytic Psychology*, 16(2): 155-178.

Boyle, T.C. (2003). *Drop City*, London: Bloomsbury.

Bradburn, N.M. (1969). *The Structure of Psychological Well-Being*, Chicago: Aldine.

Brandt, R. (1979). *A Theory of the Good and the Right*, New York: Oxford University Press.

Brentano, F. (1969). *The Origin of Our Knowledge of Right and Wrong*, O. Kraus (ed.), R. Chisholm (English edition ed.), trans. by R. Chisholm & E. Schneewind, London: Routledge & Kegan Paul.

Brink, D. (1989). *Moral Realism and the Foundations of Ethics*, Cambridge: Cambridge University Press.

Brink, D. (2006). Some Forms and Limits of Consequentialism, in *The Oxford Handbook of Ethical Theory*, D. Copp (ed.), Oxford: Clarendon Press.

Broad, C.D. (1930). *Five Types of Ethical Theory*, London and Boston: Routledge and Kegan Paul.

Bruner, J. (1960). *The Process of Education*, Cambridge, MA: Harvard University Press.

Buss, D.M. (2000). The Evolution of Happiness, *American Psychologist*, 55: 15-23.

Casebeer, W.D. (2003). Moral Cognition and its Neural Constituents, *Nature Reviews Neuroscience*, 4: 841-846.

Casebeer, W.D. & Churchland, P.S. (2003). The Neural Mechanisms of Moral Cognition: A Multiple-Aspect Approach to Moral Judgment and Decision-Making, *Biology and Philosophy*, 18:169-94.

Castelli, L., Zogmaister, C., Smith, E.R. & Arcuri, L. (2004). On the Automatic Evaluation of Social Exemplars, *Journal of Personality and Social Psychology*, 86: 373-387.

Chaiken, S., Liberman, A., & Eagly, A.H. (1989). Heuristic and Systematic Information Processing Within and Beyond the Persuasion Context, in *Unintended Thought*, J.S. Uleman & J.A. Bargh (eds.), New York: Guilford Press, p 212-252.

Clark, A. & Chalmers, D. (1997). The Extended Mind, *Analysis*, 58: 7-19.

Clinchy, B. (1975). The Role of Intuition in Learning, *Today's Education*, 64: 48-51.

Craig, A.D. (2004). Human Feelings: Why are Some More Aware than Others?, *Trends in Cognitive Sciences*, 8: 239-241.

Coghill, R.C., McHaffie, J.G. & Yen, Y.F. (2003). Neural Correlates of Interindividual Differences in the Subjective Experience of Pain, *Proceedings of the National Academy of Sciences USA*, 100: 8538-8542.

Crisp, R., (2002). Sidgwick and Intuitionism, in *Ethical Intuitionism: Re-evaluations*, P. Stratton-Lake (ed.), Oxford: Oxford University Press, p 56-75.

Crisp, R. (2006a). Hedonism Reconsidered, *Philosophy and Phenomenological Research*, LXXIII(3): 619-645.

Crisp, R. (2006b). Well-Being, in *The Stanford Encyclopedia of Philosophy (Winter 2006 Edition)*, E.N. Zalta (ed.). Available from: <<http://plato.stanford.edu/archives/win2006/entries/well-being/>>. Accessed on 21/12/2008.

Crisp, R. (2006c). *Reasons & the Good*, New York: Oxford University Press.

Critchley, H.D., Mathias, C.J. & Dolan, R.J. (2001). Neural Activity in the Human Brain Relating to Uncertainty and Arousal during Anticipation, *Neuron*, 29: 537-545.

Critchley, H.D., Wiens, S., Rotshtein, P., Öhman, A. & Dolan, R.J. (2004). Neural Systems Supporting Interoceptive Awareness, *Nature Reviews Neuroscience*, 7: 189-195.

Cushman, F., Young, L. & Hauser, M.D. (2006). The Role of Conscious Reasoning and Intuition in Moral Judgments: Testing Three Principles of Harm, *Psychological Science*, 17(12): 1082-1089.

Daniels, N. (1979). Wide Reflective Equilibrium and Theory Acceptance in Ethics, *Journal of Philosophy*, 76(5): 256-282.

Daniels, N. (1996). *Justice and Justification: Reflective Equilibrium in Theory and Practice*, New York: Cambridge University Press.

