

ITEM NO: 1930788



1930788

**FOR
REFERENCE ONLY**



**HAND BOUND
BY
F.C. BOOKBINDER
WWW.FCBOOKBINDER.CO.UK
TEL. 01792 581384**

Redefinitions of Death

Gareth Southwell

A dissertation submitted to the University of Wales, Newport
in candidature for the degree of Doctor in Philosophy

School of Health and Social Sciences

2008

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed *G. J. Don*
Date..... *20/10/2008*

STATEMENT 1

This dissertation is being submitted in partial fulfilment of the requirements for the degree of
..... *PhD* (insert MA, MBA, PhD) as appropriate

Signed *G. J. Don*
Date *20/10/2008*

STATEMENT 2

This dissertation is the result of my own independent work/investigation, except where otherwise stated.

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed *G. J. Don*
Dated *20/10/2008*

STATEMENT 3

I hereby give consent for my dissertation, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed *G. J. Don*
Dated *20/10/2008*

Summary

Redefinitions of Death

Gareth Southwell

In this dissertation I consider the current controversy surrounding the definition of death in certain rare situations which arise in response to advances in medical technology. In relation to these, I criticise various approaches which seek a resolution to such problems, arguing that all of them involve unjustified and unexamined assumptions as to the nature of death, and other related concepts. In chapter 1, I introduce and define the nature of the problem, arguing that whilst the problem can be seen to spring from technological advances, these merely reveal an inherent ambiguity. In chapter 2, I examine the so-called 'strictly-biological approach', and argue that its premise (that we can treat death as a purely factual matter) is flawed. In chapter 3, I support this conclusion with a broader attack upon conceptual essentialism, of which the strictly-biological approach can be seen as a foremost example. I also argue that this means that the nature of the problem is one that is not amenable to the sort of conceptual analysis that many might use to resolve the problem. Chapter 4 looks at the idea that biological function – the central criterion of the strictly-biological approach – cannot be considered an intrinsic, mind-independent feature of the world (and therefore, neither can a strictly-biological definition of death). Chapters 5 and 6 look at non-strictly-biological attempts to define death (what I term 'partly-biological' views) – namely the capacity for consciousness and personal identity respectively – and argue that both these approaches, far from resolving the problem, merely shift it to a different ground. In chapter 7, I present a different picture of death as an 'observer-relative' feature of the world (to use John Searle's terminology), and argue that the resolution of the problem must have more in common with practices (e.g. in sport) where similar ambiguities are occasionally faced. In chapter 8, I further explore the consequences of the observer-relative status of death, arguing that this means that a much wider degree of variance and mutability is possible in relation to the related concepts of 'self' and 'death', and that certain religious viewpoints and scenarios in science-fiction literature embody just such a 'dialogue with death' and present us with what I term notions of 'the expanded self'. Finally, I briefly outline some of the consequences of my arguments for medicine and public policy decisions, and suggest certain avenues for future research, arguing that rather than seeking to arrive at a single, unified definition of death, we should instead search for ways of coping with multiple parallel 'redefinitions' of death.

Summary

Redefinitions of Death

In this dissertation I consider the current controversy surrounding the definition of death in certain rare situations which arise in response to advances in medical technology. In relation to these, I criticise various approaches which seek a resolution to such problems, arguing that all of them involve unjustified and unexamined assumptions as to the nature of death, and other related concepts. In chapter 1, I introduce and define the nature of the problem, arguing that whilst the problem can be seen to spring from technological advances, these merely reveal an inherent ambiguity. In chapter 2, I examine the so-called ‘strictly-biological approach’, and argue that its premise (that we can treat death as a purely factual matter) is flawed. In chapter 3, I support this conclusion with a broader attack upon conceptual essentialism, of which the strictly-biological approach can be seen as a foremost example. I also argue that this means that the nature of the problem is one that is not amenable to the sort of conceptual analysis that many might use to resolve the problem. Chapter 4 looks at the idea that biological function – the central criterion of the strictly-biological approach – cannot be considered an intrinsic, mind-independent feature of the world (and therefore, neither can a strictly-biological definition of death). Chapters 5 and 6 look at non-strictly-biological attempts to define death (what I term ‘partly-biological’ views) – namely the capacity for consciousness and personal identity respectively – and argue that both these approaches, far from resolving the problem, merely shift it to a different ground. In chapter 7, I present a different picture of death as an ‘observer-relative’ feature of the world (to use John Searle’s terminology), and argue that the resolution of the problem must have more in common with practices (e.g. in sport) where similar ambiguities are occasionally faced. In chapter 8, I further explore the consequences of the observer-relative status of death, arguing that this means that a much wider degree of variance and mutability is possible in relation to the related concepts of ‘self’ and ‘death’, and that certain religious viewpoints and scenarios in science-fiction literature embody just such a ‘dialogue with death’ and present us with what I term notions of ‘the expanded self’. Finally, I briefly outline some of the consequences of my arguments for medicine and public policy decisions, and suggest certain avenues for future research, arguing that rather than seeking to arrive at a single, unified definition of death, we should instead search for ways of coping with multiple parallel ‘redefinitions’ of death.

Contents

<i>ACKNOWLEDGEMENTS</i>	V
1. INTRODUCTION	1
WHY 'MOSTLY DEAD' WON'T DO.....	1
SOME PROVISOS	2
A HYPOTHETICAL CASE.....	4
THE NATURE OF THE PROBLEM: THE SEARCH FOR SIGNIFICANT FEATURES.....	7
COMPETING CRITERIA.....	10
a) <i>Spontaneous function</i>	10
b) <i>Functional integration</i>	10
c) <i>Responsivity</i>	11
d) <i>Consciousness</i>	12
A MESSY DIVORCE: ORIGINS OF CONCEPTUAL AMBIVALENCE	14
OUTLINE OF THE ARGUMENT	16
2. A BIOLOGICAL DEFINITION OF DEATH	19
INTRODUCTION.....	19
THE STRICTLY-BIOLOGICAL VIEW.....	20
DEATH AS A BIOLOGICAL FACT.....	21
THE CARDIO-PULMONARY METHOD: JONAS AND EVANS AND PRE-RATIONAL ATTITUDES	22
ORGANISMIC DEATH: LAURENCE C. BECKER AND THE 'BEING/HAS BEEN' BOUNDARY	25
BRAIN-STEM DEATH: PALLIS AND LAMB AND FUNCTIONAL INTEGRATION	30
TOM RUSSELL AND HOMEOSTASIS	38
SUMMARY OF PROBLEMS WITH THE STRICTLY-BIOLOGICAL DEFINITION	41
CONCLUSION.....	42
3. DEATH AND THE PROBLEM OF CATEGORIZATION	44
INTRODUCTION.....	44
FALSE BEGINNINGS	44
INTUITIVE, NATURAL, AND THEORETICAL KINDS	47
CONFLICT AND RESOLUTION IN THE DEVELOPMENT OF CATEGORIES	49
CONCEPTUAL REALISM AND CONCEPTUALISM	52
DEATH AND THE EXCLUDED MIDDLE.....	58
ESSENTIALISM, SIGNIFICANT FEATURES, AND FAMILY RESEMBLANCES	59
ONTOLOGICAL RELATIVITY AND OPACITY OF REFERENCE	66
THE REDUNDANCY OF CATEGORIES AND THE DEATH OF NATURAL KINDS.....	70
CONCLUSION.....	74
4. THE STATUS OF DEATH	76
INTRODUCTION.....	76
SOME USEFUL DISTINCTIONS	77
DEATH AS AN INTRINSIC FEATURE OF THE WORLD	79
THREE ARGUMENTS FOR INTRINSIC FUNCTION	80
MILLIKAN AND PROPER FUNCTIONS.....	82
FUNCTION AND TECHNOLOGY	86
CONCLUSION	87
5. DEATH AND CONSCIOUSNESS	89
INTRODUCTION.....	89
CAPACITY FOR CONSCIOUSNESS	89
CONTINGENT AND OBSERVER-RELATIVE FEATURES.....	90
THE STATUS OF CONSCIOUSNESS	92
LIMITS OF THE ENQUIRY	93
THE NATURE OF CONSCIOUSNESS	94
EPISTEMIC OBJECTIVITY	98
CESSATION OF CONSCIOUSNESS AS A CRITERION FOR DETERMINING DEATH	100

WHY CHOOSE CONSCIOUSNESS?.....	103
CONCLUSION.....	104
6. DEATH AND PERSONAL IDENTITY	106
INTRODUCTION.....	106
GREEN AND WIKLER’S ONTOLOGICAL ARGUMENT.....	106
OBJECTIONS TO GREEN AND WIKLER.....	107
KAREN GERVAIS AND <i>REDEFINING DEATH</i>	110
RESPONSES TO GERVAIS.....	113
THE TWOFOLD DEFINITION OF DEATH.....	114
THE GENERAL PROBLEM WITH PERSONAL IDENTITY ARGUMENTS.....	117
GERVAIS ON PERSONAL IDENTITY.....	118
CONCLUSION.....	121
7. AN OBSERVER-RELATIVE DEATH	123
INTRODUCTION.....	123
A NOTE ON PHYSICALISM.....	123
AN OBSERVER-RELATIVE DEFINITION OF DEATH.....	125
DEATH, FACT, AND VALUE – PROBLEMS WITH SEARLE’S VIEW?.....	126
DEATH AS AN EPISTEMICALLY-THICK CONCEPT: THE PARTLY-BIOLOGICAL VIEW.....	128
SOCIAL REALITY: CONSTITUTIVE AND REGULATIVE RULES, AND CONVENTIONS.....	130
THE GAME OF DEATH.....	131
A SPORTING ANALOGY: FOOTBALL AND DEATH.....	132
CONCLUSION.....	134
8. DEATH AND IMMORTALITY	136
INTRODUCTION.....	136
DEATH AND CHOICE: MORALITY.....	136
WITTGENSTEIN’S RIVER.....	138
DEATH AS THE DEPARTURE OF THE SOUL.....	140
THE DECONSTRUCTION OF THE SOUL.....	142
THE EXPANDED SELF.....	145
DEATH AND IMMORTALITY.....	148
WITTGENSTEIN’S CONSERVATISM.....	151
DEATH AND THE FUTURE: SCIENCE-FICTION.....	154
THE IMMORTAL BODY.....	155
PROBLEMS AND SOLUTIONS.....	156
TECHNO-UTOPIA AND GROUP MINDS.....	162
CONCLUSION.....	165
9. CONCLUSION: REDEFINITIONS OF DEATH	167
INTRODUCTION.....	167
SUMMARY OF ARGUMENTS.....	167
AN INFORMED DEATH.....	170
BIBLIOGRAPHY	174

Acknowledgements

Firstly, I must thank Dr Gideon Calder, my long-suffering supervisor, for his encouragement, support, and general willingness to go the extra mile. Without his keen-witted critique and refusal to be hoodwinked by my blather I would not now be writing this. His friendship and advice have been – and will no doubt continue to be (unless he moves house and changes his name) – invaluable.

Thanks must also go to everyone in the Department of Health and Social Sciences at the University of Newport. Thanks to Anne, Dee, Nuala, and everyone in the office, who were ever patient and helpful in the face of my total ignorance of University procedures. I benefited greatly from the conversation and advice of staff and fellow students, but thanks especially to Steve Smith and Sheila Spong, for their conversation and ideas over the years, and for giving me an exemplary mock viva (so exemplary, I didn't sleep for the next two weeks!). Thanks also to Enzo Rossi for his feedback, to Carl Sykes for organising everything and fielding my hapless queries, and to Bronwen Stone and everyone in the library, who were constantly informative and helpful. Lastly, thanks to Phil Cole and Geoffrey Scarre for agreeing to be my internal and external examiners respectively, and for taking the time to wade through 200 pages of death! Despite being almost blind with panic and anxiety, I did recognise the efforts they both went to in order to put me at my ease, and for that, and their thoughtful and challenging response to my work (which has provided stimulus for future research), I am also extremely grateful.

Friends and family have all been very supportive of me over the time of my studies. PhD research is a long and lonely road, and it can turn anyone into an antisocial grump. For five years I have been, constantly or by turns, preoccupied, short-tempered, lazy, manic, and despondent. Many have helped me though this: Mam and Mike, Dad and Doreen, and everyone in the extended network of families (including Gwen, John, Ness and the rest of the Outlaws), have always been there for me (for *us*), and I think my Gran, my mother and I all appreciated the neat synchronicity which led me to undertake my viva at the same campus where my Gramps and her brothers learnt to teach. It is perhaps inevitable that writing a thesis on death would eventually lead me to thoughts of my own dead, but it is especially tempting to think that, in falling over the finish line I have carried their dreams a little further, and that they would be happy for me – or, dare I say it, *are* happy for me? For Gramps, then, and for Gran and Grandad Cymmer, and for all “precious friends hid in death's dateless night” (Will's always there when you need him!): this is for you.

Finally, it is with immense gratitude and love that I dedicate this thesis to Jo, Eliot, and Tesni, who had to put up with me for the duration, and sacrificed many of their creature comforts so that I could follow my impractical and financially unprofitable interests. It is only fair that I now repay them by getting a high-powered job in the city. Or something.

1. Introduction

Inigo Montoya: We need a miracle. It's very important.

Miracle Max: Look, I'm retired. Besides, why would you want someone the King's stinking son fired? I might kill whoever you want to make the miracle.

Inigo Montoya: He's already dead.

Miracle Max: [Intrigued] He is, eh? I'll have a look. Bring him in. [They enter. *Miracle Max* examines *Westley*.] I've seen worse.

Inigo Montoya: Sir... Sir.

Miracle Max: Huh?

Inigo Montoya: We're in a terrible rush.

Miracle Max: Don't rush me sonny. You rush a miracle man, you get rotten miracles. You got money?

Inigo Montoya: Sixty-five.

Miracle Max: Sheesh! I never worked for so little... except once and that was a very noble cause.

Inigo Montoya: This is noble sir. His wife is... crippled... children on the brink of starvation...

Miracle Max: Are you a rotten liar.

Inigo Montoya: I need him to help avenge my father, murdered these twenty years.

Miracle Max: Your first story was better. He probably owes you money, huh? Where's the bellows cramp? I'll ask him.

Inigo Montoya: He's dead. He can't talk.

Miracle Max: Oooooohhh! Look who knows so much, eh! It just so happens that your friend here is only *mostly* dead. There's a big difference between mostly dead and all dead. Please open his mouth. [He inserts the bellows] Now, mostly dead is slightly alive. Now, all dead... well, with all dead, there's usually only one thing that you can do.

Inigo Montoya: What's that?

Miracle Max: Go through his clothes and look for loose change.¹

Why 'Mostly Dead' Won't Do

The above excerpt from the film *The Princess Bride* illustrates nicely the problems we are faced with in determining death. In fact, those things which fuel the comedy of the situation are the very concerns that it is my present purpose to investigate. Why

¹ Goldman (1995), 'The Princess Bride', scene 13.

should it seem comically absurd to talk of someone being ‘mostly dead’, or indeed of there being such gradations of death? It is common to think as if life and death were two sides of the same coin; two mutually exclusive yet directly related opposites. To admit of gradation is therefore to court either the comic or the absurd, and not the actual or the possible.

One of my central purposes in this thesis is therefore to ask what status such assumptions have: do they have a logical or empirical force? What might happen if we were to suppose that the life-death dichotomy need not hold? Therefore, my main concern will be to seek to answer the question of what *sort* of thing death is. I believe that certain controversies that have arisen comparatively recently in the field of medical science (i.e. in the last half-century or so) have provided a key opportunity to answer this fundamental question. Heretofore, I contend, important insights into the way in which we define death have been hidden, obscured by our technological inability to maintain and revive those individuals who had suffered certain physical crises. However, with the advent of such technology – and their attendant ethical and definitional dilemmas – the *nature* of our concepts of life and death has been a little revealed. It is in this gap, I argue, through which we have been allowed to peep into the ‘inner workings’ of the life-death dichotomy itself, that we can begin to gain a deeper understanding of the debate. As Brante and Hallberg point out:

During a controversy, the scientific community becomes less opaque; different styles of reasoning are uncovered, and assumptions that during ‘normal scientific’ periods are regarded as self-evident become visible, as they are both attacked and undermined, and have thus to be consciously defended. This process sheds new light on the possible relativity of presuppositions and truth claims in science.²

In this sense, therefore, whilst I will have a view on whether current attempts to resolve the problem are successful, my primary aim is investigative, and concerns the nature of death itself. In this light, I may say in advance that I find many of the current theories to have missed what I take to be the true character of the problem, and thereby to have misunderstood the nature of the controversy. In other words, most commentators have been at pains to patch up or else explain away what they see as a disquieting aberration: life and death should be separate, distinct, and immutable concepts. What’s more, even where a greater amount of mutability has been admitted, the full consequences of this have not been followed up.³

My ultimate goal will therefore be to take what I see as the necessary conclusions to be drawn from this situation, and to extend them to their ultimate logical possibility. As such, the later chapters of this thesis will explore what might be thought of as the possibilities of life and death. However, it first needs to be shown why there is a problem at all.

Some Provisos

² Brante and Hallberg (1991), p.390.

³ For instance, Gervais (1986) admits that death cannot be defined by reference to the mere facts of biology, but goes on to argue for a variation of the traditional view of death should be adopted (I shall return to this later).

Having spoken quite generally so far of definitions, debates, and problems, I would like now to provide a clear context within which to explain these references. Firstly, however, it should be noted that, in saying that there is a controversy surrounding the definition of death, I am in fact referring to a very specific set of circumstances which take place in the emergency and intensive care units of our hospitals. In these environments, we are presented with borderline cases by the mere fact that such individuals are on the brink of death, and we are trying to save them. However, whilst I am not therefore questioning the determination of death in all cases, it should be understood that the conclusions reached in relation to these examples will – I argue – have a bearing on the notion of death as a whole.

Secondly, in speaking of ‘death’, in most cases I mean ‘human death’. For reasons which will become apparent, it is debateable whether there can be one uniform definition of death which covers all living creatures. Therefore, whilst I will have something to say on this subject, my main concern will be the death of human beings.

Thirdly, a controversy surrounding the *definition* of death should not be confused with difficulties concerning the *determination* of death: problems of definition relate to the difficulties in identifying the necessary and sufficient features which constitute ‘death’ (the concept); problems of determination, however, involve difficulties in applying criteria based on features already identified. The former is primarily a philosophical concern; the latter primarily a technical one. For instance, it may be argued that, even if death is defined as the permanent absence of brain-stem function, there are currently (some argue) substantial difficulties involved in applying this as a general rule in all cases.⁴ Therefore, whilst I will occasionally have cause to consider such technical problems, I will be mostly concentrating on the controversies surrounding definition.

In relation to this last point, it can be seen that any notion of the ‘indeterminacy of death’ must be carefully formulated. For instance, in Edgar Allan Poe’s story, ‘The Premature Burial’, the author states:

The boundaries which divide Life from Death are at best shadowy and vague. Who shall say where the one ends, and where the other begins? We know that there are diseases in which occur total cessations of all the apparent functions of vitality, and yet in which these cessations are merely suspensions, properly so called. They are only temporary pauses in the incomprehensible mechanism. A certain period elapses, and some unseen mysterious principle again sets in motion the magic pinions and the wizard wheels.⁵

Here, the line dividing life and death is seen as indistinct and – so it is implied – mysterious and unknowable. However, of what does such indeterminacy consist? Is it a diagnostic problem, or a definitional one? Obviously, whilst the two are linked, it is important to try to distinguish them. Accordingly, where, in the following chapters, I suggest that there is a problem with a certain definition of death, it should not be assumed that I am saying that ‘we will never be able to diagnose death’, or that ‘we

⁴ See, for example, Russell (2000), p.28.

⁵ Poe (1986), ‘The Premature Burial’, p.348.

will never know what death is'. I shall argue, ultimately, that the reason why the definition of death is problematic is more because of its status as a concept than the presence of any 'mysterious principle'. Therefore, what I argue for is the potential *variability* of the concept of death due to the nature of the concept. Furthermore, that there should be any deep mystery which precludes us from *ever* diagnosing death with any certainty, is a sceptical position to do with the possibilities of empirical truth, and – the common type of diagnostic problems aside – this is not a question that I will be primarily concerned with (for, in light of my overall argument, such sceptical problems may take on a different significance).

A Hypothetical Case

In light of the above, I will now illustrate the problem with a hypothetical example. However, before doing so, let us now consider some of the ways in which advancements in medical technology have given rise to definitional controversy:

- Supplementation and Support of Spontaneous Function: In cases of respiratory failure – which may involve simply a difficulty in breathing (such as arises from a lung infection) or (in the extreme case) cessation of the function of the brain-stem – it is possible to intubate (i.e. place a tube into the respiratory tract) and to ventilate the lungs. This allows for the maintenance of breathing and the artificial oxygenation of the blood, which in turn allows the heart to continue beating and the prevention of brain damage. In cases where there is no brain damage and the condition temporary, this is generally sufficient to maintain the patient in a stable condition until possible recovery. Similarly, kidney dialysis can supplement the performance of a damaged kidney, and pacemakers can correct and stabilise an irregular heartbeat. These are only a few of the ways in which medical technology is continually causing us to re-evaluate which features constitute death.⁶
- Organ Transplant: The now almost-routine instance of heart-transplant operations (as well as transplantation of other organs) means that, in some cases, survival comes down to the mere availability of viable transplant donors. Furthermore, the nature of technological advance means that transplantation must eventually one day be supplemented or replaced by artificial means (e.g. robot hearts, etc.). The significance of the advent of transplant procedures for the definitional problem is that we can now no longer rely on the permanent failure of organs (the brain aside) as a criterion of death. Furthermore, it may be argued also that the fact that the function of certain organs cannot currently be replaced is merely a technical problem which will one day be overcome.⁷

⁶ For an overview of the purposes and practices of an Intensive Care Unit (ICU), see Walton (1987), pp.33-8, which the above account is based on.

⁷ Of course, the advent of transplant technology is one of the main reasons for the debate. See Russell (2000), pp.7-11. I will discuss the potential limits to the transplant and replaceability scenario shortly, especially in relation to criticisms made by Lamb (1978, 1985). See also Lock (1990, 1995, and 1998), who details the cultural reaction to organ transplant in (e.g.) Japan, and highlights the problems that the new technology has given rise to. That the problems are different in different societies, of course, is a further indication here that we are dealing with conflicts in culturally defined concepts.

There have, of course, been many other developments – diagnostic procedures, drug treatments, nasogastric feeding, defibrillators, cardiac by-pass pumps, and so on – which have increased our ability to rescue and maintain the existence of individuals in states of perilous health or near-death states. However, I do not plan to document all of these, but simply to suggest how such advances have together defeated the previously prevalent cardio-pulmonary definition of death, whilst also noting that the host of such changes together have given rise to the possibility of situations where the determination of death is, at best, controversial.

To illustrate such a possibility, imagine the following hypothetical situation:

A young woman is attending a party and, after complaining of feeling unwell, is taken home by friends and put to bed. After fifteen minutes or so, the friends check on her and discover that she is not breathing. An ambulance is called, and mouth-to-mouth resuscitation is performed, resulting in resumed respiration.

On admission to hospital, it is discovered that the cause of respiratory failure was most probably vomiting during sleep (which is thought to have been due to alcohol), which had in turn caused her to choke and cease breathing. She is put on a ventilator to aid her breathing and, whilst her body resumes normal function, she is otherwise unresponsive, showing no reaction to pain and light, and is diagnosed as being in a coma. During this time, an EEG (electroencephalograph) reveals that brain activity is present, but abnormal and at a low level.

After a week, she is moved into the ECU (Extended Care Unit) and she is fed by a tube. After a while, relatives and staff notice that she begins to respond to external stimuli – her eyes occasionally open, and she appears to go through a sleep-wake cycle; furthermore, at certain times she elicits certain sounds (grunts, coughs, even laughs). However, it is quickly realised that these responses are not based on awareness, and that they merely represent the functioning of the non-conscious lower brain. She is therefore considered to have moved out of a coma and into a persistent vegetative state. Furthermore, when a decision is made to take her off the ventilator to test for spontaneous respiratory function, she continues to breathe unaided.⁸

So, what is going on here? Is the patient alive or dead? According to modern medical practice, she is most definitely alive. Firstly, let us go in search of some criteria.

In 1968, the report of Harvard Medical School's Ad Hoc Committee⁹ proposed four main criteria for what it termed 'brain-death' (i.e. whole-brain death):¹⁰

⁸ Some readers may recognise here the case of Karen Ann Quinlan (more or less). See, for instance, Walton (1987), pp.43-5; and, for more detail, New Jersey Supreme Court (1975-6).

⁹ Chaired by Henry Beecher. See Beecher (1968).

¹⁰ Obviously, there has been much discussion as to whether or how to amend and reform these criteria, mostly notably by the University of Minnesota Health Centre (Cranford, 1978) and the President's Commission (Lynn, 1981). However, the proposed changes are mostly procedural (to do with duration of testing, etc.), and differ very little in actual identification of criteria.

- (A) Lack of response to external stimuli (e.g. light, pain) and internal need (e.g. hunger).
- (B) Absence of spontaneous movement and respiration.
- (C) No reflex responses.
- (D) A flat EEG (electroencephalograph) reading – not essential, but useful for confirmation.¹¹

In relation to our above case, we can see that, whilst there was initially a lack of response to pain and light (A), the patient eventually regained these and other responses (C), the EEG – whilst abnormal – was not flat (D), and it was even later found that the ventilator was not required (B). From this, it can be concluded that the brain damage suffered by the patient was mainly to the cerebral cortex (housing the capacity for cognitive function), and that the brain-stem itself (which is responsible for maintaining many of the automatic responses displayed) was largely intact.

In light of the above facts, we can therefore conclude that the patient is in what is termed a ‘persistent vegetative state’ (PVS). It should be noted here that PVS is not the same as irreversible coma: in the former, the brain-stem is at least partially intact, whereas in the latter, it will not be. Coma is, by definition, an *unresponsive* and *sleep-like* state; therefore, by definition, PVS is not coma. Coma commonly lasts around three weeks (at most), after which coma patients either develop asystole (absence of spontaneous contraction of heart muscle), or enter PVS (indicated by the onset of a sleep-wake cycle) – either on the way to recovery, or (as with our patient) to persist there. PVS is therefore akin to neocortical (or ‘higher-brain’) death, in that both awareness and cognitive activity are absent.¹²

The relationship between the various degrees of brain damage, brain death, PVS, and coma can therefore be understood as different points along a continuum:¹³

- Whole-Brain death: In 1959, Mollaret and Goulon¹⁴ coined the term ‘coma dépassé’ (irreversible coma) to describe patients with ‘no spontaneous respiration, no reflexes, polyuria and low blood pressure if norepinephrine was not given continuously, and the absence of all EEG activity.’¹⁵ Therefore, without such drug treatment and the help of the ventilator, the patient would stop breathing and die. This became the basis for subsequent concepts of

¹¹ ‘The electroencephalograph (EEG) records the electrical activity of the cerebral cortex from leads attached to the scalp. An isoelectric or flat reading is the finding required to establish this criterion for brain death.’ Walton (1987), p.70.

¹² For these distinctions, I have followed Pallis (1985) - see pp.34-5. However, it should be noted that many commentators describe PVS as a *type* of coma; sometimes, as in Browne (1985), this is due to confusion or imprecision; at other times, however, it is just due to the use of the term ‘coma’ in a wider sense. Therefore, I shall use coma in Pallis’s sense.

¹³ In listing the range of abnormal brain states which may be relevant to an understanding of this discussion, Russell (2000) also includes Akinetic Mutism (p.41). For my purposes, however, the condition may be grouped with cases of PVS, since Akinetic Mutism involves a degree of intact brain-stem function together with little or no cognitive awareness (an essentially PVS-like state), though the causes are different.

¹⁴ Mollaret and Goulon (1959). Coma dépassé also included death of the nervous system, so it is not identical with the modern notion of whole-brain death.