Daniels, N. (2003). Reflective Equilibrium, in *The Stanford Encyclopedia of Philosophy (Summer 2003 Edition)*, E.N. Zalta (ed.). Available from: <<http://plato.stanford.edu/archives/sum2003/entries/reflective-equilibrium/>>. Accessed on 21/12/2008.

Darwall, S. (1997). Self Interest and Self Concern, *Social Philosophy and Policy*, 14: 158-178.

- Dasgupta, N., McGhee, D.E., Greenwald, A.G. & Banaji, M.R. (2000). Automatic Preference for White Americans: Eliminating the Familiarity Explanation, *Journal of Experimental Social Psychology*, 36(3): 316-328.
- Davidson, R. (1992). Emotion and Affective Style: Hemispheric Substrates, *Psychological Science*, 3: 39-43.
- Davidson, R. (2000). Affective Style, Psychopathology and Resilience: Brain Mechanisms and Plasticity, *American Psychologist*, 55: 1196-1214.
- Davidson, R., Jackson, D. & Kalin, N. (2000). Emotion, Plasticity, Context and Regulation: Perspectives from Affective Neuroscience, *Psychological Bulletin*, 126: 890-906.
- de Groot, A.D. (1965). *Thought and Choice in Chess*, The Hague: Mouton.
- Denes-Raj, V. & Epstein, S. (1994). Conflict Between Intuitive and Rational Processing: When People Behave Against their Better Judgment, *Journal of Personality and Social Psychology*, 66: 819-829.
- Dennett, D. (1988). Quining Qualia, in *Consciousness in Contemporary Society*, A. Marcel & E. Bisiach (eds.), New York: Oxford University Press, p 42-77.
- Depue, R.A. & Morrone-Stupinsky, J. (2005). A Neurobiological Model of Affiliative Bonding: Implications for Conceptualizing a Human Trait of Affiliation, Plus Commentary, *Behavioral and Brain Sciences*, 28(3): 313-395.
- de Quervain, D.J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A. & Fehr, E. (2004). The Neural Basis of Altruistic Punishment, *Science*, 305: 1254-1258.
- Derbyshire, S.W., Whalley, M.G., Stenger, V.A. & Oakley, D.A. (2004). Cerebral Activation during Hypnotically Induced and Imagined Pain, *Neuroimage*, 23: 392-401.
- Dewey, J. (1922). *Human Nature and Conduct: An Introduction to Social Psychology*, New York: Holt Publishing.
- Diener, E., Seligman, M.E.P. (2004). Beyond Money: Toward an Economy of Well-Being, *Psychological Science in the Public Interest*, 5: 1-31.
- Dijksterhuis, A., Bos, M.W., Nordgren, L.F. & van Baaren, R.B. (2006). On Making the Right Choice: The Deliberation-Without-Attention Effect, *Science*, 311: 1005-1007.
- Dolan, P. & White, M.P. (2007). How Can Measures of Subjective Well-Being be Used to Inform Public Policy?, *Perspectives on Psychological Science*, 2(1): 71-85.
- Donner, W. (1971). *The Liberal Self: John Stuart Mill's Moral and Political Philosophy*, Ithaca: Cornell University Press.
- Edwards, R.B. (1979). *Pleasures and Pains: A Theory of Qualitative Hedonism*, Ithaca: Cornell University Press.
- Ekman, P., Friesen, W.V. & O'Sullivan, M. (1988). Smiles when Lying, *Journal of Personality and Social Psychology*, 54: 414-420.

- Ekman, P. & Rosenberg, E.L. (eds.) (1997). *What the Face Reveals: Basic and Applied Studies of Spontaneous Facial Expressions Using the Facial Action Coding System (FACS)*, New York: Oxford University Press.
- Feldman, F. (1997). *Utilitarianism, Hedonism, and Desert: Essays in Moral Philosophy*, Cambridge: Cambridge University Press.
- Feldman, F. (2004). *Pleasure and the Good Life*, Oxford: Clarendon Press.
- Fischer, J.M. & Ravizza, M. (1991). Responsibility and Inevitability, *Ethics*, 101(2): 258-278.
- Fiske, S.T., & Neuberg, S.L. (1990). A Continuum of Impression Formation, from Category-Based to Individuating Processes: Influences of Information and Motivation on Attention and Interpretation, in *Advances in Experimental Social Psychology*, Vol. 23, M.P. Zanna (ed.), New York: Academic Press, p 1-74.
- Frank, A. & Netherlands State Institute for War Documentation (1989). *The Diary of Anne Frank: The Critical Edition*, Doubleday.
- Frankena, W.K. (1963). *Ethics*, Prentice Hall: New Jersey.
- Frankena, W.K. (1973). *Ethics*, 2<sup>nd</sup> edition, Prentice Hall: New Jersey.
- Furrow, D. (2005). *Ethics*, New York and London: Continuum.
- Gans, H. (1988). *Middle American Individualism: Political Participation and Liberal Democracy*, New York: Free Press.
- Garner, R. (2005). What's in a Name? Persuasion Perhaps, *Journal of Consumer Psychology*, 15(2): 108-116.
- Gert, J. (2004). *Brute Rationality: Normativity and Human Action*, Cambridge: Cambridge University Press.
- Gilbert, D. (2006). *Stumbling on Happiness*, New York: Alfred A. Knopf.
- Gilbert, D.T. (1989). Thinking Lightly about Others, in *Unintended Thought*, J.S. Uleman & J.A. Bargh (eds.), New York: Guilford, p 189-211.
- Gilovich, T., Griffin, D.W. & Kahneman, D. (2002). *Heuristics and Biases: The Psychology of Intuitive Judgment*, Cambridge: Cambridge University Press.
- Goldberg, P. (1983). *The Intuitive Edge: Understanding and Developing Intuition*, Boston: Houghton Mifflin.
- Gosling, J.C.B. (1995). Hedonism, in *The Oxford Companion to Philosophy*, T. Honderich (ed.), Oxford and New York: Oxford University Press.
- Greene, J.D. (in press). The Secret Joke of Kant's Soul, in *Moral Psychology, Vol. 3: The Neuroscience of Morality*, W. Sinnott-Armstrong (ed.), Cambridge, MA: MIT Press.
- Greene, J.D., Haidt, J. (2002). How (and Where) Does Moral Judgment Work?, *Trends in Cognitive Science*, 6: 517-523.

- Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M. & Cohen, J.D. (2004). The Neural Bases of Cognitive Conflict and Control in Moral Judgment, *Neuron*, 44: 389-400.
- Greene, J.D., Sommerville, R., Nystrom, L.E., Darley, J.M., Cohen, J.D. (2001). An fMRI Investigation of Emotional Engagement in Moral Judgment, *Science*, 293: 2105-2108.
- Greenwald, A.G. & Banaji, M.R. (1995). Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes, *Psychological Review*, 102: 4-27.
- Griffin, J. (1986). *Well-Being: Its Meaning, Measurement and Moral Importance*, Oxford: Clarendon Press.
- Griffin, D.W. & Ross, L. (1991). Subjective Construal, Social Inference, and Human Misunderstanding, in *Advances in Experimental Social Psychology*, M. Zanna (ed.), Vol. 24, San Diego: Academic Press, p 319-359.
- Hagafors, R. & Brehmer, B. (1983). Does Having to Justify One's Judgments Change the Nature of the Judgment Process?, *Organizational Behavior and Human Performance*, 31: 223-232.
- Haidt, J. (2001). The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment, *Psychological Review*, 108: 814-834.
- Haidt, J. (2006). *The Happiness Hypothesis: Finding Modern Truth in Ancient Wisdom*, New York: Basic Books.
- Haidt, J., Bjorklund, F., & Murphy, S. (2000). *Moral Dumbfounding: When Intuition Finds No Reason*, Unpublished manuscript, University of Virginia.
- Haidt, J. & Joseph, C. (2007). The Moral Mind: How Five Sets of Innate Intuitions Guide the Development of Many Culture-Specific Virtues, and Perhaps Even Modules, in *The Innate Mind Vol. 3: Foundations and the Future*, P. Carruthers, S. Lawrence & S. Stich (eds.), Oxford University Press.
- Hare, R.M. (1973). 'Rawls' Theory of Justice – I, *Philosophical Quarterly*, 23(90): 144-155 and 'Rawls' Theory of Justice – II, *Philosophical Quarterly*, 23(92): 241-252, reprinted with revisions in *Essays in Ethical Theory*, Oxford University Press, 1989, p 145-174.
- Hauser, M.D., Cushman, F., Young, L., Jin, R.K.-X. & Mikhail, J. (2007). A Dissociation Between Moral Judgments and Justifications, *Mind & Language*, 22(1): 1-21.
- Higgins, E.T., Rholes, W.S. & Jones, C.R. (1977). Category Accessibility and Impression Formation, *Journal of Experimental Social Psychology*, 13: 141-154.
- Hirata, J. (2003). Putting Gross National Happiness in the Service of Good Development, *Journal of Bhutan Studies*, 9: 99-139.
- Hobbes, T. (1650). *Human Nature: Or the Fundamental Elements of Policy*, in *The English Works of Thomas Hobbes*, Vol. 4, W. Molesworth, (ed.), London: John Bohn, 1839.
- Hoorens, V. & Nuttin, J.M. (1993). Overvaluation of Own Attributes: Mere Ownership or Subjective Frequency?, *Social Cognition*, 11: 177-200.
- Huemer, M. (2005). *Ethical Intuitionism*, Palgrave Macmillan.