¹⁵ Machado, et al. (2007), p.198.

whole-brain death, which may be thought of as the most extreme form of irreversible coma.

- Brain-stem death: This is where damage is localised to the brain stem and where the function of other parts of the brain remains intact (although inactive). So, whilst higher brain capacity may be present, since a functioning brain-stem is necessary for higher cognitive functions to work, then effectively the whole brain is dead.¹⁶ This may be compared to the starter motor in a car: there may be no fault with the engine as a whole, but without this central component the whole thing will not be able to function. Such patients must be maintained on a ventilator, and are considered to be in a state of irreversible coma. However, in some cases it may be difficult to determine the full extent of the damage, and patients may slip between this and other states (e.g. PVS). However, this simply reveals the occasional diagnostic problems involved, and for patients whose brain stem has been damaged beyond a certain point, there is no positive prognosis (though just what this point is, and when it has been reached, are problematic diagnostic issues).¹⁷
- Persistent Vegetative State (PVS): In these cases, the function of the brain-stem is wholly or partially intact, allowing for the continued presence of many functions of the autonomic nervous system (e.g. swallowing, coughing, blinking, emitting occasional noises, moving limbs, etc.). However, since there is often damage to the higher brain (neocortex), there is no cognitive awareness attached to these actions and responses. The main distinction between PVS and the above two states is therefore that the automatic processes that are absent in those cases are present in this one. Of course, this will differ from case to case, but PVS patients are usually characterised by the presence of a sleep-wake cycle and a range of functions (e.g. response to stimuli) generally associated with an at-least-partially-intact brain stem. For these reasons, as noted above, some commentators resist the temptation to think of PVS as a type of coma (e.g. 'vegetative coma').¹⁸ It should also be noted that patients do occasionally recover from this state, but because there are different causes of it, and different degrees of brain damage, this obviously differs case by case. However, in relation to this hypothetical case, we shall assume for the purposes of argument that the state is irreversible, and that there is significant neocortical damage, resulting in the permanent loss of the capacity for consciousness.¹⁹

We can, in relation to our patient, therefore, conclude that this is clearly a case of PVS: there is intact brain-stem function (with some possible impairment – hence the initial need for the respirator), but there is a lack of cognitive awareness.

The Nature of the Problem: the Search for Significant Features

¹⁶ Pallis (1983).

¹⁷ See Russell (2000), pp.27-8.

¹⁸ See, for instance, Pallis (1985), pp.33-4 and Walton (1987), pp.51-2.

¹⁹ For an overview of the relation between PVS and other states, see Russell (2000), pp.39-56.

Having outlined the different possible types of brain death, and illustrated one possible hypothetical scenario, let us now consider what sort of problem such a scenario might present for the definition of death.

Firstly, as already stated, the medical profession currently accepts that such a person is definitely alive. Therefore, controversies in such situations generally centre not on whether the patient is dead, but on whether it is ethical to let the patient die (i.e. that life support should be removed).²⁰ However, as we shall see later, there are a growing number of commentators²¹ who, focusing on the capacity for cognitive consciousness and notions of personal identity as central to a living human being, argue that such individuals as envisaged in the scenario are – to all intents and purposes – dead, and should therefore be considered so. Furthermore, in terms of the practical consequences of this definition, the repugnance felt by some at the thought of removing the organs of an individual whose brain stem is functioning, but who will never regain consciousness, is not necessarily shared by another, who might take the permanent absence of the capacity for consciousness to be ‘dead enough’. For instance, as Roland Puccetti argues:

When reasonably assured of a loved one’s neocortical death, it would not have the slightest interest for me that this person was still breathing when prepared for burial, however grisly it might seem to those who have to do that. And I should hope those close to me would feel the same way in my own case. If someone suggested to me that my body might survive death of the neocortex for several months or years, provided it were fed and cleaned properly, etc., that would have no greater appeal to me than preservation of my appendix in a bottle of formaldehyde. For in the sense in which life has value for human beings, I would have been dead all that time. What we are talking about is not a living person but a breathing corpse. And if the notion of burying a breathing corpse is repulsive, then I suggest we simply stop it from breathing.²²

Puccetti therefore implies that, in such situations, we should consider what value a human being’s life has – both to the individual himself and to others.

Such an approach as Puccetti’s therefore provides an interesting contrast to that of what I shall term the more traditional ‘strictly-biological’ approaches²³ because, in talking explicitly of value, it seems to take a significant step away from consideration of death as determinable in a strictly-biological way (i.e. objectively and independently of value concerns, or ‘extra-biological’ concerns – such as ethical arguments or issues of personal identity). Puccetti uses such words as ‘appeal’ and ‘repulsive’, considers the relevance of his own and others’ feelings, and the value that life has for those concerned. For Puccetti, then, the neocortical definition of death is to be preferred not because it is significant in a strictly-biological sense, but rather because conscious human life is ‘the only sort *worth* having’.²⁴ And yet, from the strictly-biological perspective, any *feelings* as to what sort of existence is and is not

²⁰ As, in fact, was the case with Karen Quinlan.

²¹ E.g. Puccetti (1976, 1988), Veatch (1975), Gervais (1986), Green and Wikler (1980).

²² Puccetti, (1976), p.252. The image here of the bottled appendix is borrowed from Flew (1955).

²³ I shall explain this term in detail in the following chapter.

²⁴ Puccetti (1976), p.256 (my emphasis).

‘worth having’ must be viewed as irrelevant (for otherwise, it would not be an objective, strictly-biological definition of death).

Admittedly, not all advocates of such an approach argue that neocortical death allows us to treat the body as we wish, and some suggest that we may still consider the body or organism itself to be alive whilst the person as such is dead.²⁵ However, whatever the details of the individual position, advocates of such an approach share a conviction that consciousness is the feature of most central importance to the question of how to define human death.

In the scenario considered earlier, therefore, what position are we to take? The options would seem to include the following:

- The patient is actually alive. Of course, this would represent the least controversial and most conservative approach. Therefore, proponents of this option would argue that the characteristics presented by the hypothetical case represent significant features of a living human being (e.g. spontaneous function, responsiveness, etc.).
- The patient is actually dead. Advocates of this approach (e.g. Puccetti) would argue that the absence of the capacity for cognitive consciousness is conclusive, and that any residual function is of no significance.
- The patient exists in an intermediary state. It may be argued that the patient actually exists in a state somewhere between life and death (a third state), or that their status is undecided. Obviously, from most perspectives, this would represent an unsatisfactory position, and would arguably simply extend the controversy (e.g. if there is doubt as to where the borderline is between life and death, then this doubt can be extended to the borderlines between life, death, and the third state itself). Accordingly, this is not a position that I am aware of anyone adopting as a *resolution* to the controversy, but rather utilising as an *illustration* of it.²⁶
- ‘Death’ is an arbitrary concept. Obviously, this would be a radical option. However, apart from (arguably) not helping to resolve the controversy we are faced with – and therefore what to do with the patient – it is one which calls for a complete reappraisal of some of the fundamental concepts which human beings employ. Furthermore, that it is an option at all, depends on the answers arrived at concerning other fundamental philosophical issues (e.g. our attitude regarding the status of the concept of death (whether it has a correlate in the ‘real’, mind-independent world), the notion of personal identity, and the nature of self). However, I shall return to this possibility in chapter 8 where I consider religious and science-fiction examples of what I shall term notions of ‘the expanded self’ and its consequence for the related notion of death.

However, the situation is not even as straight forward as these options seem to imply, because – I shall argue – what we are faced with in defining death are a range of

²⁵ E.g. Green and Wikler (1980).

²⁶ E.g. Jonas (1978).

competing criteria, of which consciousness is merely one. For instance, it may be alternatively argued that the continuing presence of spontaneous function is significant, or that the functional integration of the organism as a whole should be taken into account. Correspondingly, our attitude to death will not only be shaped by what sort of thing we take death to be (i.e. strictly-biological, socially-constructed, etc.), but also what characteristics we take to be essential expressions of it.

Competing Criteria

But what are these criteria, and which – if any of them – are sufficient for our purposes? Without attempting to be exhaustive, therefore, we may identify the following criteria:

a) Spontaneous function

The notion of spontaneous function is an obvious one for the definition of life, the ability to breathe unaided being one sign that the brain stem is operational. More generally, the ability of an organism to function autonomously is generally taken as an indication that it is, in some sense, alive. However, in itself, in its presence or its absence, it would not seem to be a sufficient condition for determining death. An individual may require something as simple as a pacemaker, yet we would not be tempted to say that the lack of spontaneous, regular functioning of their heart meant that they were dead. Similarly, whether someone requires kidney dialysis because of failure of those organs is not a deciding factor in determining their status. Certainly, if the person is conscious, then a lack of spontaneity in one or more organs is totally irrelevant. Of course, it might be coherent to imagine that almost *all* organs (apart from, maybe, certain parts of the brain) might be replaced by technological substitutes so as to create a ‘Cyborg’. Such a possibility, as Puccetti notes, whilst ‘technically outrageous, [...] does not require overriding any physiological realities.’²⁷ Furthermore, Jonas argues that

if we could do for the disabled brain – let’s say, the lower nerve centers only – what we can do for the heart and lungs, viz., *make* it work by the continuous input of some external agency (electrical, chemical, or whatever), we would surely do so and not be finicky about the resulting function’s lacking spontaneity: the functioning as such would matter.²⁸

Lack of spontaneous function, then, whilst generally taken as a sign of life, would seem to represent an insufficient singular condition for the declaration of death.

b) Functional integration

One sign of a healthy, living organism is that each of its separate organs or parts plays a functionally integrated role in the health and operation of the

²⁷ Puccetti (1976), p.257.

²⁸ Jonas (1978), p.54.

whole. In other words (some argue), the fact that a dead person still possesses a perfectly functioning set of eyes, or intercostal muscles, or even heart, does not affect the fact that they are dead *because the parts no longer serve a role in a functionally integrated whole* (which is, of course, the ultimate rationale for allowing organ transplantation). As certain commentators argue,²⁹ a lack of functional integration can therefore be considered a sufficient criterion for determining death. However, others³⁰ have pointed out that functional integration without the possibility of consciousness leaves out something central to our concept of a living human being, and that all we are left with is a 'breathing corpse'.³¹

c) Responsivity

A key feature of a living organism is that it can be seen to respond to its environment. For humans, the apparent level of response is therefore frequently used to establish not only whether an individual is alive or dead, but also – in cases of coma – to determine what functions are still active. As a criterion for pronouncing death, this concerns not only higher-level responses, such as conscious vocalisation, or signs of cognitive activity, but the more fundamental, automatic responses, such as pupil dilation, reflex muscle-response, and certain physiological reactions to painful stimuli. The test for these lower-level responses is, in effect, a test for the functioning of the brain stem itself. The deeper question here, however, concerns which sort of responses are important and why. In the famous case of Karen Quinlan, due to 'irreversible brain damage',³² the patient existed in a persistent vegetative state with 'absence of awareness of anything or anyone around her.'³³ Excerpts from court proceedings at the time report the consensus of the medical opinion of the doctors attendant upon her:

A general synopsis of their testimonies indicates that they found Karen comatose, emaciated and in a posture of extreme flexion and rigidity of the arms, legs and related muscles which could not be overcome, with her joints severely rigid and deformed. During the examination, she went through awake and sleep periods but mostly awake. The eyes moved spontaneously. She made stereotyped cries and sounds and her mouth opened wide when she did so. Cries were evoked when there was noxious stimulation. She reflexed to noxious stimuli. Her pupils reacted to light, and her retinas were normal [...] All agree she is in a persistent vegetative state.³⁴

As can be seen from the above, certain responses were present, but others were absent. Furthermore, whilst medical opinion at the time was that 'there is no

²⁹ E.g. Lamb (1985), Becker (1975), Russell (2000), Pallis (1983).

³⁰ E.g. Gervais (1986), Puccetti (1976).

³¹ Puccetti (1976), p.252.

³² New Jersey Supreme Court (1975), p.549. As already stated, this is the loose basis for our hypothetical situation.

³³ Ibid., p.548.

³⁴ Ibid., p.549.

evidence she can continue to exist physically without the respirator',³⁵ ironically, when the decision was taken to remove Ms Quinlan from the respirator, she commenced breathing of her own accord, living on for another 10 years.³⁶ Certainly, the level of response was enough to convince those involved that Ms Quinlan could in no way be classed as dead – even though, as was mistakenly thought at the time, her brain stem could not maintain spontaneous respiration. The level of response, therefore, was the main factor in deciding the patient's status, in that it raised her above the category of those with 'no discernible central nervous system activity',³⁷ (a criterion which would be required for declaration of brain death as stated by the Ad Hoc Committee of the Harvard Medical School).

There would seem to be a range of reactions to this approach. On the one hand, there are those, such as the advocates of neocortical death,³⁸ who would not consider the sort of responsivity here described as a sufficient sign of life. At the other extreme, there are those,³⁹ who would argue that the organism is still alive even when some of these responses are absent, because the functional integration of the organism as a whole is still present. Given this range of disagreement, it is difficult to see in what way responsivity can be used as a conclusive and uncontroversial criterion for determining death. I will return to these issues in the next chapter.

d) Consciousness

As already touched upon, the idea that the absence of the capacity for consciousness should determine death is associated with neocortical death. The neocortex houses the capacity for cognitive function and rational thought, but a functioning neocortex is not in itself sufficient for consciousness. The ascending reticular activating system (or reticular formation) provides the *basis* of consciousness, but as Puccetti points out:

While a functioning reticular formation in the upper brainstem is necessary to consciousness accompanying neocortical functions, without the neocortex such paleocortical structures can maintain only the crudest sort of wakefulness, and nothing like conscious human life.⁴⁰

This sort of attitude, however, as already noted, goes well beyond the strictly-biological approach, and whether an organism is alive is a very different question to whether it is capable of consciousness – let alone, as in Puccetti's

³⁵ Ibid., p.550.

³⁶ There are a number of cases where long-term coma victims have come back to full consciousness after many years – for instance, see Leidig (2003). Obviously, whilst in Karen Quinlan's case the medical evidence suggested that she no longer possessed the neurological means to regain full consciousness, there have been cases where the diagnosis has proven incorrect, which is obviously a further factor in the decisions and hopes of the families involved in the decision-making process. For misdiagnosis involving PVS (regarding which, more shortly), see also Andrews et al. (1996).

³⁷ Beecher (1968), p.85.

³⁸ E.g. Puccetti (1976), Gervais (1986).

³⁹ E.g. Becker (1976), Lamb (1985).

⁴⁰ Puccetti (1976), p.252.

example, whether that consciousness is anything more than the 'crudest sort of wakefulness'. Obviously, this would also be true for criteria involving the capacity for consciousness, such as notions of personal identity (in terms of psychological continuity) or the capacity for social interaction. The suitability of this criterion is therefore a complex issue which is intrinsically linked to the question of what status the concept of death has, and therefore to what extent definitions of death can be considered objective. I shall deal with this issue in chapter 4, and specifically in relation to consciousness in chapter 5. There, and also later on (in chapter 8), I will look at the possibility that consciousness need *not* be considered an essential feature of a living human being.

However, the theoretical argument aside, there is also the question of the degree of diagnostic certainty which may be arrived at in relation to severely brain-damaged patients. For instance, a recent study in a hospital unit specialising in the rehabilitation of severely brain-damaged individuals has shown that misdiagnosis of PVS is more common than might be thought.⁴¹

Of the 40 patients diagnosed as being in the vegetative state, 10 (25%) remained vegetative, 13 (33%) slowly emerged from the vegetative state during the rehabilitation programme, and 17 (43%) were considered to have been misdiagnosed as vegetative.⁴²

Furthermore, of those misdiagnosed,

Fifteen patients were able to make choices, such as selecting their previously preferred music tapes from a choice of three. Twelve (71%) patients were able to spell out their own short messages using the listener scanner technique.

Eleven patients (65%) were able to carry out simple one stage and two stage mental arithmetic tasks such as "What is 10 divided by 5?" and "What is 8 minus 4 minus 1?" Eleven patients (65%) were oriented in time, place, and person, and one patient was only oriented in person. Thirteen patients (76%) were able to recall a name from a choice of three options; and eight (47%) patients were able to use listener scanner techniques to write a letter to relatives.⁴³

Would Puccetti consider the above to be examples of 'the crudest form of wakefulness'? Presumably not, but it remains an interesting question – diagnostic problems aside – as to where the advocate of neocortical death draws the line.

As can be seen from the above, therefore, there is no general consensus as to which criteria are essential or inessential to the definition of death. However, what has been largely ignored in the debate so far, I shall argue, are the reasons *why* such differing

⁴¹ Andrews, et al. (1996).

⁴² Ibid.

⁴³ Ibid.

criteria should come into competition in this way. There are, I shall argue, compelling philosophical reasons not only as to why such conflicts occur, but which also suggest why the approach of most commentators to date has been misguided, or even shallow (philosophically speaking).

I shall therefore now attempt to indicate what I think are some of these reasons.

A Messy Divorce: Origins of Conceptual Ambivalence

The above four criteria represent fairly uncontentious aspects of what would generally be thought of as a living human being. In other words, an individual whose organs form a part of a functionally integrated organism, which in turn functions spontaneously, is responsive to its environment, and facilitates the conscious cognitive function of the individual concerned, would be generally considered by most to be a living human being. Of course, as already suggested, we may reduce any or all of these signs by a certain amount and still uncontentiously retain this description; a human being with reduced mental capacity is not the less a human being – or alive – for all that; artificial limbs, organs, and other mechanical aids to functioning similarly do not force us to revoke such an individual's status.⁴⁴ However, there does come a point – a threshold – where the reduction in the degree, or even complete removal, of one or another aspect of a whole, fully-functioning human being provides occasion for controversy. It is at this point, I shall argue, that the previously unified concept is 'split', and its heretofore inseparable aspects are set at variance to one another. Therefore, whereas it was previously unthinkable that a human being lacking spontaneous cardiac and respiratory functions could be kept 'alive', or that an individual in PVS might live on for years (indeed, that there could *be* such a state), medical technology can now allow such situations. In consequence, such aspects of the human being which were once thought of as being inextricably dependent upon one another – and thus, their cessation as being necessarily coincident – can now exist separately.

So, in a sense, medical advances have provided as many problems as solutions; or, as Robert Schwager puts it,

the very success of modern medical science raises both ethical and legal questions where previously none existed. For in the past the ability to save life was so limited that doing so rarely became problematical. Now, however, our abilities in this area are so prodigious that questions about their appropriate use arise daily.⁴⁵

If such a scenario is considered in relation to the conflicts within common or traditional notions of 'life', 'death', 'person', etc., it can be seen that what has happened through advances in medical technology is that the way in which the concept (or term) 'living human being' should be correctly applied is now informed by competing criteria. Up until now, the criteria and the occasions for applying the term have coincided; an individual with no cognitive function could not be kept alive.

⁴⁴ This is not to say, of course, that such attempts at status reappraisal do not or have not occurred – one need only think of how the reclassification of the mentally ill and infirm as 'non-persons' was used by the Nazis in the 1930's to justify a programme of active euthanasia.

⁴⁵ Schwager (1978), p.39.

However, occasions now exist that allow the divergent application of the term in relation to different criteria. Thus, rather like a messy divorce, what was previously ‘ours’ must now be separated into ‘his’ and ‘hers’, and criteria agreed for that purpose.⁴⁶ The establishment of agreed criteria is contentious: should ‘But my mother gave us that!’ override an agreement to simply divide things up equally? What about sentimental versus financial value? However, unfortunately, few would agree that the definition of death should be resolved by the equivalent of a pre-nuptial agreement.⁴⁷

To anticipate my argument somewhat, let us seek to give the situation a philosophical footing. An analogous philosophical concern can be found in Wittgenstein’s analysis of what is involved in ‘following a rule’.⁴⁸ Wittgenstein imagines the following situation:

[Person] A writes series of numbers down; [person] B watches him and tries to find a law for the sequence of numbers [...] A has written down the numbers 1, 5, 11, 19, 29; at this point B says he knows how to go on. What happened here? Various things may have happened⁴⁹

By ‘various things’ Wittgenstein not only means ‘different ways of expressing the same rule’, but also – arguably – different rules. For instance, we might express the rule as ‘subtracting the penultimate number in the series from the ultimate one, adding 2 to the result, then adding the result to the last number’ (i.e. $19 - 11 = 8 + 2 = 10 + 19 = 29$). However, as Wittgenstein points out, he might just simply think:

“What is the series of differences?” He finds the series 4, 6, 8, 10 and says:
Now I can go on.⁵⁰

Of course, in a sense, the two rules are the same (in that they produce the same result). In mathematical terms, we would say that they are equivalent.⁵¹ However, what if certain circumstances arose which revealed that, actually, B was not applying A’s rule at all, but only one which in all but exceptional and unusual circumstances produced an identical result? We would probably say that B had not really grasped A’s rule. However, what if the rule in question was not one which A had thought up, but was a generally applied term familiar to both A and B?

⁴⁶ I will explain this idea later on in a more technical way by utilisation of Quine’s notion of the ‘opacity of reference’ of a term.

⁴⁷ Joking aside, of course, a so-called ‘living will’ might act in a similar way – as long as the fulfilment of its wishes were within the law of the land. However, this would represent, at most, one individual’s attitude to death and quality of life, and would therefore be likely to differ between individuals. Furthermore, to leave the determination of death up to individuals would be an extremely unsatisfactory position scientifically, to say the least. I shall consider some of these questions in the last chapter.

⁴⁸ For a full discussion of this issue, see Kripke (1982). Kripke is sometimes criticised for misinterpreting Wittgenstein for his own purposes here (hence the coining of the term ‘Kripkenstein’ by some commentators), but I think for a general introduction to the problem his exposition is fine as the controversy centres around what Kripke takes to be Wittgenstein’s solution to the problem.

⁴⁹ Wittgenstein (1967), p. 59 (section 151).

⁵⁰ *Ibid.*, p.60.

⁵¹ Though we might still debate the point here as to whether they *meant* the same thing (in that they are in a sense different expressions of the rule). It depends, of course, on to what extent Wittgenstein is read as proposing a behaviourist-type account of meaning (as, for instance, Quine is).

My point here is that the parallel with the definition of death scenario lies in the fact that up until medical practice allowed situations within which individuals might diverge in their application of the term 'living human being' (or 'living person'), the 'rule' applied was (to all intents and purposes) 'opaque'. That is, the features of being alive which were significant to the rule-user were, in a sense, hidden. In fact, it is perfectly possible that *there was no consideration of certain features as more significant than others*, and that the 'rule' applied was the same in each case (and from individual to individual), but only differed *after* the advent of situations which allowed for the diverse application of the same terms (though of course, both options are possible).

But why should this problem not be resolvable through conceptual analysis, or through the application of the understanding of the biological process of death which science has – or at some future time, will have – afforded us? The reason why neither of these possibilities represents a solution, I shall argue, is twofold: (1) there is a necessary 'opacity' to our linguistic concepts which novel situations can expose as ambivalent (where what I have called 'conceptual ambivalence' comes into play); (2) whilst there are established means of resolving such ambivalence in certain cases, the concepts we are concerned with (i.e. 'death', 'person', etc.) are – in Hilary Putnam's phrase – 'epistemically thick'⁵², in that they involve both factual and evaluative aspects. Furthermore, it may be the case that the concept of death may actually share more of the value side of the divide than the factual side; or even that the concepts which interweave with it (e.g. personal identity, autonomy, self, etc.), show that the concept of death itself is far more a socially-constructed concern than has previously been thought.

We must therefore face the possibility that the type of conflict that competing definitions of death represent is something which, as Brante and Hallberg suggest, 'cannot be terminated by rational means':⁵³

Can we achieve a better understanding of this controversy by employing the concept of incommensurability? Are the problems, perspectives and standards of the proponents and opponents partly different? Do the mutually adversary perspectives constitute such dissimilar 'discrete entities' that they cannot rationally be compared? The difficulty in communicating indicates that this is indeed the case.⁵⁴

I will return in detail to the notion of conceptual ambivalence in chapter 3 where I look at how it relates to the problem of category membership. Also, I will deal with the possibility that death is a socially constructed (or, observer-relative) feature of the world (and therefore not necessarily purely rational) in chapters 4 and 7. However, before progressing to the main treatment, I will now provide a brief chapter outline.

Outline of the Argument

Firstly, in Chapter 2, I will consider the various attempts to resolve this problem to date, and consider primarily the attempt to define death as a strictly-biological

⁵² Putnam (2002), pp.28-45.

⁵³ Brante and Hallberg (1991), p.410.

⁵⁴ Ibid., p.404.

problem. This will therefore concern approaches which resist the appeal to what may be termed extra-biological criteria (e.g. personal identity, moral arguments), and attempt to identify those biological criteria which may be considered necessary and sufficient for the determination of death.

In chapter 3, in light of what I take to be the failure of the strictly-biological attempt, I look at some reasons why I think such a failure is necessary. In doing this, I define the problem as one which is reflected in the general epistemological issue of category membership. Furthermore, utilising certain approaches stemming from Quine and Wittgenstein, I argue that such a problem is – due to the nature of logic and language – one which renders the problem of death irresolvable in purely rational terms (or at least, according to a positivist-type view of rationality).

As a consequence of the conclusions of the previous chapter, chapter 4 will look at the question of what status the concept of death has. To do this, I utilise some of John Searle's epistemological framework, and argue that – even if such a realist perspective as Searle's were true – death cannot be considered anything other than a human-centred, 'observer-relative' feature of reality. I furthermore argue that it is for this reason that the concept of death cannot be ultimately grounded in absolutely rational terms, and which therefore also gives rise to the problems of category membership considered in the previous chapter. Finally, in concluding that death is an observer-relative phenomenon, I also reveal the true reasons why any attempt to define human death in strictly-biological terms must fail.

Chapter 5 begins the consideration of extra-biological features of the human being in relation to the definition of death, and begins the analysis of 'partly-biological' views. Firstly, the status of consciousness itself is considered, and its relation to Searle's distinction between 'intrinsic' and 'observer-relative' features of the world is discussed. In addition, contemporary controversies regarding the nature of consciousness are examined, and their significance to the death debate is weighed up. Finally, I consider the question of whether the permanent absence of the capacity for consciousness may provide us with a necessary and sufficient criterion for human death.

In chapter 6, I consider various personal-identity-based attempts to present a successful definition of death. Also, and aside from the analysis of individual accounts, I consider the question of whether *any* personal-identity-based approach can free itself from the type of problems and limitations identified in previous chapters.

Having concluded the observer-relative status of the concept of death, chapter 7 looks at what this means in general terms for our application of it. Death is therefore compared with other similar types of concept and the general debate concerning the distinction between facts and values. Finally, light is shed on controversies concerning our application of the concept of death via analogy with philosophically equivalent scenarios (e.g. sport and games).

Chapter 8 begins to 'stretch' the concept of death, and to consider it in relation to extensions of its possible application. Religious and science-fiction scenarios are considered, and various examples are used to illustrate the problems we face in setting limits to the concept. Given the possibility that the concept of death has more breadth

than may have first been imagined, are certain concepts of 'immortality' made viable? Is it even possible that such interpretations may allow for a more rational foundation for certain religious doctrines?

The final chapter returns to the everyday world and briefly looks at what possible consequences there may be for the foregoing conclusions in relation to practical concerns (i.e. medicine and law). Given that the concept of death is part of the socially-constructed public domain, and that we are faced with multiple possible parallel definitions of death, what rights or responsibilities are attendant upon individuals or organisations concerning death? Is there room for individual choice concerning our interpretation of what death is? Or must the wise legislate for the protection of the populace?