- Hume, D. (1777). *An Enquiry Concerning the Principles of Morals*, T.L. Beauchamp (ed.), Clarendon Press, 1998.
- Hume, D. (1754). *A Treatise of Human Nature*, D.F. Norton & M.F. Norton (eds.), Oxford University Press, 2000.
- Hurka, T. (1993). *Perfectionism*, New York: Oxford University Press.
- Hursthouse, R. (1999). *On Virtue Ethics*, Oxford University Press.
- Irvine, W.B. (2006). *On Desire: Why We Want What We Want*, Oxford University Press.
- James, W. (1902). *The Varieties of Religious Experience*, New York: Modern Library, 1994.
- Jones, J. T., Pelham, B.W., Mirenberg, M.C. & Hetts, J.J. (2002). Name Letter Preferences Are Not Merely Mere Exposure: Implicit Egotism as Self-Regulation, *Journal of Experimental Social Psychology*, 38(2):170-177.
- Kagan, S. (1992). The Limits of Well-Being, in *The Good Life and the Human Good*, E.F. Paul, F.D. Miller, Jr. & J. Paul (eds.), Cambridge: Cambridge University Press, p 169-189.
- Kagan, S. (1998). *Normative Ethics*, Boulder: Westview Press.
- Kahneman, D. (1999). Objective Happiness, in *Well-Being: The Foundations of Hedonic Psychology*, D. Kahneman, E. Diener & N. Schwarz (eds.), New York: Russell Sage, p 3-25.
- Kahneman, D. (2000). Experienced Utility and Objective Happiness: A Moment-Based Approach, in *Choices, Values, and Frames*, D. Kahneman & A. Tversky (eds.), New York: Cambridge University Press p 673-692).
- Kahneman, D., Diener, E. & Schwarz, N. (eds.) (1999). *Well-Being: The Foundations of Hedonic Psychology*, New York: Russell Sage.
- Kahneman, D., Knetsch, J.L. & Thaler, R.H. (1991). Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias, *The Journal of Economic Perspectives*, 5(1): 193-206.
- Kahneman, D., Krueger, A.B., Schkade, D., Schwarz, N. & Stone, A. (2004). Toward National Well-Being Accounts, *American Economic Review*, 94: 429-434.
- Kahneman, D. & Sugden, R., (2005). Experienced Utility as a Standard of Policy Evaluation, *Environmental and Resource Economics*, 32: 161-181.
- Kahneman, D. & Tversky, A. (2000). *Choices, Values, and Frames*, Cambridge: Cambridge University Press.
- Kahneman, D., Wakker, P.P. & Sarin, R. (1997). Back to Bentham? Explorations of experienced utility, *Quarterly Journal of Economics*, 112: 375-405.
- Kamm, F. (1993). *Morality, Mortality, Volume I: Death and Whom to Save From It*, New York: Oxford University Press.