2. A Biological Definition of Death

Introduction

In the search for a clear and uncontroversial definition of death, the first port of call would intuitively seem to be an attempt to isolate sufficient biological criteria. In most people's eyes, a living, functioning body is the essential prerequisite for 'being alive'; it would only be natural, then, to consider the definition of death to involve certain biological facts. For centuries, death was determined by the cardio-pulmonary method; that is, whether the patient was seen to maintain spontaneous heartbeat and respiration. Comparatively recently,¹ much controversy has been provoked by the suggestion that some form of 'brain death' provides a more accurate means of determining death.² However, whilst some of the consequences of adopting such a criterion will be of interest later, my initial focus in this chapter will be to investigate the very notion that a *strictly*-biological definition of death is in fact possible. I say 'strictly' here because I wish, first of all, to distinguish between two ways in which biological criteria may play a part in the definition of death. Firstly, it may be argued that the determination of death is a strictly-biological matter, and that the fulfilment of biological criteria provides both necessary and sufficient conditions for declaring an individual 'dead'. I shall therefore refer to this hereafter as the 'strictly-biological view'.

On the other hand, however, it may be contended that whilst certain biological criteria may provide *necessary* conditions for the determination of death, they do not provide *sufficient* ones; in other words, the assertion that certain biological facts about the patient's condition provide us at most with criteria which are relevant to – and may even be necessary for – the determination of death, but which are not sufficient in themselves to constitute death.³ For instance, it may be argued that, in cases where brain damage has rendered cognitive brain states impossible, but the body and 'lower' brain are still alive (such as in cases of persistent vegetative state), then the decision to declare the patient dead is not a strictly-biological one, but rather rests on certain extra-biological factors such as our concept of a person, or moral considerations as to what rights such a being is entitled to. In other words, there is a sense here that even where there is no dispute over the facts, what the facts *mean* is still open to debate –

¹ Recently, that is, in relation to the whole of human history: the Ad Hoc committee of the Harvard Medical School to Examine the Definition of Brain Death published its report advocating brain death as a criterion for death, 'A Definition of Irreversible Coma', in 1968. See Beecher (1968).

² It is not my concern here to enter into the debate about whether such brain death actually represents a new criterion, or whether in fact the 'traditional' cardio-pulmonary method was actually brain death 'in disguise', but rather to investigate whether a strictly-biological definition of death is in fact possible. The actual status of brain death criteria in relation to the traditional methods of determining death is discussed at some length in chapter 2 of Gervais (1986), where she rejects the view that there is any strict commonality between the two.

³ I would distinguish here between those criteria which are *necessary* for the determination of death, and those which are *relevant* to it. For instance, it is a fact that many would consider to be relevant to the determination and definition of death that an organism may lack spontaneous function, but that fact may not in itself be necessary for death, for some may hold that death may take place whilst spontaneous function is still intact.

perhaps because, ultimately, death is not a thing which may be strictly determined by 'the facts'. I shall refer to this as the 'partly-biological view'.⁴

Obviously, adherents to the first view will deny that it is wise, or even possible, to make such a distinction. It may be argued that whilst there may be controversy over precisely *which* criteria are necessary, the fact is that medical science has – or one day will have – resolved these controversies, and provided us with a clear (and strictly-) biological definition of death. The second view, on the other hand, seems to reject the idea that death can be explained by a reductive appeal to purely biological criteria, suggesting instead that any successful definition must necessarily involve features and concepts which take into account the wider features of being a human person (such as being conscious, having social relationships, personal identity, and so on). However, this rejection would seem to come with its own problems: if the concept of death *cannot* be defined in a strictly-biological manner, how is consensus as to its meaning and application to be arrived at? Does such an approach leave the door open to the dangers of indeterminacy, whereby what death is becomes a battleground of competing values and conventions?

I will examine in subsequent chapters the relative merits of the various partly-biological views, but now I will first address the question of whether death can be considered a strictly-biological phenomenon.

The Strictly-Biological View

It is not my place to enter into medical debates concerning *which* biological criteria may or may not constitute the loss of a certain capacity, or to decide under what circumstances the death of a particular organ takes place; and whilst such issues as whether consciousness resides in a certain part of the brain, or whether localisation of certain cognitive functions is in fact possible may occasionally be relevant to my discussion, I cannot ultimately pass judgment upon the truth of any of these questions. In one sense, therefore, in dealing with the issue of identifying sufficient biological criteria for the determination of death, I am completely reliant upon the truth of the picture of human functioning established by experts in the field of medical science.

However, having said this, it is important to distinguish between so-called 'medical facts', and any ontological or moral consequences which these may be said to have. As philosophers, it is with these inferences that we should be primarily concerned, and one of the main purposes in exploring the problematic issues which surround the determination of death in certain cases is to establish at just what point such issues become a legitimate philosophical concern. Therefore, in order to do this, we must along the way determine to what extent – if any – death can be considered a strictly-biological matter. In this sense, there will therefore be occasions when it will be necessary to distinguish between assertions of medical fact (on the one hand), and the ontological, moral, and metaphysical conclusions which some commentators think are entailed by them (on the other). Consequently, I will question the legitimacy of such

⁴ This is similar to a point made in Gervais (1986), p.2, where she argues that there is a 'decision of significance' involved in deciding which biological criteria are relevant to the determination of death. This point is of some importance, and I shall return to it in more detail in subsequent discussion.

inferences, and whether, in fact, they are so closely entailed by statements of medical facts as such philosophers think.

To begin with, it should be acknowledged that the strictly-biological view has (arguably) the weight of common sense and public opinion on its side.⁵ For many, the notion that death is anything *other* than a purely biological matter therefore sounds like a strange way of talking. Isn't it straightforwardly obvious that the determination of death merely involves biological criteria? What role might a philosopher play in this debate anyway? There are, of course, a large number of philosophical treatments of the subject of death, but a large percentage of these to date deal with what might be termed the existential consequences of our ultimate demise.⁶ However, in the context we are concerned with here, death is problematic in a different way, for – at least on the face of it – it is not what death *means* to us that is important, or how we should face life in consequence of death's inevitability, but what in fact it *is*, and how it may be clearly defined.

However, the first question to address here is what is meant by 'a strictly-biological definition of death'. I will take this phrase to refer to a view of death that denies the relevance of certain moral and metaphysical concerns – whilst they might be relevant in the wider context (e.g. the ultimate fate of the organism) – to its definition. The proponents of such definitions make two main assumptions: (1) the definition of death is a factual concern involving biological criteria only, and (2) it is possible in this way to operate in a realm of facts divorced from 'external' considerations (e.g. moral or metaphysical concerns).

Since (2) is the wider, more theoretical assumption, I will first address (1), using the issues which arise there as a jumping off point for the more abstract debate (whilst I will touch upon (2) towards the end of this chapter, it will be more directly addressed in subsequent chapters).

Death as a Biological Fact

The treatment of death as a factual concern involving only biological criteria supposes that the controversies surrounding this debate can be allayed by appeal to purely biological matters of fact. As Karen Gervais puts it:

The biologists have claimed that it is possible to put forth a strictly-biological definition of death. They assume that death is essentially a biological phenomenon, and that the effort by moralists to define human death fails because it confuses a quality-of-life issue with a definitional problem.⁷

⁵ In Western secular society at least.

⁶ See, for instance, Luper-Foy (1987), Nagel (1979), Rosenbaum (1989), Cox (1999), and Schultz (2000), which consider such questions as whether or not death can always be considered a misfortune for the person concerned, the significance and legitimacy of various attitudes towards death, and other existential questions.

⁷ Gervais (1986), p.46. Note that by 'biologists' and 'moralists', Gervais is characterizing two of what she thinks of as the three main approaches to the definition of death (the third she labels the 'ontologists'). However, in doing this, Gervais is merely following a tripartite distinction first made by Green and Daniel Wikler (1980). Thus, what I term the 'strictly-biological view' covers the biologists,

The implication of such a biological approach is therefore that this ‘definitional problem’ can be solved by the isolation of the correct criteria. Versions of the strictly-biological approach therefore usually concern the sufficiency of one of the following sets of possible criteria:

- The cardio-pulmonary method (the ‘traditional’ definition): the cessation of heartbeat and breathing.
- Organismic death: the death of the whole organism.
- Whole-brain death (the definition currently adopted in the US): the irreversible cessation of all brain activity (due to neurological damage caused by anoxia (lack of oxygen) for example).
- Brain-stem death (the definition currently adopted in the UK): when ‘the brain stem has been damaged from whatever cause in such a way and to such a degree that its functions (which include the neural control of cardiac and pulmonary function) are irreversibly destroyed’.⁸
- Neocortical death (or ‘higher-brain’ death): the irreversible loss of the capacity for cognitive thought, rationality, consciousness, etc. (though functional integration may still be possible).

I will therefore now briefly consider these approaches in the light of prevalent criticism, before going on to point out what I personally consider to be the ultimate problems for the strictly-biological view in general.

The Cardio-Pulmonary Method: Jonas and Evans and Pre-rational Attitudes

Prior to the advent of medical technology that allowed for the artificial maintenance and prolongation of life (e.g. the ventilator, heart-transplant techniques, etc.), the traditionally accepted criteria for determining death were the cessation of spontaneous respiration and heartbeat. Thus, in Shakespeare’s *King Lear* we find Lear seeking to confirm the presence of life in his daughter, Cordelia, by the presence of breath:

Lend me a looking-glass;
If that her breath will mist or stain the stone,
Why then she lives.⁹

The misted mirror might therefore be thought of as one of the first technological aids in the diagnosis of death.

However, as Tom Russell rightly points out, the diagnosis of death has long been a problematic and complex issue, and it should not be assumed that there was ever a ‘Golden Age of Heart and Lungs’.¹⁰ Rather, we should take the persistence of the cardio-pulmonary criteria as a sign of the importance of these features in our concept of a human being. For instance, spontaneous respiration and heartbeat go a long way

whilst the ‘partly-biological view’ covers the moralists and ontologists, and any other non-strictly-biological view.

⁸ Department of Health (1998), p.1.

⁹ Shakespeare (1972), V.3.259-61.

¹⁰ Russell (2000), p.11.

towards explaining the disquiet we feel at the living, breathing, but (in some cases) irreversibly unconscious patient in a persistent vegetative state.

This said, given the growing number of arguments advanced in support of brain-death criteria (some of which we shall consider shortly), and its increasing acceptance in medical practice, it might be assumed that the cardio-pulmonary criteria must now be considered obsolete. However, this is not wholly the case, and there are still advocates of a heart-lung centred approach. Two such advocates that I will briefly consider here are Hans Jonas and Martyn Evans.

Firstly, Jonas¹¹ argues that adopting brain-death criteria ignores the significance of the functioning of the rest of the organism for our concept of a living human being. In this regard, the continued beating of the heart and potential for spontaneous respiration in a brain-dead patient is still a sign of the ‘residual continuance of the subject’,¹² and must therefore be respected. In this sense, Jonas is primarily concerned with resisting the potential abuses of the existence of beating-heart potential transplant donors.¹³

Similarly, Evans argues that

Disputes about which functions ‘count for life’ [...] are not, and never have been, purely scientific. They are disputes among rival conceptions of what human life essentially is: that is, they are essentially metaphysical disputes. Since science is, and must remain, silent on such disputes, how can they be settled?¹⁴

For this reason, Evans considers ‘our primitive, pre-rational reactions and attitudes’¹⁵ to provide us with grounds on which to base our moral attitudes to the near-dead. So, because of the moral repugnance that is generally felt at the thought of the ‘cremation of the “beating heart cadaver”’,¹⁶ we must reject the adoption of brain-death criteria which would allow this to happen.

Therefore, modern proponents of the cardio-pulmonary method (such as Jonas and Evans) need not necessarily be proposing an alternative method of *determining* death (as if they advocated a return to the (outdated) practices of a hundred years ago), but are rather seeking to set a conservative threshold to the *declaration* of death. The issue is therefore not that some version of brain-death criteria does not in fact account for the cessation of those features which it is claimed to be responsible for; in other words, Jonas and Evans do not seem to be arguing that there is more going on in the human being – in a biological sense – than brain-death advocates assume. Rather, it would seem that they simply see the notion of the person as more deeply embedded in our primitive reactions and attitudes (as Evans calls them) than do the brain-death advocates. Furthermore, ignoring these reactions does not leave us with a scientifically respectable definition of death, because the alternative attitudes (i.e.

¹¹ Jonas (1974).

¹² Ibid., p.58.

¹³ As is notoriously the case in China – see, for example, Wang (2002).

¹⁴ Evans (1990), p.229.

¹⁵ Ibid., p.230.

¹⁶ Ibid.

those which involve brain-death criteria) are nonetheless based on similar, non-rational attitudes as to what is significant about a person.

This is a position with which I have a particular sympathy – though I do not necessarily wish to propose a return to the cardio-pulmonary method, or indeed to advocate the sort of appeal to primitive reactions that Evans and Jonas seem to suggest. However, I do feel that the recognition of what might be called ‘extra-biological features’ of the person – which includes personal identity, social role, the emotional reaction provoked by their status, etc. – provides a better way-in to understanding what human death really is.¹⁷

I will return to this point repeatedly over the course of the following chapters, but for the moment it will suffice to note that an approach such as that of Jonas and Evans serves to widen the debate in fruitful ways. So, whilst there are those aspects of (for instance) Evans’s approach which directly address the more technical issue of whether or not brain-stem death is an empirically clear enough diagnostic candidate, there is also the recognition that what we are faced with here are competing criteria (some of which are biological, and some of which are *extra*-biological). So,

brainstem criteria for human death must in any case be perceived as in competition with other rival criteria for death since scientific criteria arise out of a given conception, and do not adjudicate among different conceptions¹⁸

This seems a Wittgensteinian point: our concepts are not neatly circumscribed by a rational boundary, but are interwoven with our pre-rational reactions within the ‘forms of life’ in which they have their setting. For instance:

It is what human beings *say* that is true and false; and they agree in the *language* that they use. That is not agreement in opinions but in forms of life.¹⁹

Such agreement therefore is not a matter of rational agreement in what we hold to be true or false, but rather in the common practices we share, in that which ‘has to be accepted, the given’ (‘forms of life’).²⁰ It is that we agree in such primitive reactions that we *can* agree rationally – if we did not, then the whole basis of meaning would disappear. A further consequence of this, of course, is that we cannot neatly divorce certain aspects of our concepts from one another (not, at least, without distorting them or acting arbitrarily), and furthermore – for instance – we should not ignore or seek to ‘explain away’ those aspects of our concepts which no longer fit a rational reductive account, for they may be an integral part of our ‘form of life’.

To return to the significance of beating heart criteria, in this sense, many advocates of brain-death criteria who have criticised the cardio-pulmonary method would seem to have missed the significance of this aspect of the debate. For instance, in arguing

¹⁷ It is an important question as to whether human death is in fact of a different kind to death in the natural world in general. I will consider this later on.

¹⁸ Evans (1994), p.4.

¹⁹ Wittgenstein (1967), p.88, s.241.

²⁰ *Ibid.*, p.226.

against Jonas and Evans, Tom Russell construes the point as being whether or not brain-death criteria are sufficient for determining death. So, his arguments focus on such questions as whether, since the hypothalamus controls the homeostasis of the organism (including spontaneous respiration, heartbeat, etc.), we can consider lack of homeostasis as coterminous with death. Russell further points out that the function of the heart itself is a reflex that can be kept going long-after the rest of the organism has failed, and that therefore we must reject the cardio-pulmonary criteria as insufficient.²¹ I shall return to Russell's position shortly, but it should be observed here that he treats Jonas and Evans's position solely as providing competing *biological* criteria for death, whereas – I argue – the position is more complicated and subtle than that: technical arguments aside, death, they suggest (and I agree), is not something which can be decided in purely biological and scientific terms.²²

It may be objected here that, in siding with Jonas and Evans on this point, I am favouring the irrational over the rational. However, I think this misses its mark in assuming that our concepts can be completely rationally grounded. I shall return to this issue in detail in the next chapter.

Organismic Death: Laurence C. Becker and the 'Being/Has Been' Boundary

Rather than focusing on specific organs or processes, some commentators have chosen to argue that death should be construed as death of the organism as a whole. In exploring this view, I will concentrate on one of its main advocates, Laurence C. Becker.

In his article 'Human Being: The Boundaries of the Concept',²³ Becker argues that the main focus of any attempt to define death should centre on the stage at which the biological organism of the human being ceases to function as a whole. In consequence, Becker rejects not only non-biological approaches to defining death (such as appeals to the quality of life of the patient, or his or her right to life, or even personal-identity-based arguments), but also biological approaches which single out the failure of specific organs within that organism (for instance, the brain or heart) as a sufficient criterion for declaring death.²⁴

A human organism is dead when, for whatever reason, the system of those reciprocally dependent processes which assimilate oxygen, metabolize food, eliminate wastes, and keep the organism in relative homeostasis are arrested in a way that the organism itself cannot reverse [...] Loss of consciousness is not death any more than is the loss of a limb.²⁵

²¹ Russell (2000), pp.140-6.

²² As Evans himself states, methodological problems aside, 'my principal objection to the brainstem conception of death is itself a conceptual one; [...] science cannot adjudicate between rival conceptions.' (Evans 1994, p.7).

²³ Becker (1975).

²⁴ As will be seen shortly, in this respect Becker's position is similar to that of Russell (2000). However, whilst both may be said to identify homeostasis as a significant feature, Becker refuses to identify a locus for this process, whilst Russell identifies the hypothalamus in the mid-brain. This said, points which count against Becker may also count against Russell.

²⁵ Becker (1975), p.353.

Becker's approach is therefore a holistic one which looks to the state of functioning of the organism as a whole. It should be noted here, however, that it is possible to distinguish between (1) 'death of the whole organism', and (2) 'death of the organism as a whole'. We can argue, for instance, that 'death of the whole organism' requires the death of each and every organ in the body, whilst 'death of the organism as a whole' merely requires the permanent loss of the integrative functioning of the whole organism.²⁶ Since the former can occur long after medical practice has uncontroversially pronounced death (especially if we define organism to include not just organ function, but life processes in general), then (1) is generally too conservative a criterion. Therefore, in talking of 'death of the organism', (2) is generally intended.

In taking this approach, Becker considers both the beginning and end of life, arguing that the beginning – or the 'becoming/being' boundary, as he calls it – should be marked by the completion of the generative phase of development of the foetus, whilst the end of it – the 'being/has-been' boundary – should be marked by the disintegration of the functional organisation of the organism. The fact that the organism is kept alive by artificial means is not in itself a reason to consider that organism dead; 'it is simply an organism kept alive mechanically'.²⁷

In the light of this definition, we can see that Becker views death as a process, rather than an event. The gradual failure of individual organs thus represents the process of dying, a process which is only complete when the system as a whole cannot be functionally maintained, either by its own means, or artificially. It is this holistic approach, therefore, which characterises Becker's opposition to the adoption of brain death as a sufficient criterion for death itself.

Brain death is not a definition of death, not even a criterion of death. It is merely a criterion for deciding when coma is irreversible.²⁸

Becker therefore wishes to distinguish between a clear, biological definition of death, on the one hand, and moral prescriptions for the near-dead on the other. To consider someone in a state of irreversible coma 'dead' is, for Becker, 'rigging the definition of death'.²⁹ Rather, Becker is content to allow his 'being' and 'has-been' boundaries to define the concept of a human being, and for the moral arguments to take place separately. This means, for Becker, that whether the patient is alive or dead does not directly entail the presence or absence of moral obligations.

People live, but sometimes in such hopeless conditions that one may morally and legally give up trying to save them. People die, but sometimes can be revived. Their death does not in itself relieve us of moral obligations toward them. The reversibility of death is more likely the moral divide. [...] The morality of dealing with the dead, whether

²⁶ See Jonas (1978), pp.53-4, where he defends himself against the charge of confusion in the use of these two terms.

²⁷ Becker (1975), p.353.

²⁸ Ibid., p.355.

²⁹ Ibid., p.356.

reversibly dead or not, is a matter for further argument. It is not settled by this definition.³⁰

It can be seen here that Becker's definition of death leads him into talk of 'reversibility'. This is somewhat puzzling: having defined death as the functional disintegration of the organism, Becker seems to imply that, because death is a process, there may be some doubt as to when exactly the functional disintegration has reached an irreversible stage: 'people die, but sometimes can be revived'. This interpretation is borne out by Becker's reference to the recovery of a 5-year-old boy who had been submerged in icy water for 22 minutes.³¹ Here, although the boy appeared completely dead, he in fact made an almost total recovery. So, Becker implies, his 'death' was 'reversed'. Death is therefore reversible in the sense that we may decide to continue heroic efforts beyond a point where it would be acceptable – legally and morally – to give up.

This is where Becker's view becomes confusing. Having been at pains to define death in strictly-biological, non-moral terms, his introduction of the concept of reversibility seems to undermine his efforts. To say that death becomes reversible in relation to legal and moral concerns would seem to suggest that such extra-biological factors have a role in determining when we might consider someone dead (if moral considerations can reverse death, then whether someone 'recovers' from death is partly a moral decision). Furthermore, in the above case, it would seem more natural to talk of the boy as not actually being dead, rather than that his death was 'reversed'. Becker has already suggested that death is a process (in as much as functional disintegration is gradual), so why then does he insist on reversibility? Processes, of course, are sometimes reversible, but it is arguably better to think of 'death' as the last stop on the journey of 'dying'; so, someone is 'dead' when the process of dying is complete. It would therefore seem to be more in keeping with his other views for Becker to talk of death as an extended and potentially shifting point (in relation to knowledge and technology) at the end of the continuum of dying. However, to present death as a distinct point which we can go beyond and return from suggests that what has been defined is not death at all, but rather something which resembles it (just as, in the boy's case, hypothermia might). In such a case, it is the process of dying which has been reversed, and not death itself.

It is somewhat surprising, in the course of reading Becker's paper, to reach this point of confusion. It seemed that, up until that point, Becker had distinguished clearly between cases of death from functional disintegration of the organism, and cases where individuals were technically alive, but where it might be morally permissible to withdraw artificial support and heroic efforts. Furthermore, given his rejection of brain death as a sufficient criterion for death itself, it would seem that moral arguments could be supplied to justify the withdrawal of artificial support from the irreversibly comatose without actually labelling them 'dead'. However, the introduction of the possibility of reversibility muddies this distinction, for instead of saying, 'Although this person is still a human being and is therefore technically alive, we may make a moral decision to remove life support', Becker now seems to be saying, 'There are certain occasions when a person may appear dead and, whilst it

³⁰ Ibid., pp.357-8.

³¹ Ibid., p.357, n.19.

would not be immoral to discontinue heroic efforts, successful continuation of those efforts may “reverse” the patient’s death’.

Whether Becker is trying to simply allow for the frailty of human knowledge in diagnosing death, or the progression of technology and its role in ‘reversing’ death, is unclear, but it would seem that he cannot on the one hand complain of the attempt to ‘rig death’ (in the case of supplying a moral divide), whilst on the other admit the possibility of reversing it according to moral criteria – especially after having been at such pains to draw a non-moral divide in the first place. Wouldn’t it have been simpler to say, in the above case, ‘We *thought* the boy was dead, but we managed to bring him back’? Becker’s approach would therefore seem to leave us with possible instances where it is open as to whether someone is actually alive or not, for even though we have all the biological facts, their interpretation is subject to the possibility of reversal.

This point of confusion aside, there remain other problems with Becker’s view. First of all, it may be argued that a purely organismic approach to death leaves out a great deal of what is significant about the concept. This is highlighted by Gervais:

Becker’s position commits us to the claim that someone who has been decapitated, yet whose body is being maintained so that “those reciprocally dependent processes which assimilate oxygen, metabolize food, eliminate wastes, and keep the organism in relative homeostasis” (353) continue in stride, is alive.³²

Here, we are left with the somewhat absurd situation that a headless body can be considered alive. But *why* is this absurd? Perhaps Becker would argue that the role of the ‘lower’ brain has been replaced by artificial means, and that what we are left with is no different to someone in a state of irreversible coma whose higher cognitive functions will never return. However, even if Becker could defend himself in such a way, Gervais has a point: the reason we find such an example initially absurd is because it violates in a number of ways our traditional notions as to what is essential to a human being: firstly, it is difficult to imagine that a body existing without a head is a human being – except maybe in nightmares or horror films (though we might imagine a head existing without a body);³³ secondly, the permanent absence of consciousness is brought home to us in the most graphic way (whereas a body with its head intact, even though it has suffered irreversible loss of higher cognitive function, still goes some way to suggesting to us – merely as a result of bodily integrity – that

³² Gervais (1986), p. 57 (the number in brackets in the quoted passage refers to the page of Becker’s article which she is quoting from).

³³ In fact, the picturing of a headless body as purposively alive is a key component of the horror we feel (as is wonderfully brought out in the film *Sleepy Hollow* – Burton, 1999). It is not simply the unnaturalness of this situation which is unnerving, but the impossibility of what this suggests (i.e. that a headless body can to all other intents and purposes be ‘alive’ and possess features common to a ‘normal’ body – e.g. intent, awareness, etc.). Ultimately, I will argue, the reason for this disquiet is the central role played by the capacity for conscious thought and volition in our concept of a human being. For instance, in Bram Stoker’s *Dracula* (1993), it is the presence of consciousness and purpose together with the absence of other conventional signs of life which completely unnerves the human characters presented with the ‘undead’ vampire.

the *person* still in some way exists).³⁴ So, whilst Gervais's point may not be decisive against Becker's views for the reasons given, the hypothetical situation does highlight that the reason we may find Becker's concept of integrated organic functioning of the organism plausible is that it is a necessary condition for the those aspects of a human being which we also find essential (e.g. consciousness, responsiveness, etc). It is only when this is made explicit to us by the decapitation example that these conceptual assumptions become apparent, and we begin to find Becker's solely biological account more difficult to accept. I am not here suggesting that such assumptions are correct, or even that they are coherent (I will leave such questions for the moment).³⁵ However, what I am suggesting is that it is difficult to see how the concept of death can ever be completely disentangled from such extra-biological views of the person. This point, which I shall explore in greater depth later on, therefore suggests that not only is a strictly-biological definition of death difficult to imagine, but it is also at odds with certain common concepts of life and death.³⁶

I will return to Becker later where I will summarize what I think are the more fundamental underlying issues with the strictly-biological approach. However, in the above discussion, I have tried not only to show that Becker's position is internally incoherent (as in his use of the idea of the 'reversibility' of death), but also to suggest that an organismic definition of death can – under certain conditions – be made to ignore key aspects of the concept of a human being; furthermore, that 'human organism' and 'human being' are not coterminous (the latter, in my meaning, having a broader, extra-biological aspect). However, before moving on to more general discussion, I will now look at three other biological approaches, all of which centre around the functional capacity of a single organ: the brain. Brain-centred approaches may be divided into three sub-categories: brain-stem, whole-brain, and higher-brain (neocortical) criteria. Discussion of the latter, since it is closely linked with arguments to do with personal identity and consciousness, I shall keep to the later chapters which are dedicated to these topics. Similarly, as regards whole-brain death criteria, we are faced with the question of whether positing the absence of total brain function is an overly conservative criterion for determining death.

³⁴ Some advocates of the cardio-pulmonary criteria reject the idea that such a decapitated body could still be considered 'alive' on the grounds that the heart is still beating, because such heartbeat as persists would not be 'seriously persistent', and its 'normally integrated environment would be disrupted' (Evans 1994, p.7). Evans's point here, of course, is that the cardio-pulmonary case is not undermined by the decapitation scenario.

³⁵ Ultimately, in later chapters, I will be very interested in whether these 'assumptions' can be cashed out rationally, and what status they in fact possess.

³⁶ It should be noted that in referring to common concepts or notions I am not seeking to accord such things any special or privileged status. Furthermore, when I talk of 'the common concept of death' or 'competing notions as to what a person is', I am not assuming that these are clearly defined concepts (though, of course, sometimes they are – as when they form part of a philosophical theory), or that there are a finite number of them, but merely that in the debate concerning the definition of death, the conflict should be seen as stemming from the application of competing notions (whatever they are). Obviously, it is the job of philosophical analysis to clarify such concepts, whether to show that they are confused or in some way borne out. So, in defining death, the role of philosophy must be to determine to what extent common concepts of what is and isn't a living human being can be in some way justified. It will be my argument, however, that such justification – whether in terms of biological criteria or conceptual analysis (as to what is meant by 'death', 'human being', etc.) – cannot ultimately resolve the issue. I shall return to this in more detail in the latter sections of this chapter.