- Keller, S. (2004). Welfare and the Achievement of Goals, *Philosophical Studies*, 121: 27-41.
- Kitayama, S. & Karasawa, M. (1997). Implicit Self-Esteem in Japan: Name-Letters and Birthday Numbers, *Personality & Social Psychology Bulletin*, 23: 736-742.
- Klein, G. (1998). *Sources of Power*, Cambridge, MA: MIT Press.
- Kolodiejchuk, B. (2007). *Mother Teresa: Come Be My Light*, Doubleday.
- Koyma, T., McHaffie, J.G., Laurienti, P.J. & Coghill, R.C. (2005). The Subjective Experience of Pain: Where Expectations Become Reality, *Proceedings of the National Academy of Sciences USA*, 102: 12950-5.
- Kirngelbach, M.L. (2005). The Human Orbitofrontal Cortex: Linking Reward to Hedonic Experience, *Nature Reviews Neuroscience*, 6: 691-702.
- Krosnick, J.A., Betz, A.L., Jussim, L.J. & Lynn, A.R. (1992). Subliminal Conditioning of Attitudes, *Personality and Social Psychology Bulletin*, 18: 152-162.
- Kymlicka, W. (1990). *Contemporary Political Philosophy: An Introduction*, Oxford: Oxford University Press.
- Laertius, D. (1972). Epicurus, in Book X of *The Lives of the Eminent Philosophers*, trans. by R.D. Hicks, The Loeb Classical Library, London: Heinemann.
- Langer, E.J. (1989). *Mindfulness*, Reading, MA: Addison-Wesley.
- Layard, R. (2005). *Happiness: Lessons from a New Science*, Penguin Books.
- Lewis, C.I. (1946). *An Analysis of Knowledge and Valuation*, La Salle, Ill.: Open Court Publishing Company.
- Lewis, D. (1989). Dispositional Theories of Value – II, *The Aristotelian Society Supplementary*, LXIII: 113-137.
- Lewicki, P. (1986). *Nonconscious Social Information Processing*, New York: Academic Press.
- Liao, S.M. (2007). A Defense of Intuitions, *Philosophical Studies*, published online on 8 August 2007. Available from <<http://www.springerlink.com/content/3507675837420008/>>. Accessed on 21/12/2008.
- Lieberman, M. (2000). Intuition: A Social Cognitive Neuroscience Approach, *Psychological Bulletin*, 126: 109-137.
- Lieberman, M.D., & Eisenberger, N.I. (2006). A Pain by Any Other Name (Rejection, Exclusion, Ostracism) Still Hurts the Same: The Role of Dorsal Anterior Cingulate Cortex in Social and Physical Pain, in *Social Neuroscience: People Thinking about Thinking People*, J.T. Cacioppo, P.S. Visser, & C.L. Pickett (eds.), Cambridge, MA: MIT Press, p 167-187).
- Lillie, W. (1955). *An Introduction to Ethics*, 3<sup>rd</sup> edition, Methuen, 1957.
- Locke, J. (1689). *An Essay Concerning Human Understanding*, Tegg, 1848.

Loewenstein, G. & Schkade, D. (1999). Wouldn't it be Nice? Predicting Future Feelings, in *Well-Being: The Foundations of Hedonic Psychology*, D. Kahneman, E. Diener & N. Schwarz (eds.), New York: Russell Sage, p 85-105.

Lunn, J.H. (1948). Chick Sexing, *American Scientist*, 36: 280-287.

MacFarquhar, L. (2007). Two Heads, from the Annals of Science, in *The New Yorker*, February 12, 2007, p 58.

Mackie, J.L. (1977). *Ethics: Inventing Right and Wrong*. Middlesex: Penguin books.

McGuire, W.J. & McGuire, C.V. (1981). The Spontaneous Self-Concept as Affected by Personal Distinctiveness, in *Self-Concept: Advances in Theory and Research*, L. Lynch, D. Mervin, A. Norem-Hebeisen, & K. Gergen (eds.), New York: Ballinger p 147-171.

Mikhail J. (2007). Universal Moral Grammar: Theory, Evidence and the Future, *Trends in Cognitive Sciences*, 11(4), 143-152.

Mill, J.S. (1861). *Utilitarianism*, Indianapolis: Bobbs-Merrill, 1957.

Mill, J.S. (1873). *Autobiography*, J. Stiller (ed.), Boston: Houghton Mifflin Company, 1969.

Monro, D.H. (1950). In Defense of Hedonism, *Ethics*, 60(4): 285-291.

Montague, P.R., Berns, G.S., Cohen, J.D., McClure, S.M., Pagnoni, G., Dhamala, M., Wiest, M.C., Karpov, I., King, R.D., Apple, N. & Fisher, R.E. (2002). Hyperscanning: Simultaneous fMRI during Linked Social Interactions, *NeuroImage*, 16(4): 1159-1164.

Moore, A. (2004). Hedonism, in *The Stanford Encyclopedia of Philosophy (Summer 2004 Edition)*, E.N. Zalta (ed.), Available from: <<http://plato.stanford.edu/archives/sum2004/entries/hedonism/>>. Accessed on 21/12/2008.

Moore, A. & Crisp, R. (1996). Welfarism in Moral Theory, *Australasian Journal of Philosophy*, 74: 598-613.

Moore, G.E. (1903). *Principia Ethica*, Cambridge University Press, 1969.

Moore, G.E. (1912). *Ethics*, New York: Oxford University Press.

Murphy, S.T. & Zajonc, R.B. (1993). Affect, Cognition, and Awareness: Affective Priming with Optimal and Suboptimal Stimulus Exposures, *Journal of Personality and Social Psychology*, 64: 723-729.

Myers, D.G. (1992). *The Pursuit of Happiness*, New York: Avon Books.

Myers, D.G. (2004). *Intuition: Its Powers and Perils*, Yale University Press.

Nagel, T. (1979). *Mortal Questions*, Oxford University Press.

Nagel, T. (1986). *The View from Nowhere*, Oxford University Press.

Nolen-Hoeksema, S. & Davis, C.G. (2002). Positive Responses to Loss, in *Handbook of Positive Psychology*, C.R. Snyder & S.J. Lopez (eds.), New York: Oxford University Press, p598-607.

- Nozick, R. (1974). *Anarchy, State, and Utopia*, Oxford: Blackwell, 1991.
- Nussbaum, M. & Sen, A. (eds.) (1993). *The Quality of Life*, Oxford: Oxford University Press.
- Nussbaum, M. (2001). The Protagoras: A Science of Practical Reasoning, in *Varieties of Practical Reasoning*, E. Millgram (ed.), MIT Press, p 153-201.
- Nuttin, J.M. (1987). Affective Consequences of Mere Ownership: The Nameletter Effect in Twelve European Languages, *European Journal of Social Psychology*, 17: 381-402.
- O'Neill, P. & Petrinovich, L. (1998). A Preliminary Cross Cultural Study of Moral Intuitions, *Evolution and Human Behavior*, 19: 349-367.
- Oswald, A.J. (1997). Happiness and Economic Performance, *The Economic Journal*, 107: 1815-1831.
- Parfit, D. (1984). *Reasons and Persons*, Oxford University Press, revised paperback edition 1986.
- Pavot, W. & Diener, E. (1993). Review of the Satisfaction with Life Scale, *Psychological Assessment*, 5: 164-172.
- Pecina S., Smith, K.S. & Berridge K.C. (2006). Hedonic Hot Spots in the brain, *Neuroscientist*, 12: 500-511.
- Pelham, B.W., Jones, J.T., Mirenberg, M.C. & Carvallo, M. (2001). *Implicit Egotism: Implications for Interpersonal Attraction*, manuscript submitted for publication.
- Pelham, B.W., Mirenberg M.C., Jones, J.K. (2002). Why Susie Sells Seashells by the Seashore: Implicit Egotism and Major Life Decisions, *Journal of Personality and Social Psychology*, 82:469-487.
- Pence, G. (2005). Hedonism, in *A Dictionary of Common Philosophical Terms*, International Edition, G. Pence (ed.), McGraw-Hill.
- Petrinovich, L. & O'Neill, P. (1996). Influence of Wording and Framing Effects on Moral Intuitions, *Ethology and Sociobiology*, 17(3): 145-171.
- Petrinovich, L., O'Neill, P. & Jorgensen, M.J. (1993). An Empirical Study of Moral Intuitions: Towards an Evolutionary Ethics. *Ethology and Sociobiology*, 64: 467- 478 .
- Pigou, A.C. (1960). *The Economics of Welfare*, 4<sup>th</sup> edition, London: Macmillan.
- Plato (1937). Philebus, in *The Dialogues of Plato*, trans. by B. Jowett, New York: Random House.
- Plato (1986). Protagoras, in *The Dialogues of Plato*, trans. by B.A.F. Hubbard & E.S. Karmofsky, New York: Bantam Books.
- Rachels, J. (2005). *The Elements of Moral Philosophy*, International Edition, McGraw-Hill.
- Rachels, S. (2003). A Defense of Two Optimistic Claims in Ethical Theory, *Philosophical Studies*, 112: 1-30.
- Railton, P. (1986). Facts and Values, *Philosophical Topics*, 14: 5-31.