The report of the Harvard Medical School's Ad Hoc Committee in 1968³⁷ first began the shift away from heart and lung centred criteria for the determination of death, and this was followed by the report to the President's Commission in 1981,³⁸ which helped establish these criteria as standard practice in the USA. However, many commentators³⁹ have argued that this represents an unnecessarily conservative standpoint, and that – without necessarily resorting to personal-identity- or consciousness-based arguments – more parsimonious criteria are a more legitimate and accurate means of determining death. Furthermore, as Peter Black notes, the whole-brain death approach contains a wide range of criteria, it being variously argued that it is 'enough to describe a state from which survival of bodily function has not been seen', 'it is necessary to predict widespread severe brain necrosis on post-mortem examination', or that there must be a flat EEG or 'an arteriogram that shows no intracranial filling'.⁴⁰ Clinical precision and medical need alone would therefore seem to argue for a refinement of brain-death criteria.

However, the move away from whole-brain centred criteria should not necessarily in itself be seen as an advance in technological knowledge and understanding, and there are those who propose retaining the whole-brain death approach based on tutiorist principles (i.e. if in doubt, take the safer course). For instance, in Walton (1987), we find:

In whole-brain death, therefore, the highest standards of probability are satisfied that there is no possibility of return to a level of function compatible with any form of awareness.⁴¹

Moreover tutiorist concerns still persist in some quarters as to the adequacy and certainty of more parsimonious criteria,⁴² so we may still argue that this debate is not yet medically resolved.

However, aside from these worries, the debate reveals similar issues to those already raised by Jonas and Evans: science cannot necessarily adjudicate on socially-embedded concepts such as personhood and quality of life, and these in turn will decide what the respective criteria *mean*. So, in asking the question, "Is whole-brain death too conservative a criterion for death itself?" we need to consider whether a more parsimonious criterion (e.g. brain-stem death) can act as a *sine qua non* for the *totality* of the range of related concepts as to what constitutes a living human being.

I will therefore consider the question of the adequacy of whole-brain death criterion by consideration of its more slim-line offspring: brain-stem death.

Brain-Stem Death: Pallis and Lamb and Functional Integration

In considering brain-stem death, I shall look at two of its main proponents: Christopher Pallis and David Lamb.⁴³

³⁷ Beecher (1968).

³⁸ Lynn (1981).

³⁹ E.g. Puccetti (1976), Pallis (1983), Veatch (1975), to note some early opponents.

⁴⁰ Black (1978a), p.338.

⁴¹ Walton (1987), p.99.

⁴² See Russell (2000), p.28, and Evans (1994), p.3.

Pallis's position, in his own words, 'seeks to combine philosophical and physiological considerations.'⁴⁴ However, the locus of these considerations is the function of the brain-stem:

The loss of the capacity for consciousness and of the capacity to breathe relate to functional disturbances at opposite ends of the brain stem, while the former is also a meaningful alternative to 'the departure of the soul'.⁴⁵

For Pallis, then, the death of the brain stem accounts for the permanent cessation of the dual aspects of the living human being: the 'departure of the soul' (in Pallis's modern sense of the capacity for consciousness and related concepts of psychological personal identity), and the loss of functional integration of the organism ('cessation of the function of the "organism as a whole"' as Pallis calls it).⁴⁶

When the human organism has irreversibly lost the crucial capacity for consciousness and the ability to breathe (and thereby to maintain a spontaneous heart beat), and when moreover it has lost such important responses to its environment as the homeostatic maintenance of temperature and blood pressure (which are mediated through or by the brain stem), in which sense can it be said to be an independent biological unit?⁴⁷

Brain-stem function, then, for Pallis, is the prerequisite of a wider concept of the organism than is found in other definitions. In as much as a functioning brain stem is a condition of consciousness, I will deal with the consequences of this later (where the role of consciousness itself as a criterion is considered). However, in relation to the other consequence of cessation of brain-stem function – the loss of functional integration – this is a concern which he shares with David Lamb. Therefore, I shall now turn to an analysis of Lamb's views, considering the conclusions reached there as applicable to this aspect of Pallis's criterion also.

Whilst also favouring a brain-stem approach, David Lamb is at pains to establish a strictly-biological definition of death. Lamb argues, in line with Becker, that the idea that death 'must be socially or legally defined by morally relevant characteristics [...] would [...] be dangerous' because 'appeals to social, religious, and moral criteria come both from advocates and adversaries of euthanasia' and 'the issue of defining death should not be confused with the quality of residual life or with the decision on when, if ever, lives should be terminated'.⁴⁸ Furthermore, Lamb states that any appeal to personal identity is inappropriate, and that any such attempt would create a

⁴³ There are many others who share this view, but for my purposes Pallis and Lamb, as two of the earliest advocates, represent the range of positions. Others, such as Russell (2000) and Gervais (1986), use brain-stem death in conjunction with (or as a basis for) other criteria, so I discuss these separately.

⁴⁴ Pallis (1985), p.33.

⁴⁵ Ibid.

⁴⁶ Ibid., p.34.

⁴⁷ Ibid.

⁴⁸ Lamb (1978), p.145.

conceptual rift with ‘one kind of death for humans and another kind for other life forms.’⁴⁹

However, Lamb’s main point of disagreement with Becker is over his attitude to brain-stem death. For Becker, the death of the brain stem is not – because of the development of artificial means of maintaining heartbeat and respiration – equivalent to the death of the organism; the brain is just another organ with a particular functional role in the organism as a whole. It is only the irreversible demise of the organism itself which constitutes death. However, Lamb considers this to be an avoidance – or ignorance – of the facts. He argues that, since ‘the brain-stem, not the heart, is recognised as the specific area which regulates all vital processes’,⁵⁰ death of the brain stem means that:

The heart and other organs can never again function naturally. At most the heart and lungs can be mechanically operated for two weeks. But during this period one would be merely ventilating a corpse. There are no reversals of brain-stem death.⁵¹

Lamb also criticizes Becker’s holistic approach by arguing that in defining death as a process with no distinct endpoint, Becker’s view ultimately entails a vague and problematic definition of death. Lamb considers this an unacceptable consequence of the traditional cardio-pulmonary definition.

Confusion over the “reversibility of death” reveals the limits of an exclusive reliance upon traditional criteria. Faced with the possibility of unlimited resuscitation, Becker is committed to talk of death as a reversible event.⁵²

As has been already pointed out, this is a problem for Becker: either someone who is revived from the brink was not really dead in the first place, or the ‘being/has-been’ boundary is a vague and shifting one. Becker seems caught between the two positions: he doesn’t want death to be defined in social, moral, or legal terms; and yet, he doesn’t want to imply that discontinuation of medical efforts is always morally culpable. The latter, certainly, is a genuine concern, as it involves the sort of extra-biological factors that Becker and Lamb wish to avoid (whether you are dead or not should not depend, for instance, on arguments to do with who most deserves the ventilator). However, given a definition of death as a process which allows for occasional ‘reversals’, this seems unavoidable.

(But why, it may be asked, is the notion of *irreversibility* so important? Aside from the already noted concern that extra-biological (e.g. moral) factors should not play a part in determining when someone is dead, it might be argued that there is nothing

⁴⁹ Lamb (1985), p. 93. I will discuss the personal identity approach to the problem later where, in discussion of the view’s proponents (such as Gervais), I consider this objection of Lamb (for, conversely, it may be that, given the widely differing characteristics of living organisms, the concepts of ‘being alive’ and ‘individual identity’ that we utilize for humans may not be easily applicable to other species).

⁵⁰ Lamb (1978), p.146.

⁵¹ Ibid., p.147.

⁵² Ibid., p.151.

wrong with the general notion of the reversibility of death in *strictly-biological* terms. So, a person may be in a state of death (just as they are in a coma), and they may recover. This is not a point that Lamb directly addresses, but we might argue that irreversibility is central to the *concept* of death, and that – without it – death is not death at all. Answering this question fully, however, involves a conceptual analysis of death (something that I shall undertake in a later chapter), so I shall not address the issue here. However, for the moment, and in sympathy with Lamb, we might question how exactly a concept of death as reversible could be distinguished from other serious illnesses. In light of this, the inscription on comedian Spike Milligan’s gravestone – ‘Told you I was ill’ – takes on a whole new meaning! Lamb’s concern, therefore, is to maintain the binary and mutually exclusive relationship between the concepts of life and death – though why he (or anyone) should want to do this, I shall examine in detail later on.)

Lamb seeks to avoid the reversibility problem by attempting to define death as a distinct event: that of irreversible brain-stem death.

The brainstem contains (in its upper part) crucial centres responsible for generating the capacity for consciousness. In its lower part it contains the respiratory centre. It is death of the brainstem [...] which produces the crucial signs (apnoeic coma [lack of capacity to breathe spontaneously]) which doctors detect at the bedside, when they diagnose brain death.⁵³

So, for Lamb, what he terms ‘systematic death’ (the definition favoured by Becker) is only ever ‘a mechanism for causing brain death and it is only lethal if it lasts long enough for certain critical areas of the brain to die.’⁵⁴ Therefore, even if systematic death occurs, the ultimate cause of death is damage to that part of the brain which ultimately controls the systematic process. In other words, aspects of the system may fail – such as heartbeat and respiration – but may be restarted. However, if the brain stem is damaged, then those processes can never be restarted so that spontaneous function is resumed. Thus, for Lamb, artificial respiration of the brain-dead is literally ‘ventilating a corpse’.⁵⁵

However, it is not solely the inability to resume spontaneous function that Lamb considers a sign of death. Firstly, Lamb argues that the brain stem is the ‘critical system’ of the organism. As such, it is responsible for the integration and functioning of the sub-systems and organs which constitute the organism. Hence:

Given that life is essentially a matter of organisation, the moment of death is not the cessation of breathing and circulation but when breathing and circulation lack neurological integration.⁵⁶

Thus the cessation of spontaneous respiration and heartbeat only entail death when the critical system that controls and regulates them (the brain stem) can no longer function. For Lamb, this is true even when respiration and heartbeat are maintained by artificial means. This is because when the brain stem dies

⁵³ Lamb (1985), p. 5.

⁵⁴ Ibid., p.4.

⁵⁵ Lamb (1978), p.147.

⁵⁶ Lamb (1985), p.26.

a point has been reached where the various subsystems lack neurological integration and their continued (artificial) functioning only mimics integrated life. [...] The death of the brain is the point beyond which other systems cannot survive with, or without, mechanical support.⁵⁷

This is an important point, and one which we shall return to shortly. In current medical experience, death of the brain stem soon results in significant functional disintegration of the organism as a whole – even where life support mechanisms are in use. Thus for Lamb, the idea that the function of the brain stem can be substituted by mechanical life-support is fundamentally wrong, for these aids do not in fact replace the function of the brain stem at all.

what they actually substitute for are the functions of the intercostal muscles and diaphragm, which without neuronal drive from the brainstem cannot function spontaneously⁵⁸

It is as if, Lamb implies, we had picked up a dead body and made lifelike motions with its arms and legs; what we have is a semblance or mimicry of life, and not the genuine article. The other important assertion that Lamb makes, therefore, is that the function of the brain stem is irreplaceable, and consequently, not only does its demise represent the loss of its integrative faculty, but more importantly the *irreversible* death of the organism. This concept of irreversibility is vital for Lamb for two reasons: (1) he wishes to avoid the notion, inherent in Becker's definition, that death is in any way uncertain or may shift according to circumstance (you're either dead or you're not); (2) in order for death to be defined in a strictly-biological sense, death of the brain stem needs to be a sufficient criterion for death *per se*.

(1) is linked to Lamb's desire for clear, medically applicable criteria that can be used to disambiguate the status of those near death. As such, brain-stem death therefore allows us to treat death as a definable event, rather than what Lamb considers to be Becker's somewhat nebulous conception of death as a process. The second reason is linked to Lamb's desire to exclude non-medical criteria from his definition. Brain-stem death *just is* the death of the organism, and he has little patience with any suggestion that this is not the case, or that the functions of the brain stem may one day be replicated by artificial means.

Before moving on to my criticisms of Lamb, it is important to point out that Lamb's equation of death of the organism with brain-stem death means that those patients whose brain-stem function is intact, but whose 'higher' cognitive capacities will never return, are still alive. This includes those in a persistent vegetative state, but more broadly those individuals who – for whatever reason – can never employ such cognitive faculties again. Lamb is quite rightly afraid here that using a criterion of lack of capacity for consciousness (or *neocortical* death) might, via a slippery slope argument, lead to the classification of the mentally deficient as 'dead'.⁵⁹ However, he

⁵⁷ Ibid., p.37.

⁵⁸ Ibid., p.39.

⁵⁹ See, for instance, Alexander (1949), where the author describes the way in which the principle of relieving suffering was exploited by the Nazis for their own political purposes. He then considers the possibility that a similar change in emphasis away from amelioration and towards rehabilitation (where

also correctly points out that the capacity for consciousness itself cannot be isolated in the 'upper' brain (or cerebral cortex), and that key components for awareness (such as the reticular activating system) are actually situated in the brain stem itself. On the other hand, Lamb rejects 'whole-brain death' as essentially too conservative, in that if brain-stem death is indeed a sufficient criterion (which he thinks it is), then such caution is unnecessary. In other words, why wait for the whole brain to die, when the brain's significant feature (as far as being alive is concerned) is the functional integration of the organism supplied by a working brain stem. Lamb's position is therefore, in a way, a refinement of Becker's position: using brain-stem death as a criterion for death *per se* allows us not only to treat death as a definite event (as opposed to a process), but also to pinpoint the *sine qua non* of true functional integration (the brain stem).

The main point at which I wish to take issue with Lamb is the idea of the irreplaceability of the brain stem – which, incidentally, is a position he would seem to share with Pallis.⁶⁰ This concept is fundamental to Lamb's definition in that it allows him to argue that there is no conflict between the cardio-pulmonary method of determining death and his own brain-stem death criterion. The latter, he simply states, is a definition 'where none previously existed'.

That is to say, in a very important sense, the traditional cardio-respiratory concept never provided adequate criteria for death and, in the light of contemporary knowledge about the mechanics of brain death, irreversible cardio-respiratory arrest was merely an indication that brain death was imminent.⁶¹

All previous indications of death then can be seen as precursors of the true cause. However, it can now be seen that, as true indicators of death, they were simply inadequate – and at times misleading. Lamb is at pains to point out that brain-stem death is not something that it is possible to recover from, and that any purported 'reversals' represent an inadequate application of the criteria (i.e. misdiagnosis), rather than a failure of the criteria themselves. Accordingly, the example that Becker uses of the boy who revives after being submerged in icy water is rightly criticized by Lamb on these grounds: it is well documented, he points out, for such cases (i.e.

the 'nonrehabilitable sick' (p.44) are marginalized, becoming 'second-class patient[s]' (p.45)) might lead to a similar situation developing in America. Alexander's point is that it is this change in emphasis itself which must be resisted; once this point is passed (as it was in Germany), then abuses of the quality of life argument become both harder to resist and more likely. A parallel contemporary situation is alleged by Wang (2002), which is a testimony given before the Subcommittee on International Operations and Human Rights of the U.S. House of Representatives. Wang is a former doctor at a Chinese army hospital (now seeking asylum), and details his experiences in removing body parts for transplant from executed prisoners, some of which had their executions 'botched' so that they were still technically alive when certain organs were removed so as to ensure that they were viable. Many of these prisoners were executed for crimes as minor as tax evasion, and the process of organ procurement seems to have been driven by bribes to officials, and a thriving black-market. Typically – in the case of kidney transplants, for instance – the cost of procuring organs was expensive, and only affordable by high-ranking officials.

⁶⁰ See Pallis (1985), p.33: 'Whereas the functions of lungs and heart can (for a while) be taken over by a machine, those of the brain cannot.' However, it is unclear whether Pallis considers this to be a conceptual impossibility as Lamb does.

⁶¹ Lamb (1985), p.19.

hypothermia) to eventually result in revival, and thus it would simply be medically wrong to consider this a case of 'reversal'.

Having established, to his own satisfaction, the irreplaceability of the brain stem, and the subsequent irreversibility of death according to this criterion, Lamb proceeds to respond to certain key objections. The one I want to concentrate on here concerns the question of the replaceability of the function of the brain stem; Lamb considers this to be empirically impossible, and yet it is surely not his place to say so. As Green and Wikler point out, in criticism of Lamb:

development of a more perfect mechanical substitute is merely a technological problem.⁶²

Furthermore, as Byrne *et al.* observe:

any declaration that a cessation of function is absolutely irreversible is a presumption, even if well-grounded, which is contingent upon the current state of medical knowledge and on the availability of adequate life-support systems in concrete circumstances.⁶³

These seem like coherent objections, and ones which undermine one of the central tenets underlying Lamb's definition. For, if it is at least *theoretically* possible to replace the functions of the brain stem, then death of the brain stem itself only provides a 'reversible' definition of death – something which Lamb was determined to avoid.

Lamb, however, rejects the very notion that the brain stem may be replaceable, considering it 'a conceptual and practical impossibility.'⁶⁴

At the current level of scientific thinking, brain transplantation does not even amount to a logical possibility since its very plausibility would entail a radical departure from accepted canons of plausibility and possibility.⁶⁵

And furthermore:

The point is that, given existing concepts of living organisms, skills, theories and resources, this is one technological problem for which no solution can be imagined in the foreseeable future. To use an expression of Wittgenstein's: if that were possible then anything could happen.⁶⁶

It is tempting to feel sympathy for Lamb here. As he points out, philosophical literature abounds with examples imagined or drawn from science fiction where such things such as brain transplants, molecule transporter technology, and doppelgangers, are all utilized to explore the limits of logical possibility. However, Lamb argues,

⁶² Green and Wikler (1980), p.113 (quoted in Lamb 1985, p.38).

⁶³ Byrne, O'Reilly and Quay, (1985-90), p. 1988.

⁶⁴ Lamb (1985), p.61.

⁶⁵ Ibid.

⁶⁶ Ibid., pp.61-2.

whereas this may give us insight into the notion of personal identity, such a technique has its limits, and its use as a get-out clause in empirical matters becomes both tedious and artificial.

However, it is equally hard to ignore the point that Lamb's critics (myself included) are making here. Lamb's assertion is an empirical one, based on contingencies concerning the current state of scientific and technological progress. His view that such procedures as imagined in science-fiction scenarios do 'not even amount to a logical possibility' jar somewhat with the concessions that the limitations we are dealing with here are 'existing concepts' and the 'current level of scientific thinking'. So, even if we take Lamb's point, this does not elevate the status of his assertion that the brain stem is irreplaceable out of the realm of the merely contingent (that is, unless Lamb wants to do away with the distinction between logical and contingent truths – which I assume he does not, merely wishing to limit the role of the former in undermining his argument).

However, this question aside, there might in fact be a reason to doubt the assertion that the artificial replacement of brain-stem function is currently wholly beyond our technological means. As Tom Russell observes:

Ventilation can be maintained in an individual with a high cervical cord transection (which has resulted in paralysis of breathing) by stimulating the nerves that normally carry the impulses needed for spontaneous respiration by means of an implanted nerve stimulator; the patient so treated appears to be breathing spontaneously. There is no technological reason why this technique could not be used in patients in whom the loss of spontaneous respiration was a result of brainstem dysfunction.⁶⁷

In addition to this, Russell also makes the observation that clinical evidence suggests that diagnosis of brain-stem death (according to present UK criteria) does not in fact entail the death of the totality of the brain stem, and the pituitary gland and other areas outside the brain stem continue to function.⁶⁸ So, it would seem, potential technological advances aside, it may even be *currently* debateable as to whether brain-stem death is as sufficient a criterion as Lamb thinks it is.

Finally, it may also be argued⁶⁹ that Lamb's definition, like Becker's, leaves out the relevance of consciousness for a definition of death. If the brain stem can be shown to be theoretically replaceable (or its dysfunction partially remedied, as Russell argues), then it ceases to be the critical system that Lamb considers it to be, since its role can be filled by artificial means. In such a case, we are left with the possibility that, as mentioned earlier, a decapitated individual whose systematic functioning was artificially maintained would still be 'alive'. As Gervais points out, the role of the higher cognitive functions are ignored, the emphasis instead resting on the integrated systematic functioning of the organism as a whole – albeit, in this case, a headless one. On the other hand, however, if we take Lamb's position to include the cessation of consciousness (in as much as a functional brain stem plays a key role in the capacity

⁶⁷ Russell (2000), p.21.

⁶⁸ Ibid., p.28. This point is also made by Evans (1994), p.3, who cites Evans and Hill (1989).

⁶⁹ As does Gervais (1986), pp.73-4.

for conscious awareness), then his criterion ceases (arguably) to be strictly-biological (a point I shall return to in chapter 5).

Tom Russell and Homeostasis

The final position I shall consider in this chapter is that of Tom Russell. In *Brain Death: Philosophical Concepts and Problems* (2000), Russell argues that the key feature of a living being is homeostasis (the ability to maintain metabolic equilibrium). So, basing his position on a definition of life proposed by the physicist Erwin Schrödinger,⁷⁰ Russell argues that the absence of homeostasis is a sufficient criterion for death because the organism's incapacity to maintain metabolic equilibrium results in the death of the organism as a functioning whole (in physics terms, this may be thought of as a state of maximum entropy, where low entropy – where the organism has 'a higher energy content than the environment' – is the optimum state). In biological terms:

It is necessary to maintain the physical and chemical properties of the fluids in which the cells lie within a narrow range in order to ensure the continued function of those cells [...] This maintenance of the "internal milieu" of the organism is referred to in humans (and other animals) as homeostasis.⁷¹

Furthermore, as a consequence of adopting this definition, Russell is able to propose a uniform definition of life (and consequently death), and to avoid the twofold definition necessitated by a position that values a solely human property (such as higher cognitive function or personhood). Russell defends this move on three counts:⁷²

- The twofold distinction between human and non-human species is not justified morally as it affords humans special treatment (i.e. it is the basis for speciesism).
- The neocortical approach excludes certain human individuals (e.g. 'anencephalic new-borns, new-borns generally, the comatose, those in the persistent vegetative state, foetuses and the severely mentally disabled.'⁷³
- The neocortical approach is a value-based distinction, and not one based on an understanding of organic function (which, Russell argues, the definition of death should be). Consequently, Russell quotes Lamb approvingly: 'A human being without identity is just as alive as any other living human being.'⁷⁴

Therefore, in rejecting a twofold definition (and with it, the neocortical approach to human death), Russell is attempting to propose a strictly-biological definition of human death which can also apply to animals and other creatures. Obviously, due to the variance in biological features among humans and animals, the physiological reasons for the maintenance (and cessation) of homeostasis will be different. However, in terms of human death, what then is the biological locus of homeostasis?

⁷⁰ Schrödinger (1944).

⁷¹ Russell (2000), p.92, but see pp.90-2 for full explanation.

⁷² Ibid., p.103-4.

⁷³ Ibid., p.104.

⁷⁴ Lamb (1985), quoted in Russell (2000), p.104.

The question now becomes one of establishing where this capacity for homeostasis resides in an intact human being. This is a relatively straightforward question to answer inasmuch as physiologists have demonstrated the fact that the hypothalamus in the human brain is responsible for the control of homeostasis, and that damage to a variety of areas within the hypothalamus results in imperfect control or loss of control of some of the facets of homeostasis. From such evidence, it is reasonable to deduce that the capacity for homeostasis resides in the hypothalamus. It follows from the preceding argument that permanent loss of function of the hypothalamus would result in permanent loss of homeostasis.⁷⁵

Obviously, therefore, whole-brain death would certainly result in permanent cessation of homeostasis, but what about brain-stem death, or some less conservative criterion? The problem, Russell points out, is that

it is simply not possible with current investigative technology to detect the permanent loss of hypothalamic function prior to the occurrence of whole brain death as measured by current standards and tests. Whilst, therefore, I have argued that permanent loss of hypothalamic function is a sufficient condition for death, in practice I must accept that whole brain death is a sufficient condition for death if death is accepted as the permanent loss of control of homeostasis.⁷⁶

The hypothalamus is a comparatively small area that sits below the thalamus in the mid-brain (and, as such, is distinct from the brain-stem). As such, its function is difficult to monitor or test, and its role in the *diagnosis* of death – whilst potentially theoretically sound (if we accept Russell’s arguments) – is currently practically unworkable. Russell’s problem, then, is a practical one: in practical terms, he must accept whole-brain death as a criterion; in theoretical terms, cessation of hypothalamic function is sufficient. Therefore, whilst whole-brain death is, as he says, ‘over-sufficient and not specific’,⁷⁷ it will have to do.

In analysing Russell’s position, however, I am not primarily concerned with whether homeostasis is a sufficient criterion for the presence of life, or even whether the hypothalamus is in fact the centre for homeostatic control. Both, to a certain extent, are biological concerns, and are therefore to be argued against on those grounds. However, what I do make objection to is his assertion that homeostasis is the essentially defining feature of the living organism, and that therefore the homeostatic functioning of the organism is the defining feature of a living human being; the former, in as far as it refers to the ‘organism’, may be fine; the latter, however, posits this criterion in relation to the ‘human being’, and as such, I shall argue, is a value-driven move.

As I hope is clear from my criticism of the biological view so far, the main problem that arises in strictly-biological approaches is that they seek to determine an essential

⁷⁵ Russell (2000), p.141.

⁷⁶ Ibid., p.142.

⁷⁷ Ibid., p.141.

feature of a human being. However, this, in itself, is a value-laden enterprise – and, as such, is open to contradiction by opposing or differing sets of values. In the next chapter, I shall argue that the problems surrounding the definition of death arise largely from difficulties in isolating the essential features of a concept. Accordingly, in Russell’s case, he wants to argue that homeostasis is *essential* to the concept of human life.

I would further argue that, since human beings and other animals are biological systems and will therefore cease to exist (*i.e.* die), then it is at least logical that this death should be expressed in biological terms. Since death is common to human beings and animals, it would follow that a function, the loss of which would uniquely constitute death, must be common to human beings and animals.⁷⁸

However, there is a fallacy here. This is the form of the argument:

- Human beings and animals are biological systems.
- The death of biological systems must have something in common.
- Therefore, animal and human death has something in common (*i.e.* cessation of homeostasis).

The problem lies in the first premise: even if, in a sense, human beings *are* just biological systems, it does not necessarily entail that their death merely consists in the death of the organism. After all, without being in any way mystical or dualistic, there is an important difference between animals and humans, and that is our capacity for cognitive consciousness. It is this capacity, which facilitates language and conceptual thought, which in turn gives rise to notions of personal identity and selfhood, social practices and human values (from which the concepts of ‘death’, ‘self’, etc. – I shall argue – cannot be wholly divorced). So, to make Russell’s first premise explicit, it should read:

- Human beings and animals are *solely* biological systems.

In other words, that is all a human being is: a biological system. And yet, isn’t it obvious that we are *more* than that? It is these very features which give rise to notions of value and the quality of life (and which lead such as Puccetti to talk of ‘breathing corpses’). It is because of the fact that a human being is so much more than a collection of organs maintained in homeostasis that the very problem of defining death has arisen. And yet, whilst Russell argues that ‘religious and philosophical concepts of death are flawed’, and which in turn leads him to propose a biological approach as the only one which is ‘workable’,⁷⁹ this would seem to entail an essentialist move which is neither wholly justified nor inescapable.

It should be noted here that, in one sense (and from one point of view), human beings *are* only biological systems; we are made of biological parts, and (some will argue) nothing more. However, even if this is true, it does not mean that what a human being *is* (their significance) can be reductively explained by reference to the biological

⁷⁸ Ibid., p.115.