- Railton, P. (1989). Naturalism and Prescriptivity, *Social Philosophy and Policy*, 7: 151-174.
- Rawls, J. (1971). *A Theory of Justice*, The Belknap Press of Harvard University Press.
- Rawls, J. (1974/5). The Independence of Moral Theory, *Proceedings and Addresses of the American Philosophical Association*, XLVII: 5-22.
- Raz, J. (2004). The Role of Well-Being, *Ethics*, 18: 269-294.
- Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G. & Kilts, C. (2002). A Neural Basis for Social Cooperation, *Neuron*, 35: 395-405.
- Robinson, T.E. & Berridge, K.C. (2000). The Psychology and Neurobiology of Addiction: An Incentive-Sensitization View, *Addiction*, 95, supplement 2: S91-S117.
- Robinson, T.E. & Berridge, K.C. (2001). Incentive-Sensitization View and Addiction, *Addiction*, 96: 103-114.
- Rosen, B. (1993). *Ethical Theory: Strategies and Concepts*, Mayfield Publishing Company.
- Ross, W.D. (1930). *The Right and the Good*, Oxford University Press.
- Rutherford, D. (2002). *Routledge Dictionary of Economics*, 2<sup>nd</sup> edition, London: Routledge.
- Ryle, G. (1954). *Dilemmas*, Cambridge: Cambridge University Press.
- Saffran, J.R., Aslin, R.N. & Newport, E.L. (1996). Statistical Learning by 8-Month-Old Infants, *Science*, 274: 1926-1928.
- Sanborn, J. (2003). The Short Course for Murder: How Soldiers and Criminals Learn to Kill, in *Violent Acts and Violentization: Assessing, Applying, and Developing Lonnie Athens' Theories*, L. Athens & J.T. Ulmer (eds.), *Sociology of Crime, Law and Deviance*, Vol. 4, Cambridge, UK: Elsevier Sciences, p 107-124.
- Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E. & Cohen, J.D. (2003). The Neural Basis of Economic Decision-Making in the Ultimatum Game, *Science*, 300: 1755-1758.
- Sasson, T. (1995). *Crime Talk: How Citizens Construct a Social Problem*, Hawthorne, NY: Aldine.
- Savery, W. (1934). A Defense of Hedonism, *International Journal of Ethics*, 45(1): 1-26.
- Seligman, M.E.P. & Csikszentmihalyi, M. (2000). Positive Psychology: An Introduction, *American Psychologist*, 55: 5-14.
- Shin, L.M., Dougherty, D.D., Orr, S.P., Pitman, R.K., Lasko, M., Macklin, M.L., Alpert, N.M., Fischman, A.J. & Rauch, S.L. (2000). Activation of Anterior Paralimbic Structures during Guilt-Related Script-Driven Imagery, *Society of Biological Psychology*, 48: 43-50.
- Sidgwick, H. (1907). *The Methods of Ethics*, 7<sup>th</sup> edition, London: Macmillan.