⁷⁹ Ibid., p.119.

organism. My point is therefore that, whilst ‘human organism’ may account, in one sense, for the totality of what a human being is, in another sense it leaves out what is significant about that organism (i.e. that it enables us to be conscious, to socially interact, to have values, etc.). This can easily be illustrated by scenarios where biological function is absent or in some other way compromised: would it matter that this were the case if (for instance) consciousness was maintained by some other means? If synthetic bodies were possible, would we consider such individuals ‘dead’? I think not. Nor, conversely (as noted earlier, in relation to Lamb) need we feel compelled to value a headless body whose functional integration (lower brain) were maintained artificially.

In conclusion, then, Russell’s position would seem to entail similar problems to others already considered. So, whilst he cogently argues – for example – that consciousness-based approaches leave us with all sorts of practical and theoretical difficulties (e.g. as to when permanent unconsciousness is reached, what we mean by consciousness, etc.),⁸⁰ this is not in itself a conclusive reason for adopting a strictly-biological approach. Furthermore, Russell’s essentialist assertion as to the nature of human beings is effectively weak – especially when it is realised that the concept ‘human being’, whilst it *does* certainly admit of a strictly-biological interpretation (in that – it may be argued – we are merely biological machines), is also one that may be interpreted according to a much wider set of extra-biological criteria.

Summary of Problems with the Strictly-Biological Definition

Let us now look at the general problems with this type of approach. Given the above objections to the respective views considered, it would seem that these strictly-biological definitions fail because they attempt a purely organismic definition of death.⁸¹ Furthermore, this failure shows that a strictly-biological concept of death – even if it were possible – would exclude key aspects of our notion of what a human being is, and therefore of what his or her death entails. This implies, arguably, not only that a strictly-biological definition will necessarily leave something out, but also that the definition of death must necessarily include what Gervais terms ‘a decision of significance’,

that is, a decision that there is a certain feature (or cluster of features) whose permanent absence constitutes the death of the person.⁸²

In the case of biological criteria, the significant features include the lack of spontaneous respiration and heartbeat, lack of responsivity, indications of brain-stem damage, etc. More importantly, perhaps, it also includes the ability to be conscious. As Gervais argues:

Since we do not declare death when the lower brain can no longer support respiration and circulation, but only when the upper brain is also permanently destroyed, the permanent absence of consciousness has an

⁸⁰ Ibid., p.118. I consider some of these issues in chapter 5.

⁸¹ Aside, of course, from Pallis’s use of brain-stem death as the indication of the cessation of the capacity for consciousness and therefore the ‘departure of the soul’ (so-called), which I shall return to later.

⁸² Gervais (1986), p.2.

undeniable and crucial bearing on the decision to declare death. If the source of organismic functioning per se is of no consequence in a decision that someone is dead or alive – as the use of the brain-death criterion implies – then consciousness has become, with the adoption of that criterion, the *sine qua non* of human life.⁸³

Gervais uses this approach to argue for a neocortical, personal-identity-based definition of death. However, because this forms the basis of an argument which utilizes notions of personal identity and the capacity for consciousness, I shall leave treatment of these issues to later on where I will discuss them in detail. It is sufficient for the moment, then, to note that it is one criticism of the strictly-biological definition that, in focusing solely on biological criteria, certain significant features which we associate with a human being – such as the capacity for consciousness – may be omitted, leaving us with the seemingly absurd consequence that an individual can still be thought of as alive who completely lacks the potential for consciousness of any sort (e.g. is headless). This apparent absurdity leads us in turn to question *which* indications of death are significant and *why* (i.e. what the absurdity consists of, and whether it may be considered a rationally grounded reaction).⁸⁴ However, as this suggests, such indications are not significant *in themselves*, but are rather imbued with significance in relation to the clutch of values and directives that inform those common notions which are being employed.

Finally, to briefly return to the question of whole-brain death criteria (as promised earlier), whilst it can be seen from the above that the use of brain-stem death criteria is problematic, this does not necessarily mean that a tutorialist return to a wider set of criteria (such as whole-brain death) necessarily solves the overall problem. As I have argued, not only are biological criteria merely a part of our wider concepts of what is meant by ‘living’ and ‘dead’, but the adjudication as to how these terms are applied is not solely medical science’s to make. There are, as Gervais, Jonas, and Evans observe (in different ways), good reasons why we cannot ignore the role that the culturally derived aspects of the concepts of life and death cannot be ignored, and which an appeal to death criteria must therefore account for. In this sense, Pallis’s other use of brain-stem death to indicate ‘departure of the soul’ will have to wait until a later point.

Conclusion

In this chapter I have distinguished between two competing approaches to the definition of death (only the first of which has so far been explored): what I have termed the ‘strictly-biological view’ (following Gervais), which attempts to define death by reference to purely biological criteria; and what I have termed the ‘partly-biological view’, which considers biological facts to be merely relevant (i.e. not in

⁸³ Ibid., p.44.

⁸⁴ Actually, I think Gervais is wrong here to criticize Lamb for denying the role of consciousness altogether. As Lamb points out, the ARAS (Ascending Reticular Activating System) plays a key role in consciousness, yet is situated in the brain stem. Thus, to say that citing a functioning brain stem as the key indicator of life is to exclude consciousness is not completely correct; but nor is it correct, as Gervais suggests, to single out the neocortex as the sole possessor of our ‘higher’ functions. The truth is that the two parts give us different things, neither of which we would want to do without: the neocortex houses the higher cognitive functions, whilst the brain stem houses – among other things – the ARAS, which allows us – among other things – to ‘wake up’ and switch our conscious attention. I will examine these arguments in more detail when I look at consciousness and personal identity later on.

themselves sufficient) to the definition of death, which must necessarily take into account other, extra-biological features of the human being (e.g. their moral status, the notion of personal identity, etc.). The former view, I have argued, necessarily leads to conceptual absurdities (such as the headless but living human example), or else provides an insufficiently clear account of death, so that it becomes 'reversible' (Becker). This latter position, whilst not perhaps objectionable in itself, leads into deeper questions concerning the concept of death (i.e. assumptions concerning its binary opposition with life, the need for finality and irreversibility that some see as essential to it), and I will address these questions in the following chapters.

However, in conclusion, I consider the strictly-biological view to be necessarily flawed. The reasons for this, I argue, are that not only does such an approach necessarily ignore certain key features of the human person (such as personal identity, the capacity for consciousness, etc.), but it is ultimately an attempt to define certain strictly-biological criteria as constituting the essence of a human being. In this sense, it also reveals an ignorance of the *type* of concept that death is. Death must, I shall argue, be considered in a manner which sees biological facts as relevant, but not in themselves sufficient for defining what death is (i.e. the partly-biological view). It is this view that I will now explore in the following and subsequent chapters.

3. Death and the Problem of Categorization

Introduction

So far, I have outlined a number of attempts at defining death from what may be termed a 'strictly-biological' perspective, together with criticisms of these definitions which suggest that the general approach is at best problematic. However, in this chapter I want to give more fundamental reasons as to why I think the strictly-biological approach fails *in general*, and that the problem faced in clearly defining death is not necessarily eliminable via conceptual analysis (of what a 'human being' is, what we mean by 'death', etc.), or resolvable through advances in scientific knowledge. In doing this, I shall treat the definition of death as a problem of categorization in as much as death is a state which we would differentiate from the state of being alive; thus, these two states are categories with (it is commonly thought) mutually exclusive membership, and in defining death we are primarily concerned with making rationally explicit those essential features which determine exclusive membership of one category or another. This is not to deny, of course, that many different types of factor influence the decision – moral concerns, metaphysical notions, practical considerations, and so on – but I merely point out that the ultimate point of the discussion of these aspects of the issue is to arrive at a conclusive *disambiguation* of problematic cases in favour of one category or another (i.e. life or death). As such, I aim here to consider the problem of category membership as a pivotal issue for *all* of these debates. Furthermore, if it can be shown that there is a fundamental problem with the very notion of distinct category membership (at least in relation to the concept of death), then it will aid our general understanding of the debate as a whole.

I will begin with the observation that, in human history, category formation is initially motivated by cultural concerns. As an anthropological claim, this is not a contention that I can defend with any rigour, but its plausibility provides a possible context for the sort of logical problems that I argue are at the centre of the categorization process, and therefore at the centre of the definition of death debate. I then go on to consider the scientific process whereby we progress from intuitive notions of category membership ('intuitive kinds') to more scientifically robust notions ('theoretical kinds'). In relation to this process, I then consider examples of category conflict (such as we are faced with in the definition of death), and try to show how such conflicts in fact stem from deeper metaphysical commitments to do with the nature of universals (specifically, the rationalist assertion that certain concepts – such as death – are immutable, and are defined by essential features). I then proceed to attack this essentialist position (and variations thereof), before outlining an alternative conception of the nature of category membership and its application to the definition of death.

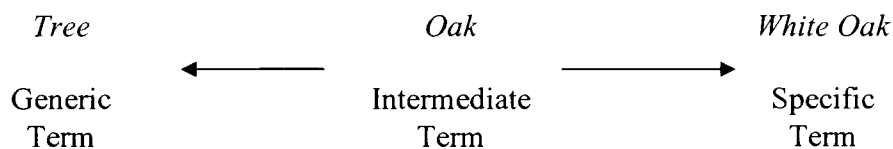
False Beginnings

Categorization is a practice which it would seem is ubiquitous for its utility, allowing us to group and distinguish things in terms of their possession of common features. Obviously, the sort of criteria employed in this process will differ according to the purpose of the grouping: for example, an anthropologist may find that particular tribes class certain plants together because they are edible, because of their location, for

their use in rituals or medicine, or for other reasons. However, overridingly, research among tribal peoples has shown that ‘organisms are grouped into named classes primarily on the basis of overall perceptual similarities.’¹ This suggestion, which seems ‘obvious (to biologists) but somewhat surprising (to cultural anthropologists),’² perhaps goes some way in accounting for the remarkable similarity in prescientific classification of natural phenomena in cultures throughout the world. For instance, regarding the plant categorization employed by peoples in geographical locations as diverse as Mexico and China, it has been found that the

broad taxa [employed] are reminiscent of classical Greek and Latin plant classification (tree, vine, herb). These three major groupings, it might be speculated, represent such distinct perceptual discontinuities that their recognition may constitute a substantive near-universal in prescientific man’s view of the plant world.³

Underlying such practices, then, would seem to be the idea that there are ‘natural kinds’ of things – in other words, that the world consists of organisms and other natural phenomena, each type of which may be distinguished from its neighbour by virtue of the possession of distinct characteristics or features. Such naïve ‘folk biology’, as it may be called, would therefore seem to be the foundation upon which later categorical sophistications – of which Aristotle’s concepts of ‘genus’ and ‘species’ would be an early scientific example – are built. Furthermore, the classification process would seem to be most concerned initially with the establishment of a certain level of intermediate prototype. For example, according to this view, the term ‘oak’ would precede not only the more abstract and more inclusive ‘tree’, but also the more specific – but more concrete – ‘white oak’, the latter two representing later sophistications upon the initial, intermediate category.⁴ Thus, the birth of a fully-fledged system of categorization can be seen to grow in two opposing directions from this intermediate point.



The wider import of this is that the picture of categorization which it presents can be seen to parallel the development of robust categories in scientific inquiry. The establishment of prototypes or generic ideas at a ‘basic level’ represents a sort of initial theory as to the composition of the world; it says, ‘there are different, distinct *types* of thing in the world’.

That the process of categorization should begin at this level is, perhaps, unavoidable. Our knowledge of what establishes a ‘type’ of thing, as well as our understanding of the underlying structures that such groupings are based on, are later sophistications; the first step in establishing groupings is to identify common elements – which, as we

¹ Berlin (1978), p.10.

² Ibid.

³ Ibid., p.14.

⁴ See Rosch (1978), and her theory of ‘basic level’ categories (from which the above example is taken); also Berlin’s complementary discussion of ‘generic ideas’ (see above).

shall see, can often turn out to be misleading: genetically, for example, we frequently find that species *X* has more in common with species *Y* (which, in terms of appearance, it doesn't resemble), than it does with species *Z* (which it does resemble).

However, it may also be argued our initial attempts at identifying natural kinds are testament to the unavoidably anthropomorphic tendencies of human understanding. Scientifically speaking, we might say that human beings have trouble dealing with both the very big and the very small, and our greatest scientific challenges therefore lie in the quantum and stellar worlds, which our 'middle' world of human desires and motivations least resembles. For instance, as James Ladyman observes, quantum physics has shown that we are fast running out of candidates for mind-independent properties:

None of the primary properties of matter listed by Locke and other corpuscularians are now regarded as true properties of the ultimate constituents of matter. Even mass is now regarded as a secondary property produced by the "rest mass" of things in a certain frame of reference. The only candidates for primary properties that physical science now ascribes to things, such as charge, isospin, spin, "colour-charge", and so on, lack any counterparts in our experience, so we can hardly say the sensations they produce in us resemble them.⁵

The reason why the modern concepts of quantum physics 'lack any counterparts in our experience' is that the concepts we use to deal with everyday life (the 'middle' world) are not ultimately fit for describing the subatomic world. So, whilst such developments may or may not reveal an inherent *incommensurability* between human understanding and the world, it does at least reveal that our first attempts at such an understanding are likely to be anthropomorphic – and ultimately false.

Therefore, humans begin their search for knowledge not from some set of mythic, absolute first principles or divine overview, but from a more conveniently human, intermediate perspective. However, as Nietzsche points out, such false beginnings are not as harmful as we might think; and point, in fact, to the real driving force behind such 'errors':

The falseness of a judgement is to us not necessarily an objection to a judgement [...] The question as to what extent it is life-advancing, life-preserving, species-preserving, perhaps even species-breeding; and our fundamental tendency is to assert that the falsest judgements [...] are the most indispensable to us [...] to renounce false judgments would be to renounce life, would be to deny life.⁶

This is a point taken up in numerous places in Nietzsche's writings – for instance, in the short essay, 'Truth and Falsity in an Ultramoral Sense', where he argues that it is

in a limited sense only that man desires truth: he covets the agreeable, life-preserving consequences of truth; he is indifferent toward pure,

⁵ Ladyman (2002), p.144.

⁶ Nietzsche (1990a), section 4, pp.35-6.

ineffective knowledge; he is even inimical toward truths which possibly might prove harmful or destroying. And, moreover, what after all are those conventions of language? Are they possibly products of knowledge, of love of truth; do the designations and the things coincide? Is language the adequate expression of all realities?⁷

Nietzsche is here making two suggestions: (1) what we commonly think of as truth is not in fact as disinterestedly objective as we like to suppose; (2) a non-anthropomorphic and 'objective' truth (if, in fact, there could be such a thing) would not necessarily serve human purposes best. In this sense (to return to the notion of categories), our immediate interests are frequently served by the identification of significant groupings which may not (and probably do not) correspond with the way the world is in itself (which, incidentally, is *a priori* unknowable – in this sense, Nietzsche's position is a form of antirealism – I will return to this point shortly). Therefore, we prefer those ways of seeing the world which have 'life-preserving consequences' and eschew both 'pure, ineffective knowledge' and truths which may prove 'harmful or destroying'. For Nietzsche, therefore, the ways in which humans see the world is riddled with convenient fictions which serve some positive, life-promoting purpose (even if that purpose is simply to hide the inherently *inimical* nature of the world from us).

Can we therefore consider the notion of category to be an example of this human need to assert an arbitrary order ('false judgements') upon an otherwise chaotic world of experience? Furthermore, can we view the type of conflict which we face in defining death to spring from this gap between human concepts and reality, as something which actually reveals this as a fact?

The question to be considered here therefore is that, given that the categorization process begins in intuitive, anthropomorphic groupings for a variety of useful, 'life-advancing' purposes (but which may ultimately prove unreliable), *does it ever get beyond that?* Nietzsche may be interpreted as saying that we do not; we simply evolve more sophisticated anthropomorphic 'falsehoods'. Ultimately, I will argue that the problem we face in determining death is caused by a reluctance to accept Nietzsche's point, and a determination to treat the problem as rationally soluble. However, before reaching that discussion, we need to give this question its full due: do we advance from value-laden, mistaken, 'intuitive kinds', to the discovery of scientifically robust, objectively true 'natural kinds'?

Intuitive, Natural, and Theoretical Kinds

The world continuously upsets many of our intuitive, psychological, and utility-driven attempts to order and compartmentalise it. The need which therefore arises to revise and adjust categories begins, of course, a refining process which is itself the birth of science. W. V. O. Quine, in his essay 'Natural Kinds',⁸ puts this very well in relation to the concept of colour:

⁷ Nietzsche (1965), p.506.

⁸ Quine (1969b).

Nothing in experience, surely, is more vivid and conspicuous than color and its contrasts. And the remarkable fact, which has impressed scientists and philosophers as far back at least as Galileo and Descartes, is that the distinctions that matter for basic physical theory are mostly independent of color contrasts. Color impresses man; [...] But color is cosmically secondary. [...] Color is king in our innate quality space, but undistinguished in cosmic circles. Cosmically, colors would not qualify as kinds.⁹

It is the inadequacy of such 'intuitive kinds', as Quine calls them, that leads – by a 'trial and error process of theorizing' – to the development of better-equipped 'theoretical kinds'.¹⁰ This helps us to develop an understanding of the deeper, more significant features that sometimes unite apparently dissimilar specimens under one kind, or conversely, separately categorize those things which on the surface appear to be of the same kind. For instance:

A crude example is the modification of the notion of fish by excluding whales and porpoises. Another taxonomic example is the grouping of kangaroos, opossums, and marsupial mice in a single kind, marsupials, while excluding ordinary mice. By primitive standards the marsupial mouse is more similar to the ordinary mouse than to the kangaroo; by theoretical standards, the reverse is true.¹¹

Intuitive kinds, therefore, act as a basis for the process of refinement and development which ultimately produces theoretical kinds; the latter form, in turn, a more efficacious means of predicting, explaining, and controlling our natural environment. So, from intuitive groupings of things according to basic, immediately apparent criteria (such as visual similarity), the process of categorization is refined to a point where classifications and groupings can occasionally seem, according to 'primitive standards', counter-intuitive.

On the one hand, then, the process involved here is therefore developmental: over time, intuitive kinds give way to theoretical kinds as our desire for categories with greater explanatory power drives us forward. Thus, a theory which predicts the physical behaviour of all falling apples is more powerful than one which just describes that of a single apple, or only of Cox's Pippins. Furthermore, a theory which can not only account for the descent of all apples, or all fruit, but for all physical bodies in freefall, is the most powerful of all. However, as the Dark Ages exemplify, this development is by no means inexorable; progress may stagnate, and may even enter into reverse, where dogmas and tenets of belief are considered inviolable.

The important consideration here for the analysis of the concept of death – at this point in my argument, at least – is whether the above picture to some degree provides a framework for understanding the problem. In other words, would the assertion that our basic notions of kind are intuitive and inadequate (i.e. based on a rudimentary criterion of similarity and category membership) provide us with a model of the conflict which we are faced with concerning certain individuals who are near death?

⁹ Ibid., p.127.

¹⁰ Ibid., p.128.

¹¹ Ibid.

However, it may be argued, even if this is a suitable model, couldn't this conflict be resolved? Mightn't it be possible, by the refinement of the criteria for membership of these categories, and thus through the development of more sophisticated theoretical kinds, to establish a procedure for generating robust categories (which we might then apply to the death debate)?

The above question is not dissimilar to the one asked by advocates of eliminative materialism in the field of philosophy of mind, where, if we eliminate the troublesome terms, we may replace them with more efficient concepts. As John Searle puts it:

If the theory goes, the theoretical entities go with it: to demonstrate the falsehood of folk psychology would be to remove any justification for accepting the existence of the folk psychological entities.¹²

Here too, in the relation to the definition of death, we are arguably faced with a 'folk belief' regarding what constitutes a living person, and there are also attempts to 'eliminate' such concepts in favour of more conclusive and reductive ones. Are such moves viable, or merely an attempt to gerrymander our conceptual boundaries? Let us consider some possible models of conflict resolution.

Conflict and Resolution in the Development of Categories

What is the process of conflict which provokes the development from intuitive kinds to theoretical kinds, and what persuades us to abandon categorical schemes based on the former in favour of the latter? If it is true that the formation of initial categories is based upon immediately apparent features (such as physical appearance), then the frustration of these categories only comes about when less obvious features become apparent. To recall Quine's example, 'color is cosmically secondary',¹³ and so humans have discovered that not only do all red things not necessarily agree in their physical composition, but also that there is no 'red stuff' that is uniformly responsible for embodying or producing the colour red. Furthermore, knowing that something is red does not help us to predict its behaviour or fix its physical composition. In short, colour classification is a poor tool for the furthering of our scientific understanding of the world.

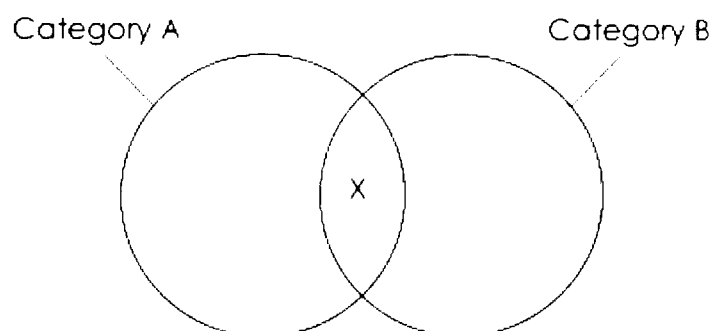
Firstly, if our notion of kind does not adequately allow us to explain why certain things are similar or dissimilar, then there will be a corresponding failure in our ability to predict the behaviour or reaction of the things in question under certain circumstances. Furthermore, such inadequacy will also lead to a limitation in our ability to utilise said things, leading either to its inappropriate use (such as arsenic as a green dye in wallpaper, or lead in petrol), or else a failure to realise a potential use (e.g. the surgical use of maggots to clean infected wounds). This is not to say that such rudimentary groupings *cannot* be useful or correct – of course they can ('brown eye-colour is genetically dominant over blue eye-colour') – but it depends on the purpose. However, for scientific purposes, intuitive and rudimentary kinds tend to be either of limited use or ultimately misleading.

¹² Searle (1992), p.6. Eliminative materialism is associated with such thinkers as Churchland (1981) and Stich (1983).

¹³ Quine (1969b), p.127.

We may talk about this process of the refinement of categories in terms of the inclusion or exclusion of a category's members according to their possession of certain properties. Thus, the utilisation of 'red things' as a category may be seen to be useless for many purposes, because the other features possessed by its members are too many and too disparate, and thus for many purposes the category itself must be abandoned. However, on other occasions, the category may persist in the light of new information regarding its members (e.g. all instances of phosphorescence may indicate the presence of the same chemical); but certain members may need to be reclassified to another existing category, be made to form a subclass within the original category, or be removed to form a discrete class of their own. As I have already pointed out, in scientific settings, such examples of reclassification, abandonment of categories, and creation of new schemes of classification, do not necessarily represent failure, but rather are a symptom of progress. This progress is, in turn, measured in practical terms: a greater ability to explain, predict, and utilize according to the new theory.

However, on certain occasions, this process can give rise to conflicts and difficulties which do not seem to be so readily resolvable. An example of such a situation would be where a would-be member of a particular category possesses, equally, sufficient characteristics to include it in either of two categories (what I will term the 'intermediate case' scenario). This can be simply represented in the following Venn diagram:



As can be seen, *X* shares a number of characteristics with both categories *A* and *B*, but which category should *X* be grouped under?

At this point, I need to preface the following discussion with some definitions. In considering the potential resolution of category conflict, I am going to consider various possibilities. However, the viability of certain possibilities will differ according to the degree to which *realist* or *antirealist* positions are held. Accordingly, I will first define what I understand by these terms.

I take *realism* (or *metaphysical realism*) to stand for the view that there is a mind-independent reality which can in turn be understood by human beings.¹⁴ However, since most realists also subscribe to the view that we cannot perceive this world directly, but only via mental representations or ideas (*idealism*), then most forms of

¹⁴ For the distinction here between realism and antirealism I have followed Ladyman (2002), chapter 5, pp.129-61, which also informs much of the following discussion.

realism are indirect (i.e. representative or causal realism). As such, this implies some sort of correspondence theory of truth – as, for example, held by John Searle,¹⁵ who argues that truth may be defined as ‘the variety of ways in which statements can accurately represent how things are’.¹⁶ Thus, realists believe that our ideas in some way *correspond to* (describe, represent, etc.) mind-independent facts or states of affairs. Furthermore, we may distinguish between realism regarding the existence of mind-independent reality (*metaphysical realism*) and realism regarding the existence of discrete concepts or universals (*conceptual realism*). The former is a still widely-held doctrine, whilst the latter is a somewhat outdated view, which is best exemplified by Plato.

On the other hand, *antirealism* may be taken to stand for the view that, whether or not such a mind-independent reality exists, we cannot know it. Most antirealists also subscribe to idealism, but, unlike realists, argue that truth cannot be understood as the correspondence of our views to a mind-independent reality, because such a reality is unknowable; they will therefore usually favour some definition of truth which argues for its origin in (for example) instinctive drives (Nietzsche), social conventions (social constructivism),¹⁷ or some other explanation which attempts to account for truth in a way that is *internal* to human linguistic and social practices (e.g. a form of *coherence* theory).¹⁸ It should be noted, however, that antirealists are not committed to the idea that the mind-independent world does not exist (as with Berkeleyan idealism),¹⁹ but merely that knowledge of a mind-independent reality is impossible, and that we must therefore find some other basis for our concepts.²⁰

In light of the above, we may therefore construe the problem of the definition of death differently to the extent that we adopt realist or antirealist approaches to the concept of categories. If, for instance, we were to consider categories *A* and *B* as fixed, mutually exclusive and irreducible, then *X* may represent a genuine problem for us (and, arguably, as it currently does in our applications of the concepts of ‘alive’ and ‘dead’).²¹ In this case, such an approach might picture a category as an irreducible feature of a mind-independent reality. However, if we are pliable in our notion of what defines categories *A* and *B*, considering them mutable or even dissolvable, then a number of ways of dealing with *X* present themselves. Of course, this latter position is not antirealist *per se*, but only in relation to categories and classification systems (that is, it is a rejection of conceptual realism). After all, one may reject the notion of categories altogether and still be a realist regarding the world.²²

¹⁵ See Searle (1995), chapter 9, ‘Truth and Correspondence’, pp.199-226.

¹⁶ *Ibid.*, p.213.

¹⁷ For example, Kuhn (1962) – according to some interpretations, at least.

¹⁸ Such as, arguably, Wittgenstein – see, e.g., Wittgenstein (1974).

¹⁹ See, for instance, Berkeley (1975).

²⁰ Accordingly, antirealism can take many forms, and philosophers can be antirealists to different degrees and in different respects. The same is obviously true of realism. However, I shall try to be specific about which aspects of these views are relevant to the point under discussion.

²¹ For sake of convenience, I have drawn categories *A* and *B* as overlapping. However, in illustrating the problem, it should be understood that *A* and *B* are *mutually exclusive* and that it is only the existence of *X* which *makes them* overlap (just as the existence of ambiguous near-death situations makes the otherwise mutually exclusive categories of ‘alive’ and ‘dead’ overlap).

²² See the section below, ‘The Redundancy of Categories and the Death of Natural Kinds’.

However, there would seem to be an underlying, implicit commitment to maintaining the binary opposition between 'life' and 'death' in many solutions to the problem. But if we regard certain categorical groupings as inviolable, then our options for conflict resolution are limited: in the absence of a third option (or, for instance, a spectrum of alternatives), we are left with binary possibilities, and any problem case must be made to fit one or the other. In relation to the binary categories of human life and death, therefore, might it not be possible to provide a resolution to the problems faced in the death debate by rejecting the basis of this binary opposition?²³

Conceptual Realism and Conceptualism

But what is the basis of this opposition? As already noted, neither metaphysical realism nor antirealism need involve a commitment to the idea that certain of our concepts are fixed and inviolable. However, in as much as antirealism is a broadly empiricist position, then for such a position to argue that our concepts are in some way fixed it would need to be shown that there is a natural limit to our conceptual definitions. Consequently, a position which holds that 'death' is a fixed concept with definite boundaries must either spring from conceptual realism, or else some empiricist position which accounts for conceptual limitation in non-rationalist terms (which we may term *conceptualism*). The latter position, however, would also seem to allow for a degree of contingency regarding the form of concepts in that they are based on non-rational or natural factors (e.g. human physiology, the features of human language, etc.), and the degree to which our concepts are fixed would therefore seem to rely on the extent to which the determining features of human life which influence concept formation are themselves fixed; accordingly, I will designate varieties of conceptualism 'liberal' or 'conservative' according to the extent that they lean toward conceptual realism or nominalism respectively (this is an important point, as will be seen later on in this chapter). For the purposes of this chapter, therefore, I will concentrate on conceptual realism and conceptualism, which I argue supply the main (implicit) assumptions underlying many attempts at resolving the present definition of death controversy. (I will consider the more general question of the truth of metaphysical realism, and its relevance to the debate, in following chapters.)

In the terms of medieval scholasticism, conceptual realism and conceptualism both take their place along a continuum on which *nominalism* may be found at the other extremity.²⁴ Nominalism may be considered as broadly antirealist in that it would seem to deny any connection between a mind-independent reality and our concepts; however, conceptualism may be either realist or antirealist (in the metaphysical sense) depending on how it views the process of concept formation. For instance, causal realists such as Locke and Searle hold to a form of conceptualism in that they believe that, although a mind-independent reality exists, our concepts are human constructs. However, this is different from nominalism in that there is a sense in which our concepts are *related* to reality, and that they are therefore not arbitrary. As William H. Youngren puts it:

²³ I will at times refer to life and death as binary concepts (you are either alive or dead), and that this is a traditional and conventional assumption. However, this is not to say that the signs of death have always been fixed and identifiable – on the contrary, anxiety and uncertainty to do with the diagnosis of death seems to stretch right through from antiquity to the present day (see Russell 2000, pp.1-11).

²⁴ See Stanilard (1972), pp.26-7; also Youngren (1980).

Instead of maintaining with the realist that universals are special immaterialist entities existing either apart from or in particular objects, or with the nominalist that (in Hobbes's words) there is "nothing in the world universal but names", the conceptualist maintains that universals are concepts or ideas in the minds of human beings. Thus the reason that a general term can be rightly applied to a particular thing is not (as for the realist) that the particular thing manifests or instantiates a certain universal or (as for the nominalist) simply that the word is in fact applied to the particular thing. For the conceptualist the reason is rather that the particular thing is related in a special way to a certain concept or mental entity: as Locke often puts it, the thing "agrees with" or "conforms to" the concept.²⁵

Furthermore, an antirealist might also be conceptualist if they hold that human concepts are to an extent determined – e.g. by cultural forces or competing values. Different varieties of conceptualism may also be distinguished depending on the *degree* to which our concepts are more or less determined. So, we might consider a conceptualist to be conservative if he considered human concepts to take a determinate form, and liberal to the extent that they allowed for a potential variance in the form of concepts.

As such, the conceptual landscape may be described accordingly:

- (A) *Conceptual Realism*. An account of general terms and the form of concepts by appeal to mind-independent universals (e.g. Plato). This may be distinguished from *metaphysical realism*, which simply holds that there is a mind-independent reality.
- (B) *Conceptualism*. The attempt – such as that of the early empiricists (Locke, Berkeley, Hume), but also later ones (e.g. Wittgenstein) – to account for general terms via empiricism without appeal to universals. In this sense, conservative forms of conceptualism (Wittgenstein) attempt to set limits to concept formation and the meaning of terms (e.g. by appealing to 'the logic of our language' and 'forms of life').²⁶ More liberal forms of conceptualism may however admit the possibility of conceptual mutability, whilst also admitting that there are factors and limits which save us from total nominalism. As already noted, conceptualists may be realist or antirealist to the extent that they see a mind-independent reality as in some way limiting or necessitating our concepts.
- (C) *Nominalism*. The position that our designation of general terms is, to some extent, arbitrary, and may be revised accordingly. However, as Youngren notes, 'it is such a difficult position to hold consistently that finding a true-blue nominalist who is also a first rate thinker is almost impossible'.²⁷

²⁵ Youngren (1980), p.707.

²⁶ See Wittgenstein (1967). Wittgenstein, and the linguistic turn in general, may be considered an inheritor of certain empirical traditions. In this sense, it may be considered a conservative variety of conceptualism. I shall discuss what I consider to be the limitations of such conservative conceptualism in chapter 8.

²⁷ Youngren (1980), p707.

(William of Ockham, who is often thought of as a nominalist, is actually a conceptualist,²⁸ and ‘even Hobbes cannot consistently hold to his nominalism’).²⁹ Consequently, nominalism often collapses into conceptualism, for few can hold that *all* our concepts are completely arbitrary. However, should such a position exist, then it would be antirealist.

Perhaps these distinctions are best illustrated by different and opposing examples. For instance, Wittgenstein argues that certain forms of discourse are meaningless because they seek to go beyond the limits of language, and – as Felicity McCutcheon points out – he believes that ‘We cannot access a reality independent of the concepts we use to think about it’.³⁰ In this sense, Wittgenstein comes close to antirealism: it is not that he does not believe in a mind-independent reality,³¹ but he believes that (like Kant) we cannot have access to it, and furthermore that we may reach a conceptual limit to our enquiries. However, Wittgenstein also rejects conceptual realism, in that he proposes that our concepts do not make sense in virtue of their relation to a rational universal, but rather within the context of human practice and discourse. In this sense, Wittgenstein is a conceptualist, in that he holds that (a) there are no rational universals, but that (b) there are factors which limit possible concept formation. (If he had rejected (b), he would be a nominalist.) I shall therefore term Wittgenstein a *conservative conceptualist*.³²

In comparison with this, we may consider Nietzsche’s approach. For instance, the notions of cause and effect are seen as explanatory conventions which may have no reality in their own right,³³ as possibly is the case with the desire to attribute a causal centre to agents or events (what he terms ‘atomism’).³⁴ For Nietzsche, then, whilst (a) he also rejects conceptual realism, he holds that (b) whilst there are factors which determine concept formation, there is a sense in which the potential variance of our concepts is much greater. So, whilst Nietzsche would also reject nominalism, we find that the factors which determine our concepts are less restrictive. Nietzsche may also be thought an antirealist – and perhaps with more cause. I shall therefore term Nietzsche a *liberal conceptualist*.

(These two versions of conceptualism will be important for my overall argument. For, whilst I may not ultimately wish to argue that we may define death in whatever way we wish (i.e. nominalism), I will eventually propose that adopting something like liberal conceptualism may allow us to view the whole debate from a new perspective. I will deal with this possibility in chapter 8.)

²⁸ Ibid., p.707-8.

²⁹ Ibid., p.708.

³⁰ McCutcheon (2001), p.25.

³¹ As has been pointed out, the *existence* of a mind independent reality is not one that many philosophers would attack; it is our access and relationship to it that is controversial. See, for instance, Page (2004), pp.255-6, and Rorty (1997), p.160. However, Wittgenstein’s antirealism is a contentious point – and McCutcheon herself rejects this view of him.

³² It may be controversial to designate Wittgenstein in this way, but it is I think defensible. After all, if we see conceptualism as holding the middle ground between realism and nominalism, then Wittgenstein’s linguistic approach (which is both anti-essentialist and seeks to prescribe limits) can be viewed in this way. See Youngren (1980), pp.724-6.

³³ Nietzsche (1990a), section 21.

³⁴ Ibid., sections 12, 17, and 20.

Returning to our general definitional problem (the ‘intermediate case’ scenario), it can be seen that the options open to us in resolving the problem will differ according to the attitude adopted in relation to the nature of our concepts. Regarding the options of the nominalist view of categories (or even a more liberal conceptualist), there are a number of moves which are possible. For instance, let us imagine that a new breed of animal (X) has been discovered, that it shares genetically a number of characteristics with dogs and cats respectively (let category A stand for ‘Dog’ and B for ‘Cat’). Now, the options for dealing with animal X are as follows:

- (1) To create a new, hybrid category of ‘Dat’ (or ‘Cog’ – or whatever) for animals just in possession of those same characteristics as X .
- (2) To redefine ‘Dog’ so as to include X (whilst redefining ‘Cat’ so as to exclude it).
- (3) To redefine ‘Cat’ so as to include X (whilst redefining ‘Dog’ so as to exclude it – i.e. the reverse of (2) above).
- (4) To combine the categories of ‘Cat’ and ‘Dog’ into a larger set, so that X is a subset of the larger category (i.e. it is another example of a species of which cats and dogs are also members).

There may be other options than these, but the above four will serve to illustrate the sort of possibilities available.

So, in relation to the above possible resolutions to the problem, we are left with two main positions:

- (a) The concept of death *cannot* be revised, because there are factors which limit concept formation (a conceptual realist or conservative conceptualist position).³⁵
- (b) The concept of death *can* be revised, because there is at least a degree of mutability to our concepts (moderate nominalism/liberal conceptualism), and at most a complete arbitrariness (extreme nominalism).

Possibility (a) is therefore the main stumbling block for any possible revision of the concept of death in order to accommodate the intermediate case scenario. Therefore, if we take the problem of the definition of death as being of this sort, then certain cases (X) seem to have equal claims for membership of either category (‘being alive’ or ‘being dead’). If we accept the validity of presenting the problem in this way, then the proponents of (b) have at least four options:

- (1) To admit that X s represent genuine intermediate cases and therefore constitute a distinct third category (‘neither living nor dead’).

³⁵ As an intermediate point on this continuum, conceptualism may be considered either as a form of moderate realism, or moderate nominalism. I shall return to this point later when I examine conceptualism in more detail.

- (2) To admit that the current way of defining ‘being dead’ is incorrect, and to adjust the criteria accordingly to include cases of *X*.
- (3) To admit that the current way of defining ‘being alive’ is incorrect, and to adjust the criteria accordingly to include cases of *X*.
- (4) To combine the categories of ‘being dead’ and ‘being alive’ so that *X* is a subset of the larger category.

As can be seen, the above options parallel those in the animal *X* scenario. Interestingly, the equivalent of options (1) and (4) in that case would represent – in the context of the definition of death – more problematic and controversial moves. However, any objection to this could only be made by a conceptual realist, who would consider such options conceptually invalid, or a conservative conceptualist, who might object on some non-rationalist basis (e.g. that there were linguistic limits to our concepts). If, however, we reject both the conceptual realist and conservative conceptualist approaches, with their attendant rigidity, we may still avail ourselves of a number of nominalist, or liberal conceptualist solutions. However, in doing so, we are presented with options which are at least highly controversial, and at worst – in the eyes of some – morally repugnant.

Most attempts at resolving the definition of death controversy to date have centred around options (2) and (3) – i.e. by redefining or extending the definition of ‘being alive’/‘being dead’ so as to include or exclude the problematic cases. (These approaches, although their metaphysical allegiances often remain unstated and unexplored, are by and large conservatively conceptualist in nature.) These options are the most popular in that they retain what appears to be a vital ingredient of any solution for the proponents on either side – that is, the mutually exclusive, binary nature of the definition (‘alive’ or ‘dead’). However, since options (1) and (4) both disrupt this opposition (option (1), by introducing a third category: ‘neither alive nor dead’; option (4), by making all categories subcategories of one greater category – and thus doing away with the distinction altogether), they have been correspondingly ignored by most advocates.

So, to summarize: I have so far argued that the problem of the definition of death can be stated as a problem concerning categorization. Furthermore, I have argued that similar types of problem only present difficulties for conceptual realists and conservative conceptualists, and that there are both realist and antirealist solutions to such problems. However, of the four identified options, when applied to the definition of death debate, two of these options would result in controversial or bizarre situations.

To illustrate this concretely, let us now return to the patient in PVS first discussed in Chapter 1. Proponents of options (2) and (3) would argue, according to a range of criteria, that the person was either alive or dead (and that our current conception of both categories should be altered). However, imagine now that someone proposes option (1) – that the person is neither alive nor dead, but actually exists in an intermediate state. According to this position, however, such a person is not in an *indeterminate* state – in other words, the proponent of option (1) is not arguing that we *do not know* whether they are alive or dead; rather, it is being argued that they

satisfy certain criteria for being in a state which is neither 'alive' nor 'dead'. For instance, this position is briefly considered by Geoffrey Scarre:

One response to such cases would be to abandon the normally unquestioned assumption that death is an all-or-nothing affair. Perhaps we should describe PVS subjects as occupying a halfway house between life and death, analogous to the way in which we might describe twilight as a state between day and night [...] As failure of brain function is a matter of degree, it may seem reasonable to regard life and death as matters of degree too, so that someone who has reached a certain point of decline might be deemed to be more dead than alive.³⁶

Of course, a proponent of option (2) or (3) can argue that we do not know – even *cannot* know – that a person is alive or dead under certain circumstances (such as when we do not have access to the required technology); this sort of approach would be in keeping with general metaphysical realism (i.e. we may be wrong, or not currently possess enough knowledge, to decide the issue, but there *is* a definite answer to be found). The proponent of option (1), on the other hand, is instead arguing that there is a genuine state between life and death which is *not* the result of a lack of knowledge on our part; in other words, indeterminate cases are not simply waiting to be disambiguated, they *are* a third state. (The latter position would therefore seem to imply some liberal conceptualist position, which allows for a degree of conceptual flexibility in line with novel situations, whilst also denying that our concepts are completely arbitrary.)

Conversely, adopting option (4) has the effect of resolving the problem by not only dissolving the mutual exclusivity of the categories of 'life' and 'death', but also of admitting a range of stages or states between which it is impossible to clearly differentiate. It is as if we were to say, 'We shall no longer have the separate colours "green" and "blue", but instead adopt one category to cover the whole spectrum of colour shades which we shall term "greater turquoise"'.³⁷ (This would be a liberal conceptualist/nominalist position).

The point I am trying to make here is that the viable options regarding the resolution of category conflict will differ depending on the conceptual approach adopted (or assumed). However, there are legitimate and established ways of resolving ambiguities and difficulties involving categorization besides options (2) and (3). The problem, therefore, exists only for those philosophers who would hold to the binary

³⁶ Scarre (2007), p.8

³⁷ Or, perhaps, to use Nelson Goodman's example, 'grue' (see Goodman 1983, p.73 onwards). Goodman's use of the term, however, is not directly used to illustrate some problem with the notion of categories, but rather to illustrate certain problems involved in formulating a sound theory of inductive reasoning. However, there is a parallel here between Goodman's problem of induction and my own case of problems involving category membership; as Goodman points out, 'the problem of prediction from past to future cases is but a narrower version of the problem of projecting from any set of cases to others.' (p.82). In relation to induction, Goodman's point is that, without a clear understanding of the difference between lawlike and accidental relationships between evidence and hypothesis, then we are left with many possible and contradictory hypotheses. In relationship to category membership, therefore, we can see that the problem can be stated as how we can identify a particular set of features (those possessed by the 'dead') so that all future cases of being dead will possess those features.

opposition of life/death, and would see this as being prescribed in some way (i.e. conceptual realism or conservative conceptualism).³⁸

Death and the Excluded Middle

Before moving on to a more detailed consideration of the type of conceptual approach which lies at the bottom of the intermediate case scenario, it is necessary briefly to consider whether this problem may be resolved by alternative *logical* concepts. It may be argued, for instance, that the need to maintain a binary opposition between certain categories (i.e. 'life' and 'death') stems from a certain view of what is logically possible. For instance, Aristotle's notion of the 'excluded middle'³⁹ proposes that something must either *be* or *not be* the case; or, in Leibniz's formulation:

*every proposition is either true or false. That is false which is the contradictory of the true; those propositions are contradictory which differ only in that one of them is affirmative and the other negative. These principles are such that it is vain to demand a proof of them.*⁴⁰

In relation to our context, therefore, a person must either be 'dead' or 'not dead', and no intermediate cases are permissible.⁴¹

Occasionally, in the history of philosophy, attempts have been made to undermine this 'law' of logic and to show that it does not obtain in all cases. So, until Bertrand Russell's theory of definite descriptions attempted to resolve them, certain paradoxes seemed to suggest that there were particular types of statement that were not either true or false (e.g. the famous statement, 'the present king of France is bald').⁴² Similarly, other philosophers have attempted to argue for 3- or many-valued logic systems,⁴³ where the Aristotelian binary opposition is replaced with one or more intermediary possibilities. As well as these examples, we may point to the many-valued logics found in certain schools of Indian philosophy and religion, where – for instance – Jainism proposes a logic with a total of seven possible forms of judgement,

³⁸ I will explore the possibilities open to a liberal conceptualist in chapter 8.

³⁹ See Aristotle (1924), 4.4; and Humberstone (2000).

⁴⁰ Leibniz (1973), 'The Nature of Truth', p.93.

⁴¹ There is some discussion of possible intermediate states in the literature, and the possible consequences for medical practice; for instance, Waylard Gaylin's use of the concept of the 'neomort' in Gaylin (1974). However, such discussions tend to treat intermediate states as problematic, rather than as a putative resolution: see, for instance, the film *The Island* (Bay, 2005), which takes the organ-harvest scenario to a shocking, sci-fi driven conclusion; Jonas (1978), pp.55-7, which echoes Galyin's vision of bodies being used as organ banks and vehicles of '*in vivo*' research, and decries the 'cowardice of modern secular society' which, through its fear of death, seeks to redefine it in the 'relentless expanding of the realm of sheer thinghood and restricted utility' (p.58). The quoted passage from Scarre (2007) above suggest that he is slightly more open to the possible coherence of the idea, so that – whilst he recognises the problems inherent in the intermediate state – he considers that 'there seems to be no compelling reason for rejecting a conceptual shift that would permit more nuanced descriptions to be given of a subject's state' (Scarre 2007, p.8).

⁴² See Bertrand Russell (1973).

⁴³ Also termed 'non-classical' logic systems, for instance, the idea of 'truth degrees' in the intuitionist logic of Kurt Godel, or the 'fuzzy' logic developed for use in certain computing contexts by (e.g.) L. A. Zadeh.

five of which contravene the law of excluded middle (e.g. 'S is P, and S is also not-P').⁴⁴

However, the opposition between 'alive' and 'dead' is not of the same type as 'A' and 'not A'. The reason for this is that there is no logically necessary reason for us to equate 'A' with 'dead' and 'not A' with 'alive'. Rather, we may only equate 'A' with 'dead' and 'not A' with 'not dead', which leaves the possibility of any other number of intermediate states which are 'not A' (i.e. 'not dead') in the same way that 'not being a cat' may involve being any number of other different things.

Therefore, whilst it would be interesting to examine these possibilities in relation to our current concerns, it is not really necessary. It is not the law of excluded middle which is responsible for the strict adherence to options (2) and (3) above, but rather the desire not to abandon the 'binary opposition' of 'life' and 'death', and therefore an underlying metaphysical commitment to the nature of concepts. As I have already argued, the only purely logical reason for maintaining this opposition is the assertion that 'life' and 'death' represent inviolable and mutually exclusive categories of which we have direct knowledge via intellectual intuition (conceptual realism) or else some empirical limit imposed by – for example – 'the logic of our language', as it may be called (conservative conceptualism). Let us now therefore consider this position in more detail.

Essentialism, Significant Features, and Family Resemblances

In the remaining sections of this chapter, I will examine various criticisms of the idea that our concepts have a definite shape or form. As such, I will be attacking two positions already discussed above: conceptual realism and conservative conceptualism. As noted there, the former springs from metaphysical realism (though metaphysical realism does not necessarily entail conceptual realism), and assumes that our concepts have a determinate form and can be objectively apprehended. Accordingly, this position is not widely held in contemporary philosophy. However, the remnants of its influence can be found in conservative varieties of conceptualism, which is an attempt to fix or limit the meaning and form of concepts without appeal to universals. As noted earlier, conceptualists may vary according to the degree of mutability that they allow, and which is in turn related to the factors which they identify as determining the limits and meaning of concepts (which in turn will be linked to their *metaphysical* realism or antirealism). Therefore, whilst conceptualism in general rejects the

⁴⁴ Chatterjee and Datta (1960), pp.80-86. Jainism is generally considered to espouse a relativist – though non-sceptical – philosophy, generally known as Syadvada. The seven logical possibilities are therefore all prefaced with 'syad', meaning 'somehow', 'in some way', or 'it is possible that' (e.g. 'somehow S is P') to indicate that the judgement is only true from a certain point of view or in a certain context. As well as affirmation and negation, therefore, Jainism allows for a third possibility – 'somehow S is and is not P' – where contradictory contexts or aspects are considered together. However, this does not imply a violation of the principle of contradiction, but merely a recognition of the chronological and aspectual contexts of logical judgements. A fourth possibility, therefore, 'somehow S is indescribable', simply suggests that no judgement is possible if no context is specified. The seven possibilities therefore embody all possible combinations of the three qualities (affirmation, negation, and indescribability). However, the debate concerning the existence, nature, and significance of such logical possibilities as Western philosophy does not traditionally recognise has a long history in Indian philosophy (within which we may include Buddhism), and marks a point of difference between different philosophical schools, serving to classify them according to their realist or nihilist tendencies.

rationalist concept of universals, its attempt to fix and limit the meaning of terms and the nature of concepts makes it a target for nominalists and other varieties of anti-essentialist. It is this notion of essentialism which I therefore consider conceptual realism and certain forms of conceptualism to share, and which I will for the purposes of this discussion consider to be a key feature of those approaches which attempt to maintain the binary opposition between the concepts of 'life' and 'death'. Any criticism of essentialism will therefore be taken to apply to both conceptual approaches. I will also more specifically argue that the strictly-biological approach discussed in chapter 2 can be better understood as an essentialist attempt to define death in purely physical terms (and which, as such, is also open to the general essentialist criticisms considered).

Essentialism is commonly characterised as the view that something possesses specific characteristics which define its essential nature. So, in relation to the concept of a triangle, we might say that its essential features consist in having three sides, internal angles that add up to 180 degrees, and so on. In other words, the 'essential properties of a thing are the ones it needs to possess to be the thing it is.'⁴⁵ Without the properties already stated, the geometrical shape under consideration would not be a triangle.

In relation to the definition of death, the strictly-biological view is therefore an example of an essentialist approach which is based solely on physiological criteria. In other words, the search for those biological features which determine conclusively whether a person is alive or not assumes that the concept of death must entail a finite set of criteria. But does it? As I have argued so far, the search for a clear definition of death is an example of the general scientific process of the refinement of categories; starting out as 'intuitive kinds', we refine our notions in order to arrive at 'theoretical kinds'. However, whilst this undoubtedly proves useful, there is a great difference between the usefulness of a category and its ultimate validity. The essentialist position therefore implies the existence of 'natural kinds' – i.e. objects or phenomena 'which have some theoretically important property in common.'⁴⁶ To view death as a natural kind is therefore to suppose that there exist a definite set of features which constitute it.

To sum up, then, the strictly-biological approach

- (1) Views death as a natural kind consisting of essential features;
- (2) Views the search for essential features as the refinement of an ongoing theoretical kind;
- (3) Implies, consequently, a conceptual realist or conservative conceptualist position in relation to our knowledge of those essential features;
- (4) Adopts an attitude to the concept of death whereby its validity as a category, its binary opposition to life, etc., is held to be true (e.g. by intellectual

⁴⁵ Yablo (2000), p.254.

⁴⁶ Daly (2000), pp.612-3.

intuition, in the case of conceptual realists, or some natural or linguistic delimiting factor, in the case of conservative conceptualists).⁴⁷

As noted above, it is the combination of these views which lead to the problem in the first place. If the concept of death were not held to be in fixed binary opposition to life (i.e. by introducing an intermediary state, etc.), and that concept were not seen as a natural kind possessing essential features, then the controversy might be resolved another way (e.g. by some liberal conceptualist approach).

If we are to attack this view – which I aim to do – we must therefore focus on the notion of essentialism. Without it, the above four positions crumble, and the problem of the definition of death is – if not resolved – then at least set on a different footing (which I shall explore later in chapter 8). Of course, the issue underlying this discussion has an ancient pedigree, and is more generally known as the problem of universals.⁴⁸ For instance, in Plato’s dialogues, we find Socrates asking questions such as ‘What is virtue?’, ‘What is courage?’, and so on. However,

Socrates always emphasised that in asking “What is X?”, he was *not* asking to be given illustrative examples of X, nor to be told what kinds of X there are.⁴⁹ What he wanted was a general definition of the quality of X, which would then serve to guide us in deciding what we ought to say about particular cases. For example, people sometimes disagree about whether a particular type of action should really be called “courageous”. According to Socrates, such disputes can be settled only by first discovering the correct definition of courage. He seems, in fact, to have regarded the discovery of such definitions as a precondition of any fully rational thought.⁵⁰

So, in considering the definition of death, we can shed light on the controversy by considering whether or not it is actually an example of the problem of universals: in other words, we may imagine Socrates asking “What is death?”, and seeking a conceptual realist’s answer in terms of the identification of essential features.

One of the most famous attacks on essentialism occurs in Wittgenstein’s *Philosophical Investigations*. In the opening sections, Wittgenstein considers certain sections from the writings of St. Augustine, which, he says, ‘give us a particular picture of the essence of human language.’⁵¹ There, Augustine presumes a one-to-one relationship between words and thoughts (or words and objects). It is a similar position to one held centuries later by John Locke, where he considers the way in which general terms come about:

⁴⁷ Incidentally, it may be denied by advocates of a strictly-biological definition of death that they hold proposition (4). However, if they do not, then they must ask themselves why they do not take another approach to defining death (e.g. a moral one).

⁴⁸ See, for instance, Stanilard (1972).

⁴⁹ Which, incidentally, is what Wittgenstein (and other non-essentialists) *do* offer – of which, more shortly.

⁵⁰ Stanilard (1972), pp.1-2.

⁵¹ Wittgenstein (1967), p.2.

Thus we may conceive how *words*, which were by nature so well adapted to that purpose, come to be made use of by men, as the *signs of their ideas*; not by any natural connexion [...] but by a voluntary imposition, whereby such a word is made arbitrarily the mark of such an idea. The use then of words, is to be sensible marks of ideas; and the ideas they stand for, are their proper and immediate signification.⁵²

The point is here that, for Locke, words are fairly arbitrary labels which designate distinct ideas. Accordingly, when we come to Locke's account of general terms (which is most relevant to our discussion of category membership), we find

That which general words signify, is a sort of things; and each of them does that, by being a sign of an abstract idea in the mind, to which idea, as things existing are found to agree, so they come to be ranked under that name; or, which is all one, be of that sort. Whereby it is evident, that the *essences of the sorts, or* (if the Latin words please better) *species* of things, are nothing else but these abstract ideas.⁵³

Thus, for instance, in the formation of the general term 'man', we 'make nothing new, but only leave out of the complex idea [we have] of *Peter and James, Mary and Jane*, that which is peculiar to each, and retain only what is common to them all.'⁵⁴ Locke is therefore proposing a conceptualist account of general terms: he denies that universals exist, other than in the sense that we abstract the general resemblances that exist among groups of individual instances. However, in doing this, he would also seem to be essentialist in that he does imagine that there is some set of properties that is 'common to them all' – in this sense, therefore, his conceptualism is essentialist (i.e. what I have termed conservative conceptualism). The reason this essentialism remains implicit, however, is because the essentialist assumption lies not in the discovery of an independently existing idea (the rationalist's universal, which the British empiricists rejected), but rather in the nature of the world (or rather our experience of it): the idea that there is an essential group of natural features designated by the term 'man'. So, whilst the linguistic sign that is attached to the general idea is fairly arbitrarily assigned, the general term itself is far from arbitrary, for – as Youngren observes – 'the mind takes its cue from the actual undisputed existence of what Mill was later to call natural kinds'.⁵⁵ It is these natural kinds, therefore, that the empiricists still appeal to. The mind therefore becomes adept at recognising these qualities which already exist in nature. Even in Hume, therefore, we find

Here, then, is a kind of pre-established harmony between the course of nature and the succession of our ideas; and though the powers and forces, by which the former is governed, be wholly unknown to us; yet our thoughts and conceptions have still, we find, gone on in the same train with the other works of nature.⁵⁶

⁵² Locke (1997), p.363.