- Silverstein, M. (2000). In Defense of Happiness: A Response to the Experience Machine, *Social Theory and Practice*, 26(2): 279-300.
- Singer, P. (1974). Sidgwick and Reflective Equilibrium, *The Monist*, 58, reprinted in *Unsanctifying Human Life*, H. Kulse (ed.), Oxford: Blackwell Publishing, 2002, p 27-50. Page references are to reprinted version.
- Singer, T., Kiebel, S.J., Winston, J.S., Dolan, R.J. & Frith, C.D. (2004a). Brain Responses to the Acquired Moral Status of Faces, *Neuron*, 41: 653-662.
- Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R.J. & Frith, C.D. (2004b). Empathy for Pain Involves the Affective but not Sensory Components of Pain, *Science*, 303: 1157-1162.
- Sinnott-Armstrong, W. (2006). *Moral Skepticisms*, New York: Oxford University Press.
- Smith, E.R., & DeCoster, J. (1999). Associative and Rule-Based Processing: A Connectionist Interpretation of Dual-Process Models, in *Dual-Process Theories in Social Psychology*, S. Chaiken & Y. Trope (eds.), New York: Guilford, p 323-336.
- Sobel, D. (1994). Full Information Accounts of Well-Being, *Ethics*, 104(4): 784-810.
- Sobel, D. (2002). Varieties of Hedonism, *Journal of Social Philosophy*, 33(2): 240-256.
- Sober, E. & Wilson, D.S. (1998). *Unto Others: The Evolution and Psychology of Unselfish Behaviour*, Cambridge, MA: Harvard University Press.
- Sosa, E. (2006). Intuitions and Truth, in *Truth and Realism*, P. Greenough & M.P. Lynch (eds.), Oxford: Clarendon Press, p 208-226.
- Sprigge, T.L.S. (1988). *The Rational Foundation of Ethics*, London: Routledge and Kegan Paul.
- Steele, C.M. (1988). The Psychology of Self-Affirmation: Sustaining the Integrity of the Self, in *Advances in Experimental Social Psychology*, Vol. 21, L. Berkowitz (ed.), New York: Academic Press, p. 261-302).
- Sumner, L.W. (1996). *Welfare, Happiness, & Ethics*, Oxford: Clarendon press.
- Thomson, J.J. (1971). A Defense of Abortion, *Philosophy & Public Affairs*, 1: 47-66.
- Timmons, M. (2006). *Conduct and Character: Readings in Moral Theory*, 5<sup>th</sup> edition, Thomson Wadsworth.
- Tindell, A.J., Smith, K., Pecina, S., Berridge, K. & Aldridge, W.J. (2006). Ventral Pallidum Firing Codes Hedonic Reward: When a Bad Taste Turns Good, *Journal of Neurophysiology*, 96: 2399-2409.
- Tooley, M. (1972). Abortion and Infanticide, *Philosophy & Public Affairs*, 2: 37-65.
- Turton, D. (2007). Fred Feldman's Pleasure and the Good Life: Concerning the Nature, Varieties, and Plausibility of Hedonism, in Book Reviews, *Australian Journal of Philosophy*, 85(2): 338-340.
- Tversky, A. & Kahneman, D. (1991). Loss Aversion in Riskless Choice: A Reference-Dependent Model, *Quarterly Journal of Economics*, 106(4): 1039-61.

- Unger, P. (1996). *Living High and Letting Die*, New York: Oxford University Press.
- Urry, H.L., Nitschke, J.B., Dolski, I., Jackson, D.C., Dalton, K.M., Mueller, C.J., Rosenkranz, M.A., Ryff, C.D., Singer, B.H. & Davidson, R.J. (2004). Making a Life Worth Living: Neural Correlates of Well-Being, *Psychological Science*, 15(6): 367-372.
- Valdesolo, P. & DeSteno, D. (2006) Manipulations of Emotional Context Shape Moral Judgment, *Psychological Science*, 17(6): 476-477.
- Valentine, C. (1929). The Relative Reliability of Men and Women in Intuitive Judgments of Character, *British Journal of Psychology*, 19: 213-238.
- van Lange, P.A.M., Tarris, T.W. & Vonk, R. (1997). Dilemmas of Academic Practice: Perceptions of Superiority among Social Psychologists, *European Journal of Social Psychology*, 27: 675-685.
- van Praag, B.M.S. & Baarsma, B.E. (2005). Using Happiness Surveys to Value Intangibles: The Case of Airport Noise, *Economic Journal*, 115(500): 224-246.
- Veenhoven, R. (2004). Happiness as a Public Policy Aim: The Greatest Happiness Principle, in *Positive Psychology in Practice*, P.A. Lindley & S. Joseph (eds.), New York: John Wiley & Sons.
- Vesey, G. & Foulkes, P. (1990). Hedonism, in *Collins Dictionary of Philosophy*, G. Vesey & P. Foulkes (eds.), London and Glasgow: Collins.
- Watkins, J.W.N. (1957). Between Analytic and Empirical, *Philosophy*, 32(121): 112-131.
- Wheeler, R.A. & Carelli, R.M. (2006). The Neuroscience of Pleasure: Focus on “Ventral Pallidum Firing Codes Hedonic Reward: When a Bad Taste Turns Good, *Journal of Neurophysiology*, 96: 2175-2176.
- Williams, B. (1973). *Utilitarianism: For and Against*, Cambridge: Cambridge University Press.
- Woodward, J. & Allman, J. (2007). Moral Intuition: Its Neural Substrates and Normative Significance, *Journal of Physiology-Paris*, 101(4-6): 179-202.
- Word, C.O., Zanna, M.P. & Cooper, J. (1974). The Nonverbal Mediation of Self-fulfilling Prophecies in Interracial Interaction, *Journal of Experimental Social Psychology*, 10: 109-120.
- Zebrowitz, L.A. & Collins, M.A. (1997). Accurate Social Perception at Zero Acquaintance: The Affordances of a Gibsonian Approach, *Personality & Social Psychology Review*, 1: 204-223.