⁵³ Ibid., pp.371-2.

⁵⁴ Ibid., p.369.

⁵⁵ Youngren (1980), p.710.

⁵⁶ Hume (1975), *An Enquiry Concerning Human Understanding*, p.54.

Therefore, even with those who explicitly oppose conceptual realism, a subtler form of essentialism creeps in.

If we now relate this to the concept of death, we arrive at a definition based on an analysis of all the things that we consider to be 'dead'. In this way, we build up a set of essential properties which may then be used to designate what we really *mean* by 'death'. However, as Wittgenstein points out, there is a deep problem with this view. The central difficulty lies in the identification of those characteristics, features, or meanings which can be said to be essential to *all* uses of that term. In a famous example, Wittgenstein discusses the possibility of identifying what all uses of the term 'game' have in common.

Consider for example the proceedings that we call "games". I mean board-games, card-games, ball-games, Olympic games, and so on. What is common to them all? [...] For if you look at them you will not see something that is common to *all*, but similarities, relationships, and a whole series of them at that. [...] Look for example at board-games, with their multifarious relationships. Now pass to card-games; here you find many correspondences with the first group, but many common features drop out, and others appear. When we pass next to ball-games, much that is common is retained, but much is lost. – Are they all 'amusing'? Compare chess with noughts and crosses. Or is there always winning and losing, or competition between players? Think of patience. In ball games there is winning and losing; but when a child throws his ball at the wall and catches it again, this feature has disappeared. Look at the parts played by skill and luck; and at the difference between skill in chess and skill in tennis. Think now of games like ring-a-ring-a-roses; here is the element of amusement, but how many other characteristic features have disappeared! And we can go through the many, many other groups of games in the same way; can see how similarities crop up and disappear.⁵⁷

Given this shifting nature of definition, the problem for the essentialist lies in identifying just those features which are common to all uses of a term (or, for our purposes, all members of a category). However, as can be seen from the example, this is arguably impossible; for, as Wittgenstein points out, the overall picture is one of 'a complicated network of similarities overlapping and criss-crossing',⁵⁸ where features which unite a certain number of games are not common to others: many games involve competition, but many do not; for some, amusement is a common feature, yet in other cases this very same feature unites activities that in other ways are very dissimilar (so much so that they almost seem to share more in common with things that we do not call games – e.g. rugby and war). Faced with this problem, we have a limited number of choices: (1) to abandon strict essentialism in favour of some other model, or (2) to redefine those things we call 'games' (or 'dead') in line with a definition motivated by (e.g.) cultural or moral factors.⁵⁹

⁵⁷ Wittgenstein (1967), pp.31-2 (section 67).

⁵⁸ *Ibid.*, p.32 (section 67).

⁵⁹ That is, to take a nominalist and or liberal conceptualist approach. This is explored in subsequent chapters.

Option (1) is what Renford Bambrough opts for in his theory of ‘family resemblances’⁶⁰ – and what he claims Wittgenstein, in introducing this metaphor, is also suggesting that we employ.⁶¹ However, there are good reasons for considering this form of revised essentialism (conservative conceptualism) to be as unworkable as pure essentialism (and, arguably, a misinterpretation of Wittgenstein). Dick Beardsmore, in his paper ‘The Theory of Family Resemblances’, provides a telling criticism of this view in terms of being a member of a family:

Unfortunately, what makes the members of a family into a family has nothing whatsoever to do with family resemblances [...] For the unity of a family is *not* the unity of a pattern of family resemblances. My sharing the Churchill family face will not make me a member of the Churchill family, and had the son of Lord Randolph lacked it, he would still be one of the family’s most distinguished members.⁶²

Beardsmore’s point is a valid and important one: category membership is not merely a case of possessing a certain *quantity* of characteristics, but is rather decided by the presence of *significant* ones.⁶³ In Beardsmore’s example, looking or not looking sufficiently like members of one’s family is, ultimately, less important than actually being genetically related, and in certain circumstances, certain features – such as, in this case, genetic relationship – have correspondingly greater weighting than others. If no difference between the significance of features is made, if ‘being genetically related’ were simply one other characteristic which we take into account along with ‘having the same nose’, ‘having the same eyes’, etc., then absurd consequences arise (such as Beardsmore’s membership of the Churchill family). Similarly, what makes an animal a member of a particular species is not just the possession of *enough* characteristics, but rather the possession of *specific* characteristics.⁶⁴ For example, from a quantitative view of its features, rugby may have more in common with war than it does with solitaire; nonetheless, we consider rugby and solitaire both to be games, whilst to consider war a game (unless in a metaphorical sense) would be highly inappropriate – the significant feature here being the presence of ‘the intention to take life’.

⁶⁰ Bambrough (1960).

⁶¹ I am not really concerned with whether Wittgenstein is in fact proposing this hypothesis at this point of the *Philosophical Investigations*, though I think it is safe to say that even if he is, he later goes on to reject it. This is not to say, however, that Wittgenstein ends up putting forward some sort of ‘extreme nominalist’ position (i.e. that the only things games have in common is that we call them ‘games’). Rather, Wittgenstein ultimately seeks to undermine the possibility of such a nominalist position being true by attacking the associated possibility of a completely private language (which a nominalist position would implicitly support). For discussion of these points in relation to the family resemblance argument, see Stanilard (1972), pp.68-83.

⁶² Beardsmore (1992), p.146.

⁶³ This is an especially important point when we come to consider the definition of death controversy because many arguments for drawing the boundary one way or another centre on the presence or absence of certain features of the individual which are seen as essential to that person being said to be alive (or dead). For the moment, however, I wish to continue to deal with the issue of categorization in abstract terms so as to better highlight the process and its attendant problems.

⁶⁴ An example of this approach would be the recognition of different species according to a DNA ‘barcode’ (i.e. a sequence of DNA which is more or less specific to each species). However, even with this there are difficulties, for the same problems of marginal and borderline cases will arise (where common ancestors are of different species) – see Amos (2005) and Lever (2008). Incidentally, this is not a fatal point for our contentious example above; ‘sharing a common ancestry with both cat and dog’ can easily be added as one of the features possessed by animal ‘X’, thus preserving its ambiguous status.

On the other hand, however, I do not think that Beardsmore is proposing a different, revised-essentialist view; he is merely pointing out that the theory of family resemblances, as a form of watered-down essentialism (for 'single common feature' read 'group of interconnected common features'), is just as unsuccessful in accounting for what general terms have in common as traditional essentialism is.⁶⁵ In this way, it can therefore be seen that many forms of conservative conceptualism are no more than the ghostly remnants of conceptual realism.

At this point it may be objected that in seeking to arrive at an absolute definition of what a 'game' is, we are merely dealing with a very general cultural phenomenon (as opposed to a strictly scientific one), and that we should not be surprised at the lack of uniformity or precision. This is an especially important point when we come to consider the nature of the boundary between life and death. When we talk of games, or any other cultural phenomenon, can we clearly distinguish between culturally relative concepts and 'scientific' ones? If so, then perhaps it may be possible to arrive at a strictly-biological definition of death; but if not, then the concept of death would appear to be open to the same mutability and variance that cultural concepts have. For instance, the 1975 film *Rollerball* envisages a popular sport where death is an accepted part of the game – hence the tagline: 'The next war will not be fought – it will be played.'⁶⁶ However, surely the same cannot be true of death? In other words, there must be something – the conceptual realist or conservative conceptualist would like to think – which limits and determines our concept of death (and therefore distinguishes the search for a correct definition of 'death' as a distinct *type* of pursuit from that of the search for the significant features which define what a 'game' is).

Of course, this is simply the view mentioned above (proposed by Quine) that we may arrive at theoretically robust kinds of thing through the critical development away from our first, instinctive notions of category (in other words, through scientific and empirical investigation, which establishes more significant resemblances than mere physical appearance – for instance). The significant features thus identified therefore seek to provide a non-relative basis for category membership.

But is such an approach really immune to the anti-essentialist criticisms which Wittgenstein's comments embody? In other words, is it possible to identify just those significant features which constitute the concept of 'death'? If construed within an analysis of our linguistic practices, then it would seem that what 'death' is, and what 'games' are, represent similarly problematic questions. However, as I have suggested, the possibility that the question concerning the definition of death may be resolved in a way that is not open to the question of what a game is rests on the idea that the two questions are of *different types*. In seeking to defend a position which argues that 'death' may be uncontroversially defined in a way which avoids the problems faced in

⁶⁵ It is difficult to say what position Beardsmore holds, exactly, since his arguments are largely focused on discrediting Bambrough. However, it is safe to say that he is, in following a certain interpretation of Wittgenstein, defending a non-essentialist position.

⁶⁶ *Rollerball* (Jewison, 1975). However, this is a now ubiquitous dystopic theme in film and literature, and other examples abound – e.g. *The Running man* (Glaser, 1987), which marries futuristic game-show and corporate-fuelled despotism, and *Tron* (Lisberger, 1982), a virtual reality twist on the theme. However, such science fiction treatments are, to a significant degree, reworkings of one of long-established death games: the Roman gladiatorial arena.

defining ‘game’ (i.e. the strictly-biological view), we are therefore seeking to privilege the former pursuit – but on what basis? In the following section, I shall argue that such a move is inherently problematic.⁶⁷

Ontological Relativity and Opacity of Reference

So far I have argued that the problem faced in defining death can be fruitfully understood as an example of the more general problem of category membership. In light of this, I have shown that different resolutions to this problem are possible depending on the conceptual approach adopted (i.e. conceptual realist, conceptualist, or nominalist). However, I have then argued that both conceptual realism and conservative forms of conceptualism share an essentialist approach which leaves them open to certain well-established criticisms. It is this essentialist commitment, I argue, which is at the bottom of many current attempts to define death (especially the strictly-biological approach). In this section, therefore, I want now to provide further reasons for rejecting essentialism, which I will also argue explain the nature of the definitional problem itself. It is these reasons, which stem from Quine’s notions of the ‘opacity of reference’ and ‘ontological relativity’, which I argue provide further corroboration of my criticisms of essentialism, and also show why a conceptual reduction of ‘death’ to strictly-biological features (and a resolution of category conflict via that method) is – at best – highly problematic.⁶⁸

In his essay ‘Ontological Relativity’, Quine points out that the only successful and coherent way in which we can identify guiding principles or criteria (such as the verification principle of the logical positivists purported to be) is where one conceptual scheme is discussed in relation to another one:

What makes sense is to say not what the objects of a theory are, absolutely speaking, but how one theory of objects is interpretable or reinterpretable in another.⁶⁹

Thus, there is always, for Quine, a ‘background theory with its own primitively adopted and ultimately inscrutable ontology’.⁷⁰ Thus, we may interpret a conceptual scheme within a background theory (such as we may identify what the chemical constitution of a substance is), or even explicitly reduce one conceptual scheme to another one (such as chemistry has been reduced to quantum physics); however, what we cannot do is seek some absolute reduction of the background theory to rational first principles *which do not in themselves involve irreducible, non-rational assumptions*.

If this is true, then the implications for the definition of death are considerable. Firstly, there would seem to be nothing to privilege a strictly-biological definition over

⁶⁷ I shall consider Putnam’s notion of the so-called ‘entanglement of fact and value’, and its relevance to this issue, in chapter 7.

⁶⁸ In the following discussion, I do not attempt to defend Quine’s theories against all his critics – though I consider some. Rather, having noted some strong criticisms of essentialism, I try to show how an alternative conception of language (such as an antirealist or nominalist might favour), would allow us a clearer understanding of the problem we face in defining death.

⁶⁹ Quine (1969a), p.50.

⁷⁰ *Ibid.*, p.51.

competing theories (such as a personal-identity theory). Secondly, if the attempted reduction leaves out what others take to be significant features for the purpose of a clear, unambiguous, strictly-biological definition (e.g. ignoring persistence of spontaneous function), then the reduction could also be considered an unsuccessful one. Thirdly, if it can be shown that the attempted reduction retains features which are not strictly-biological (as I will argue the notion of function is in the next chapter), then the reduction will also fail on this count. This last point brings to mind the other well-noted problem that besets such reductive enterprises – i.e. the difficulty involved in distinguishing between analytic and synthetic properties.⁷¹

However, all this talk of systems, theories and schemes has led some philosophers to criticise the intelligibility of this picture. Donald Davidson,⁷² for instance, has pointed out that such philosophers as Thomas Kuhn, Paul Feyerabend, and B. L. Whorf (and, of course, Quine himself), have argued that there can exist differences between conceptual schemes that are so great as to represent an incommensurability in terms of translation. Thus, for instance, Whorf⁷³ has argued that certain aspects of the Native-American Hopi language suggest an ontology different to that of English, and which therefore represent a metaphysics which cannot ‘be calibrated’ with our own.⁷⁴ Furthermore, such a picture provides an antirealist backdrop to the way in which such concepts as death and self evolve, and accordingly what status they have.

The categories and types that we isolate from the world of phenomena we do not find there because they stare every observer in the face; on the contrary, the world is presented in a kaleidoscopic flux of impressions which has to be organized by our minds — and this means largely by the linguistic systems in our minds. We cut nature up, organize it into concepts, and ascribe significances as we do, largely because we are parties to an agreement to organize it in this way — an agreement that holds throughout our speech community and is codified in the patterns of our language.⁷⁵

Of course, the above picture is not necessarily tied to the notion of differing conceptual schemes that Whorf and Quine favour. We may, in fact, consider the above merely as a picture of *the whole* of language (in which case it takes on a Wittgensteinian flavour). As such, it is an antirealist view which would allow for a liberal conceptualist or nominalist perspective. However, let us first consider the relevance of differing conceptual schemes for the general argument.

Davidson is sceptical of such methods of analysis and the very notion of differing conceptual schemes. For instance, he argues, if Whorf uses ‘English to convey the contents of sample Hopi sentences’, and ‘Kuhn is brilliant at saying what things were like before the revolution using – what else? – our post-revolutionary idiom’, then

⁷¹ The seminal treatment of which is, of course, Quine (1961).

⁷² Davidson (1984).

⁷³ Whorf (1956a); also, see Whorf (1956b), where he concludes (similarly to Quine) that the grammar of a language may be relative to that language (and therefore distinct from others). This idea is known as the Sapir-Whorf hypothesis, as it is generally considered to be a position that first put forward (independently) by Whorf and Edward Sapir.

⁷⁴ Quoted in Davidson (1984), p.184.

⁷⁵ Whorf (1956b), pp.213-4.

isn't there an inherent contradiction involved here? In other words, if we can understand the 'alien' ontology enough to know that it is different, then don't we simply understand their ontology? If this is true, then the differences are not really those between incommensurable conceptual schemes, but rather those which may exist between people holding different beliefs. As Davidson points out, in any situation where communication takes place there must exist a charitable interpretation of another's utterances. As he puts it:

charity is not an option, but a condition of having a workable theory [...] whether we like it or not, if we want to understand others, we must count them right in most matters.⁷⁶

Thus, for communication to be possible at all, there must be a common ground of shared beliefs. So, whilst, as the use of the word 'charity' implies, there is obviously room for error and misinterpretation, 'it may be more natural to speak of differences of opinion' rather than radically different conceptual schemes.⁷⁷

Davidson's point here is a sound one, I think. However, I do not think that in rejecting talk of conceptual schemes we need also dismiss the aspects of the idea of ontological relativity which – I will shortly argue – can be used to bring out important features of the controversy concerning the definition of death.

Firstly, Quine's notion of the 'indeterminacy of translation' can be seen to involve the related, but not inseparable notion, of 'inscrutability of reference'; it is the latter which leads, it may be argued, to the former. Thus, because terms are, beyond a certain point, inscrutable and cannot be analysed further (or at least, not without involving further terms which are themselves open to analysis), we may be sceptical about whether terms to be translated *from* another language are indeed equivalent to certain terms in the language *into which* they are to be translated. So, Quine's famous fictitious example involves having to translate the foreign word 'gavagai' into English. Faced with certain clues, and through interrogation of native speakers, the translator settles on 'rabbit' as an apt translation. The problem, however, for the translator, is that, unbeknown to them, 'gavagai' actually means 'rabbit part' and not simply 'rabbit'. However, because the translation of it as 'rabbit' seems to fit all uses of the term 'gavagai', the discrepancy is never noticed. Furthermore, even if it were, there is nothing to preclude the existence of other discrepancies, which ultimately stem from the utilisation of an ontology alien to ours. But how to test for this ontology? We cannot, Quine argues, because in doing so we simply supply our own. On one level, of course, Quine's point (like Wittgenstein's parallel point about rule following)⁷⁸ concerns difficulties involved in ostensive definition: the translator infers the meaning of 'gavagai' from seeing it used by native language users. However, there is a deeper point here (and with Wittgenstein also) which applies to the general use of terms in *any* language: no matter how explicit we seek to be in our explanation and definition of such terms, there is always the possibility that there is a lurking 'opacity' which certain circumstances will reveal (I shall return to this shortly).

⁷⁶ Davidson (1984), p.197.

⁷⁷ Ibid.

⁷⁸ See Wittgenstein (1967), p. 59 (section 151) and following, and also the discussion of this point in chapter 1 above.

One difficulty in Quine's argument should be noted at this point. How can the ontologies be different if, at bottom, the terms of reference are inscrutable? In other words, what is there for the two respective ontologies *to differ about* if there is no 'ultimate fact of the matter' concerning the reference of terms? If you say 'gavagai' and I say 'rabbit', and in every practical situation we agree in our use of the term (i.e. it throws up no confusion or conflict of which either party becomes aware), then where in fact does the difference in ontologies lie? The difference between the two ontologies envisaged by Quine may therefore involve a misunderstanding which it is not impossible to resolve, but merely highly difficult. It therefore represents a practical epistemological hurdle – i.e. a problem concerning how I might be sure – rather than one which it is impossible to resolve *in principle*. I may go my whole life believing that, by 'gavagai' and 'rabbit' respectively, we mean the same things. However, because that situation has not arisen does not mean that it *could* not arise.

The issue here, then, is whether epistemologically, Quine's gavagai problem represents a mere difficulty, or an impossibility. In other words, is it just unlikely that we may find out, or is it in fact out of the question on logical grounds? It would seem that Quine wants to argue for the latter case. If so, then there would seem to be a further difficulty: how does he know? For, paradoxically, if there is an 'alien ontology' that we can, by definition, never be a party to, then how can we know that it is incommensurably different from our own? Surely this implies knowledge, at least to some extent, of the alien ontology in question?

In utilising Quine's arguments for my own purposes I will therefore be careful to distinguish between these two forms of interpretative difficulty: the first involves an epistemological difficulty which it may be possible, under certain circumstances, to resolve; the second type involves an impossibility which, by definition, cannot be overcome. The first position, I argue, is simply a natural consequence of the inscrutability of reference, whilst the second cannot be held without the type of contradictions just mentioned. The second position not only implies, paradoxically, that we can have knowledge of this 'alien' ontology (in order that we may *not* understand it), but also that there exists, independently and unknowably behind the observable actions of the other-language user, a shadowy ontology. The significance of this last point is that Quine holds a behaviourist-type view of language and how it works. Thus, how can there be an ontology which is not reflected in the thought and (potential) behaviour of the other-language user? Surely, if this presented any significant problem for equivalence of terms, it would eventually be revealed in the interactions of the two sets of language users?

For these reasons, and in line with Davidson, I therefore hold the notion of competing conceptual schemes to be, at best, problematic. However, for my purposes, I believe that some version of Quine's notion of the inscrutability of reference is important to an understanding of the definition of death debate. Firstly, however, it is important to establish that the notion of the inscrutability of reference does not live and die with the notion of ontological relativity. We do not need to propose the existence of different ontologies, and consider translation between different 'languages', in order to show that there are times when people mean different things by certain terms. On occasions, this misunderstanding can be cleared up through further investigation and communication. However, on other occasions, the discrepancy does not become evident until very special circumstances reveal it.

For instance, imagine that science were to develop the technology for brain transplants. Consequently, there would then arise controversies as to the identity of individuals who possessed the brain of one person and the body of another; more intriguingly, imagine that transplantation of *half* of someone's brain were possible, and that one half might live in one body, and the other half in another; or even that complete copies of individuals could be made so that the copy retained the original's memories.⁷⁹ In such situations, where once there was agreement as to the use of a term (in this case, 'her', 'yours', 'Mavis', 'person', etc.), now there is divergence. Different parties will supply different arguments as to the defining criteria of personhood in order to evaluate the status of the different cases. Before the advent of the technology, however, the use of the terms had remained inscrutable, and the divergence latent.

It is tempting here to assume that there is a pre-existing explicitness about these terms which certain situations bring out (e.g. as a conceptual realist might hold). Thus, for instance, in the brain transplant scenario, one advocate always, in reality, held that the body was the important thing in determining identity, whereas the other always secretly held that it was the brain. Thus, when circumstances allowed these two previously-conjoined things to exist separately, the respective 'secrets' were revealed (just as, in Wittgenstein's example of following a rule discussed in chapter 1, the two individuals would at a certain point diverge in their answers as the two divergent applications of the rule were revealed; or, in Quine's scenario, a situation arose in which 'gavagai' and 'rabbit' could be seen to mean different things).

Though this may, on occasion, be the case, it is not necessarily so. Sometimes – and this, I think, is the situation regarding the definition of death – situations arise which are new and unforeseen. Thus, the new development provokes a divergent interpretation in a term of reference (in this case, 'being dead') which, up until the advent of the new situation, did not in fact contain a 'secret' divergent application, but was in effect genuinely 'opaque'. Thus, in relation to the determination of death, the advent of new medical technology and advances in medical care have allowed such previously unthought-of scenarios as a 'brain dead' individual, some of whose spontaneous functioning is artificially maintained. It is such advances in technology, coupled with the inscrutability of the terms of reference used, which have allowed for a 'schism' to arise in what we mean by 'death'. Furthermore, such inscrutability is not even theoretically avoidable. Firstly, as Quine has argued, terms can only be cashed out relatively; so, whilst we may define terms in relation to other concepts (and so make them more explicit), we may not do this absolutely (that is, we will always reach a point where our definitions seem arbitrary, or otherwise rely on the irrational). Secondly, given that we do not know the future, we cannot know in what way our current understanding of terms may be lacking. The most obvious example of this is technology, where the possibilities it throws up are sometimes completely unforeseen (so, for example, it would have been difficult and extremely unlikely for us to have foreseen and incorporated into our use of the term 'being alive' the consequences of advancements in medical care).

The Redundancy of Categories and the Death of Natural Kinds

⁷⁹ These examples are taken from Parfit (1984).

Before rounding off this chapter, I would like finally to consider one possible solution to the problem of categorization. Earlier, I outlined the way in which the development of robust scientific categories (theoretical kinds) stems ultimately from the refinement of intuitive natural kinds. However, the development of such scientific standards may lead ultimately, Quine points out, to the disappearance of the notion of kinds altogether. For instance,

One can define water-solubility by simply describing the structural conditions of that mechanism. [...] once we can legitimize a disposition term [such as “water-soluble”] by defining the relevant similarity standard, we are apt to know the mechanism of the disposition, and so by-pass the similarity.⁸⁰

In other words, knowing precisely in what way things are similar (are of a particular kind) leaves us in a position to dispense with the notion of a ‘kind of thing with certain properties’ (e.g. water-solubility) in favour of a reductive explanation of that similarity in terms of more fundamental composition and interaction (e.g. molecular structure and behaviour). On the other hand, taxonomies that still find the retention of kind terminology useful can employ such terms in a fully informed way by virtue of the mere existence of an underpinning reductive explanation. For instance, in utilization of the term ‘species’, zoologists or biologists

are in a position to define comparative similarity suitably for this science by consideration of family trees. For a theoretical measure of the degree of similarity of two individual animals we can devise some suitable function that depends on proximity and frequency of their common ancestors. Or a more significant concept of degree of similarity might be devised in terms of genes. When kind is construed in terms of any such similarity concept, fishes in the corrected, whale-free sense of the word qualify as a kind while fishes in the more inclusive sense do not.⁸¹

This, then, is the goal of science – but is it one which is universally realisable? One caveat noted by Quine is that such a programme of reduction as makes the concept of kinds scientifically robust is only achievable in a relative sense (what he terms ‘ontological relativity’). In other words, the disposition of physical objects (e.g. water-solubility) is reducible to a mechanistic explanation only where the criteria of similarity can be explicitly stated in terms of the explanatory theory (e.g. the language of chemistry). However, for the purposes of a more fundamental reduction, the door is left open: chemistry may of course be explained in terms of more fundamental processes (e.g. the language of physics – hence the old quip that ‘chemistry is just a branch of physics’), but that in turn begs another reduction, and so on.

I have already dealt with why I think it is impossible that anything but a relative reduction is possible. However, I would like now to consider whether a more modest form of reduction might not be of benefit in relation to the determination of death debate. Firstly, as I have already implied, such reductions may be of use in scientific

⁸⁰ Quine (1969b), p.135.

⁸¹ *Ibid.*, p. 136.

contexts where the sort of explanations they enable can in fact replace categorical explanations. So, for example, an evolutionary biologist seeking to understand a particular animal may find that it falls between two categories. However, in understanding it in terms of its genetic inheritance, it is possible to do away with the categorical problem altogether. So, instead of asking 'is or isn't it a mammal?' the biologist need only concern himself with its genetic family tree (some members of which were mammals, some not), whilst treating the category of 'mammal' merely as a term of convenience (useful, but ultimately theoretically inadequate). Thus, the categorical problem (In which category, by virtue of its traits, does that animal belong?) can be circumvented by an understanding of the more profound reasons why the animal comes to possess those traits in the first place.

When applying this method as a possible solution to the definition of death debate, it is important first of all to realise that it does not necessarily enable us to resolve category membership. It may enable us to refine certain categories (such as affiliating an animal previously thought dissimilar with another category based on common ancestry), but this only applies when we already know what enables membership of that category (i.e. the possession of significant features). However, since this is the issue in question, the reductionist method is of little help here. Secondly, all this method could do, essentially, is to describe certain states (the possession of certain characteristics), in terms of more fundamental processes (e.g. the interaction of biological systems). But whether these processes represent life or death would in turn be an open question that a reductive explanation could not resolve, in as much as we would have to have already decided what death was in order to understand that that was what certain processes represented. Ultimately, then, the reductive method is, at best, of value only in clarifying how already classified processes work (i.e. how putrefaction takes place), and at worst, indifferent to the whole categorization process. As such, the method cannot help us understand why one state should be called 'dead' and another 'alive' independently of previously-taken categorical decisions.

If the above is true, then what we are witnessing, in effect, is the death of the notion of natural kinds. However, this is not a demise accepted by everyone. For instance, David Wiggins has argued that there are 'lawlike principles'⁸² which govern the identity of certain kinds. This, Frederick Doepke has pointed out in his defence of Wiggins, would ensure that

no such thing could become a member of a different kind. No body of water or bit of Gold, for example, could change its atomic structure, and no human person could become a member of another species.⁸³

Wiggins's position may therefore be seen as a combination of a realist belief in the mind-independence of certain entities (i.e. that such things as trees, water, gold, etc., actually exist as natural kinds of thing), combined with the more conceptualist admission that the form of our concepts is determined by human perceptual limits.⁸⁴ This view, therefore, represents

⁸² Wiggins (1980), pp.77-86.

⁸³ Doepke (1992), p.89.

⁸⁴ *Ibid.*, p.91. For my purposes here, since they both defend similar views, I shall consider Doepke and Wiggins together.

a common-sense metaphysics in which continuants [between different candidates for examples of the same kind] are known directly. Stones and trees and dogs are things that we know in the first instance simply by perceiving them. We do not, in this metaphysics, first perceive stages of them, and then go on to string these together, as we first perceive people and then go on to form concepts of groups of them, such as tribes, committees and nations.⁸⁵

The approach of Wiggins and Doepke, therefore, is to resist the temptation to consider *any* ‘continuants’ from the natural kind as first perceived as being significant. So, it is not just that persons A and B have red hair, freckles, are called Smith, etc., that make them the same person, but the fact that the survival of such features (‘continuants’) are linked in a lawlike way (nomologically). Thus, if we may identify those continuants which are governed by lawlike principles, then we have also succeeded in identifying those features which are significant to the identity of the natural kind in question.

In other words, imagine that we wanted to establish whether person A was the same as person B, where A had gone into a teletransporter and B had come out (to use Parfit’s example).⁸⁶ According to Wiggins and Doepke, what would be sufficient for identity would be the presence of features in B which were ‘continuants’ of features in A. However, these could not just be any features, but specifically those features which were essentially A in the first place. But what are these features? Wiggins’s answer is that

the determination of a natural kind stands or falls with the existence of lawlike principles, known or unknown, that will collect together the extension of the kind around two or three good representatives of the kind. In order for the name to stand for a natural kind, everything depends on whether there is some nomological grounding for the application of the name. If there is, and if the predicate is worthy to survive as a natural kind term, then the holding of the relevant principles is constitutive of its exemplification by its instances. To be something of that kind *is* to exemplify the distinctive mode of activity that they determine.⁸⁷

Wiggins’s notion of natural kind therefore employs a ‘pre-empiricist notion of substance’, whereby we can simply

pick out (say) a tadpole and identify its kind by saying ‘the members of the kind are anything relevantly like this’, while leaving the actual standards of relevance to be scientifically determined at macrolevel and microlevel⁸⁸

We are back, then with the notion of intuitive kinds considered earlier. We begin with an intuitive notion of a thing, which is then refined through ‘standards of relevance to be scientifically determined’. In our teletransporter scenario, therefore, what defines

⁸⁵ Ibid., p.91.

⁸⁶ Parfit (1984), p.199-200, and subsequent sections.

⁸⁷ Wiggins (1980), p.80.

⁸⁸ Ibid., p.83.

person A are these scientific standards of relevance. For instance, we might argue that the atoms of A were exactly reconstituted in B in the same configuration, and that because of the lawlike connection which governs the movements of the atoms between A and B, there is also identity.

In relation to the concept of death, we may apply Wiggins's views as follows: person X is alive because he possesses features which are relevantly similar to other living people; conversely, he is dead because he lacks those features. But what, here, are relevant features? It may be all very well to argue that our intuitive notion a natural kind can be scientifically determined at a later date when what we are talking about is a tadpole (though even here it may be problematic, as already noted in the above discussion of species); however, where we are speaking of a human being, then the role of science in 'filling in the details' of what makes a natural kind distinct is highly problematic. Firstly, as argued above, such a move would seem to privilege scientific discourse above others; secondly, as also noted, it would imply that we can in fact clearly distinguish between those features of a living human being which are amenable to scientific analysis, and those which have some cultural basis (such as personal identity, the notion of self, and so on). So, rather than reviving natural kinds, such views as those of Wiggins and Doepke merely seem to rely on the hope that our intuitive notions of kind can be cashed out in purely scientific terms. In the next chapter, however, I will consider important reasons why I think scientific analysis cannot ultimately help us with the notion of natural kinds – at least, not in a way that is divorced from the general marketplace of cultural concepts. Furthermore, I will later argue that it is this kinship with non-scientific and cultural concepts that must eventually allow for a much wider conception of death than has previously been admitted.⁸⁹

Conclusion

Throughout this chapter, I have attempted to view the problem of the definition of death from the perspective of conflicts involving category membership. Consequently, I have tried to show that, in relation to death, this search can be seen as the attempt to identify those essential features which determine death as a natural kind. However, I suggest, our concepts have not been built up from rational first principles, but have rather evolved from intuitive kinds. A consequence of this is that there would appear to be no essentialist fact of the matter as to what 'death' is: the criticisms of essentialism (and revised essentialism) are quite conclusive, and – I argue – we must reject both conceptual realism and the conservative variety of conceptualism which attempts to maintain the 'ghost' of essentialism.

The failure of essentialism (I argue) also in turn necessarily leads to the failure of the concept of natural kinds: if we cannot identify essential features, then we cannot identify natural kinds. We may, of course, argue that this failure arises as a consequence of our inability to directly perceive the true nature of death (which is nonetheless an intrinsic feature of a mind-independent reality) – the possibility that

⁸⁹ Of course, neither Wiggins nor Doepke is explicitly proposing a definition of death here, and – to be fair – their views might also be used in conjunction with a personal-identity theory of death, and are therefore not necessarily tied to a strictly-biological definition. However, as I will later show in chapter 6, such a view is no less problematic.

this is true of death is something that I consider in the next chapter. However, in terms of our notion of death, it would seem that there are no conclusive ways in which we can fix the form of the concept – or, in fact, maintain its binary opposition with ‘being alive’ (hence allowing for the possibility of the existence of genuine intermediate states, or other possible conceptions). But does this leave us with nominalism? Not necessarily so, and in later chapters I will explore what I take to be the consequences of a non-essentialist conceptualism for the definition and application of the concept of death.

It is as a direct result of this defeat of essentialism – whether or not the doctrine of ontological relativity is true – that something like Quine’s ‘opacity of reference’ may help us to understand how problems such as we face in the definition of death can come about: previously ‘opaque’ terms (as all terms must be at some level) are made ‘explicit’ in unforeseen ways, and which set their previously interwoven features in potential opposition to one another. Furthermore, this referential opacity may also be seen to spring from the nature of concepts themselves. Concepts arise as ‘intuitive kinds’, and it is *because* they evolve from the ultimately non-rational grounds of human culture (as Nietzsche, Quine, and Wittgenstein would agree) that they are opaque: therefore there can be no ultimate, rational ‘fact of the matter’, no completely successful reduction, and ultimately – in a conceptual realist’s sense – no such thing as a natural kind.

In the next chapter, I will attempt to conclusively show that – even if we adopt a metaphysical realist’s ontology, and even if the above problems were ignored – the concept of death must be considered as relative to human culture.

4. The Status of Death

Introduction

In the previous chapter, I outlined the problem concerning the definition of death in terms of a category conflict resulting from an essentialist position concerning the problem of universals. I also argued that whilst this sort of conflict is not confined to the definition of death problem, there are aspects of it which are peculiar to it (or, perhaps, peculiar to the *type* of concept that death represents). For instance, I have tried so far to show that there are traditional (or at least logically possible) ways of resolving such category conflicts which have been largely ignored by philosophers in relation to this context due to underlying and unexplored metaphysical assumptions (i.e. conceptual realism or – more likely – an essentialist conceptualism). For this reason, the problem in determining death is still commonly presented as a search for significant criteria which are sufficient to disambiguate the status of problem cases in favour of either one category or another ('alive' or 'dead').

In saying this, however, it should be noted that I am not necessarily proposing that other, intermediate categories should be introduced into the scenario; nor am I essentially interested in identifying the set of significant criteria that render the status of such cases unambiguous; rather, I am concerned to show that, if we are to understand the nature of this debate, we need to analyse why it is that the majority of approaches to the subject still treat it as one which has to be solved in terms of binary, mutually exclusive categories. And yet, as I hope I have shown, the essentialist assumptions upon which it is based are at best problematic, and that a more fruitful understanding of the issues may be arrived at if we abandon the search for essential features, step back, and attempt to understand the nature of the problem. I therefore argue that this suggests that the desire to maintain a binary opposition between life and death is not based on any logical constriction, but rather on a non-rational (e.g. cultural or moral) need.

I have supported this approach by pointing out that the main method that has been suggested might resolve this problem (the strictly-biological definition) is itself problematic. For instance, in seeking to identify criteria which are sufficient for determining that someone is dead, we already seem to assume that not only is death a strictly-biological fact, but also that it is possible to define it in terms which are independent of any value-based concerns that we may have.¹ In this current chapter, I wish therefore to settle the question of what *sort* of concept death is. I will do this by attempting to show that, even if we adopt a realist ontology (such as that proposed by John Searle), death must still be considered as a human-dependent, 'observer-relative' aspect of reality.

In doing this, I will therefore consider various ways in which death might be considered to be an 'intrinsic' feature of the world, and attempt to show that – no matter what features of the definition of death are proposed as being intrinsic – there

¹ I dealt with this latter point somewhat in the last chapter where I argued, in line with Quine, that any principle which could be utilized for the purposes of clearly distinguishing between fact and value must in turn be either itself partly value-based, arbitrary, or irreducible (i.e. the idea of ontological relativity).

is a necessary sense in which the significance of such features is always supplied by a human agent.

Incidentally, some might wonder why I have taken this route: might it not be simpler to argue for an antirealist position, and show that therefore there are no intrinsic features *at all*? Whilst I take this point, there are good reasons for also rejecting it: firstly, presenting a case for rejecting realism outright would be a more major undertaking than merely showing that death is not an independent feature of the world; secondly, the proposition that all features of everything are socially constructed (or 'observer-relative') is too strong a case for the consideration of the status of death – and one that I might not necessarily wish to hold. My point, rather, is that, even if we adopt a realist position such as Searle's, Death cannot be considered an intrinsic feature of it; needless to say, therefore, that neither would it be an intrinsic feature of an antirealist point of view (since it does not contain any). So, whilst I do not necessarily hold to Searle's position, it is a useful way of answering the question of the status of death *regardless* of the position held in the realist/antirealist debate.

Finally, it should be noted that in this chapter I deal with the notions of metaphysical realism and the question of the existence of death as a mind-independent property of the real world. Therefore, these questions should not be confused with the debate concerning conceptual realism and antirealist approaches to concept formation considered in the previous chapter. There, I dealt with the concept of death; here, with the idea that there might exist something independent of our concepts, and to which our concepts refer.

Some Useful Distinctions

I think the first question that we need to answer in investigating these issues is what status we think death has. To aid in doing this, I am now going to introduce some distinctions used by John Searle in his book, *The Construction of Social Reality*.² Searle is what may be termed a *realist* in respect of certain issues, in that he believes that there is a clear sense in which the world exists as a human-independent reality. A consequence of this view is that Searle believes that the features of the world can be grouped into two classes: those which are *intrinsic*, and those which are *observer relative*. Intrinsic features include many of the things that science considers to be truths, such as the mass of a physical object, its chemical composition, etc. On the other hand, the observer-relative features of a physical object would include such things as its function, its aesthetic appeal, its monetary value, etc. Intrinsic features can therefore be termed *ontologically objective* in that they exist independent of individuals or human society. Observer-relative features, however, could not exist without a potential observer (or society of language users) within which such features are acknowledged. In this sense, such features are *ontologically subjective* in that they exist only as a consequence of the subjective ascription of meaning or function.

To illustrate this simply, in a world that did not contain humans, there would be no goals, tax returns, beauty, or freedom; however, there may still be nuclear radiation, carbon dioxide, trees, and thunder storms. The distinction, therefore, is between features which are inherent in physical objects, and those which are ascribed to it

² Searle (1995), pp.7-13.

solely in the course of human interaction with them. It may be objected at this point that our knowledge of physical objects is observer relative in that ‘inches’, ‘miles’, ‘seconds’, ‘kilograms’, etc., would not exist without human culture. However, whilst this is true, Searle is careful to differentiate between a physical phenomenon’s ontologically-objective features, and the ontologically-subjective terms used to describe them. So, whilst there have not always been humans around to measure the distance from the earth to the moon (in miles or kilometres), it does not follow that there has not always been some distance that existed between here and there. This is an important point for the definition of death: for death to be considered to be an intrinsic feature of the world (and ontologically objective), there must be a sense in which our ontologically-subjective conceptions *correspond* to it. However, if death is not an ontologically-objective (intrinsic) feature, then there is nothing for our terms to correspond to. In this sense, ‘death’ would be a social construction, or contingently relative to some other factor which influences concept formation.

A further distinction which Searle draws is between *epistemically-subjective* and *epistemically-objective* statements. An epistemic statement concerns knowledge and is subjective or objective according to whether it involves opinion or fact (put crudely). So, ‘Mount Snowdon is beautiful’ is an epistemically-subjective statement (it involves opinion), whilst ‘Mount Snowdon is the tallest mountain in Wales’ is epistemically objective (it involves fact).³

Utilizing these terms, we may neatly categorize everything from quasars to cricket matches. For instance, consider in football the concept of a ‘goal’. Since football would not exist without human beings and their social and linguistic practices, then I think it is safe to say that the definition of a ‘goal’ falls into the category of observer-relative things. Furthermore, since a goal is something which may be decided conclusively by empirical evidence, then we may also consider it to be epistemically objective. This is not to say, however, that people do not disagree about whether or not a goal has been scored – of course they do. However, in this case, their disagreement revolves around the determination of matters of fact, and not about subjective preference or ascriptions of value. To conclude, then, the decision that something is a goal is an epistemically objective statement about an observer-relative and ontologically-subjective event.

But where then does death stand? In terms of a medical/scientific model of death, it is tempting to argue that that it is an intrinsic feature of the world which would exist whether there were human beings or not (so, animals, plants, or perhaps even stars, may be said to die without there having to be a human concept of death to apply to them). In this light, the present confusion over the definition of death under certain conditions would be on a par with certain problems within quantum physics or genetics; in other words, we may not be certain how to resolve them at the moment, but their ultimate resolution is a scientific concern – i.e. a matter of ascertainable fact.

³ I realise that there are other accounts of the notion of fact than the one that Searle assumes here (for instance, which question the very notion of independent states of affairs). However, my purpose here – as already stated – is not to defend Searle, but to use him as a framework for discussing the potentially intrinsic status of death within a realist’s framework (i.e. it may turn out that, even if such a realist perspective were true, death may still not be considered an objectively existing, intrinsic feature of the world). However, I will return to this question in relation to Searle’s views in chapter 8.

But may death, in Searle's sense, be considered an intrinsic feature of the world?

Death as an Intrinsic Feature of the World

Death has long been defined according to the cessation of certain vital functions – heartbeat, respiration, and now (somewhat controversially, admittedly) brain function. In this sense, it has involved what may be termed the *functional definition*. Furthermore, in as much as a strictly-biological definition of death is reliant upon cessation of function, then we may equate these two definitions. If we are to assess this functional definition, or some variant thereof, then we therefore need to establish whether the notion of function itself can be considered an intrinsic feature of the world. When we talk about the function of the heart, for instance, are we describing intrinsic features of a physical object, or are we simply attributing an observer-relative function to it?

An example of an observer-relative function would be where we term a particular arrangement of bits of wood (or other material) a 'table'. There is nothing intrinsic about the standard function of this object, and we may in fact use any suitable surface as a table. Other things, however, seem already to possess function – such as, for instance, the human heart. In a sense, then, the functions of such things as the heart appear to be 'discovered' rather than attributed. In Searle's terminology, such a discovery is an example of 'non-agentive' function. In other words, its function has not been ascribed to it by a particular agent (as when I use any flat surface as a table), but is suggested by the interrelation of its parts in a causal way. Searle, by the way, argues that even though such functions are 'discovered', they are not ultimately intrinsic features of the world, because we ascribe value to what are simply causal systems. That the heart pumps blood is a fact; that its *function* is to pump blood, is a value-based ascription of a teleological purpose or role. This is an important and subtle distinction: Searle is not arguing that there is no difference between biological objects, other naturally occurring things, and manufactured items – between, for example, an eye, a stone, and a desk – but merely that we attribute function to each of them in different ways. A desk is intentionally designed to serve a particular purpose; a stone may, by chance, possess qualities which make it useful for some purpose (e.g. a large, flat stone may make a good writing surface); however, an eye, by virtue of the causal interrelation of its parts, is said to fulfil the function of seeing. However, whilst differences for the reason for the ascription of function exist, it is nonetheless true in all cases that the function is ascribed, *and not inherent in the object*. In other words, without human society to attribute function and purpose to these objects (or to 'discover' function or purpose), they would have no inherent function (whereas, they would still possess mass, structure, a causal interrelationship of parts, and other physical qualities).

Searle's assertion here is at odds with a number of approaches which seek to account for the sort of teleology we find in non-agentive function; approaches which, in Mark Bedau's phrase, attempt to 'sanitize' what he and Searle see as the value-based ascription of function (i.e. the observer-relative aspect).⁴ This generally amounts to seeking to replace the appearance of value (or teleological explanation) in the function

⁴ Bedau (1992). For the purposes of this discussion, I will utilise Bedau's threefold characterisation of the approaches he identifies and also some of his criticisms of them.

with some 'purely descriptive form of causal analysis'.⁵ We may identify, here, three main approaches to this problem: the mentalist approach, the systems approach, and the etiological approach.⁶ All three approaches attempt a reduction of the notion of function into terms which avoid the ascription of teleological purpose, either by a human agent, or in a religious sense (i.e. divine design); function, then, it is argued, is a natural feature of the world. I will now evaluate these three approaches with a view to assessing the general plausibility of the idea that function can be in some way intrinsic.

Three Arguments for Intrinsic Function

Firstly, the mentalist approach attempts to account for teleology in terms of the rational purposes of a designer. So, a teleological explanation of the function of a kettle would account for it in terms of the intentions of its designer. However, when we come to non-agentive functions (i.e. ones which are merely 'discovered'), we find the mentalist having to say something like, 'Something has function if it *appears* that it has been designed by a rational agent for a certain purpose.'⁷ This may be called an 'as if' type of argument, in that non-agentive functions are looked upon *as if* they were designed. So, whilst there is not necessarily any suggestion of the existence of a designer (e.g. God), the idea of function is analysed as the *as-if* existence of features which, if someone *had* designed them, could be said to represent intelligent craftsmanship. This proposition is problematic enough in itself, I feel (how can we separate agentive actions (design) from an agent?), but the more central problem with this approach, of course, is that (as Bedau observes) it assumes the notion of *good* design, or of something's being *more* or *less* fit for purpose; in this way, the approach admits observer-relative values into the discussion. After all, design is not, by its nature, automatically 'good' or life serving; we could equally justifiably picture an inept designer, or one whose purposes were eccentric or bizarre.⁸ 'Good design' is therefore not something which we would find 'in the world', independently of human opinion as to what good design is. Thus, ultimately, the mentalist approach would seem to rely upon observer-relative values (or else, the existence of an actual deity).

Secondly, the systems approach attempts to explain apparent teleology in terms of goals and goal-directed systems.⁹ For instance, the function of the heart may be accounted for purely in terms of the constraints of a physical system; in other words, it circulates blood because that is how the system maintains its functional equilibrium. In this sense, the systems approach attempts to account for teleology in terms only of

⁵ Ibid., p.781.

⁶ In his rejection of the notion of intrinsic function, Searle only considers and rejects the etiological approach of R.G. Millikan and Larry Wright. In the following, therefore, I will seek to outline all three approaches with a view to showing how they are all inescapably value-laden (and hence observer-relative), or else inadequate as definitions of function (e.g. because they are too broad).

⁷ Another alternative, of course, is to argue that non-agentive functions (from a human perspective) have purpose through the fact that they were created by God to fulfil a certain goal (i.e. religious teleology). So, whilst it may be argued that mentalism cannot account for non-agentive function, 'divine mentalism' can. However, this begs not only the question of whether God exists, but also how we may know His purposes. This latter problem leaves even the religious believer in a position where some other, non-religious reason for attributing function must be found, or else it must remain a mystery.

⁸ Bedau (1992), pp.783-4.

⁹ Ibid., pp.784-5.

'intrinsic causal dynamics'.¹⁰ The problem with this approach, however, is that such notions as 'equilibrium', 'the goal of a system', etc., seem to be no more than a restatement of the teleological function that such talk was meant to replace (and be equally value-laden). We can illustrate this by showing that any state which a system passes through may be said to be the goal of that system – a goal which is, furthermore, at odds with the traditionally assumed purpose of the system (an example, of course, would be the disease of an organism). Furthermore, the idea of a system itself seems problematic (at least in relation to function), for we can imagine 'systems' that have the necessary causal interrelation, but which appear merely by some accident of nature (such as – to quote Bedau's example – the formation of a bridge over a river by the random effects of natural forces – such as erosion, earthquake, etc.).¹¹ The systems approach, then, would seem not only to require some normative judgement as to what the 'proper' goal of a system is, but also to provide too broad a definition of a system so as to include events whose goal is incidental or a random act of chance (e.g. the fashioning of a useable bridge by natural forces).

Finally, the third approach to be considered is the etiological account, as proposed by Larry Wright.¹² In this version, something can be said to possess a certain function if (and only if) the thing's performance of certain actions leads to that function being fulfilled. This is a fairly straightforward definition, for it is basically saying that to ask for the function of a thing is to ask for an explanation of why it is there, does the thing it does, etc. To ask, 'What is the function of the heart?' is therefore to ask for an explanation of how the heart works, what the effects of its actions are, how it came into being and continues to exist, etc. As Wright notes, this is very different from asking, 'What is the heart good for?' for

if "Why do animals have livers?" is a request for a function, it cannot be rendered "What is the liver good for?" Livers are good for many things which are not their functions, just like anything else. Noses are good for supporting eyeglasses, fountain pens are good for cleaning your fingernails, and livers are good for dinner with onions. No, the *function* of the liver is that *particular* thing it is good for which explains why animals have them.¹³

Wright's etiological account therefore seeks not only to avoid the difficulty involved in distinguishing between accident and true function, but also attempts to sidestep a number of other pitfalls that previous accounts have fallen into: the tendency to confuse function and goal, the failure to distinguish between 'a function' and 'the function', and the inability to provide an analysis which can account equally for natural and conscious functions.¹⁴ The reason it is possible to do the latter, Wright argues, is that 'in virtually all of the usual contexts, *X* was designed to do *Z* simply entails that *X* is there because it results in *Z*.'¹⁵ Thus, both types of function ultimately

¹⁰ Ibid., p.785.

¹¹ Ibid., p.785.

¹² Wright (1973). For Searle's discussion of this, see Searle (1995), pp.16-23.

¹³ Wright (1973), pp.155-6, italics in the original.

¹⁴ Ibid., pp.140-3. Incidentally, where Wright talks about 'natural' and 'conscious' functions, Searle talks about 'non-agentive' and 'agentive' functions respectively.

¹⁵ Ibid., p.165.

admit of an etiological account (i.e. one that explains function in terms of how it came about).

However, one objection to this approach – provided by Bedau¹⁶ – argues that such etiological descriptions may account for a thing’s function where in fact that thing had no teleology at all. For instance, a stick which is floating down a stream comes to be pinned against a rock and is kept there by the backwash it creates. From Wright’s point of view,

all of the accident counterexamples can be avoided if we include as part of the analysis something about how *X* came to be there (wherever): namely, that it is there *because it does Z* – with an etiological “because”.¹⁷

Furthermore, the etiological ‘because’ is exemplified by such sentences as, ‘It exploded because it got too hot’,¹⁸ a usage which seems to fit the stick and backwash example. It therefore seems that Bedau’s example meets Wright’s conditions for identifying a teleological-type purpose in the system, and thus distinguishing between accident and function, and yet provides us with something that we would not wish to call teleological. Wright’s analysis, therefore, would seem to fall foul of one of the very pitfalls he was determined to avoid: the inability to distinguish between accident and genuine function. This is fatal for the whole account, for, as Bedau points out,

Given that the heart example and the stick example involve a similar sort of etiology, why is only the heart teleological?¹⁹

Millikan and Proper Functions

The baton of etiological explanation of function is, however, taken up by Ruth Millikan.²⁰ Like Wright, Millikan is concerned with accounting for function in terms of its causal history. However, Millikan’s focus is primarily dependent on a Darwinian account of the evolution of biological function. So, for Millikan, what she terms ‘proper functions’ (what Searle would think of as ‘non-agentive functions’, and Wright as ‘natural functions’) exist by virtue of fulfilling either one of two conditions:

1. An item, *A*, originated ‘as a “reproduction” (to give one example, as a copy, or a copy of a copy) of some prior item or items that, *due* in part to possession of the properties reproduced, have actually performed [function] *F* in the past, and *A* exists because (causally or historically because) of this or these performances.’²¹
2. Or, ‘*A* originated as the product of some prior device that, given its circumstances, had performance of *F* as a proper function and that, under

¹⁶ Bedau (1992), p.786.

¹⁷ Wright (1973), p.156.

¹⁸ Ibid.

¹⁹ Bedau (1992), p.787.

²⁰ See Millikan (1984) and (1989).

²¹ Millikan (1989), p.288.

those circumstances, normally causes *F* to be performed by *means* of producing an item like *A*'.²²

An item fulfilling the first of these conditions is termed by Millikan a 'direct proper function' in that it inherits its purpose directly by virtue of being a copy of something that had the same causal properties, and performed the same apparent function.²³ So, a heart's function is to circulate the blood, because not only does it actually do that, but it also possesses the same properties, and is a direct copy of, hearts that did the same job in direct ancestors. However, 'derived proper functions' work in a different way; we may say, for instance, that when a chameleon changes its skin colour to blend in with its environment, it is utilizing a device whose proper function (blending in with the background) is, in a sense, 'indirect', in that such skin-pigment-changing functions as the chameleon possesses 'have not proliferated as a result of producing, specifically, brown and green skin patterns' and, unlike the function of the heart, is 'a proper function only as adapted to a given context.'²⁴ Hence, since the proper function in this case exists in *relation* to something else, it cannot be direct; the function is therefore provided by the context (it is accidental, and not purposive), whereas direct functions exist (in an evolutionary sense) *because* of the performance of this function in the past (and the part that the performance of this function has played in the evolution and survival of the organism). However, for Millikan, both derived and direct proper functions may be contrasted with accidental function (such as the function of the nose in holding up eyeglasses).

Fascinating as it is, however, Millikan's account of function is open to a number of criticisms. Most importantly, as is pointed out by Searle,²⁵ it fails to account for the normative aspect of our idea of function. So, when we say, for instance, that a heart is malfunctioning, it is implied that there is a state of functioning that is worse or better than another. However, in Millikan's account, she can only attempt to account for the *existence* of hearts, or the *existence* of a chameleon's capacity to alter the pigmentation in its skin. However, in doing this, the notion of something being good at something, of optimum performance, of disease, etc., is left unaccounted for. This, it may be argued, is a necessary consequence of any attempt to account for teleology in terms of 'brute facts' (i.e. intrinsic features of the world), and – by extension – any attempt to apply this type of reductive explanation to what would seem to be, at least in part, value-laden and observer-relative terms.²⁶

Another problem for Millikan, however, is how she can adequately distinguish between derived and direct proper functions, or indeed between proper and accidental functions. For instance, imagine the following situation: in my son's bedroom there is a toy figure of a soldier which has a motion detector (what is sometimes called a 'room guard'). Therefore, whenever something comes near it, it activates a recorded sound, and says 'Who goes there?' So, one day, my son leaves for school, leaving the toy switched on. As a result, the house cat – which usually sleeps on my son's bed – is afraid to go in the room. Now, I come home from work early – before my son – and

²² Ibid.

²³ See Millikan (1984), chapter 1, pp.19-38.

²⁴ Ibid., p.40. For more on derived proper functions, see chapter 2, pp.39-49.

²⁵ Searle (1995), p.18.

²⁶ I will come back to this point later when I will look at the consequences of this position (which, incidentally, is one favoured by Mark Bedau) for death.

discover the situation. How do I know whether my son has done this deliberately? I can surmise, of course, but the only way I can be sure is to ask him. In effect, this is the only way I can distinguish between the accidental and conscious action of leaving the toy turned on.²⁷

Now, in relation to Millikan, we can see that she faces a similar question. How can she distinguish between the accidental and deliberate nature of a function? In the above example, to establish that the toy was the direct or indirect cause of keeping the cat out of the room, we would need to know my son's intentions (perhaps he only intended to keep his sister out of the room). So, to return to the example of the chameleon, to distinguish between this type of indirect function and the direct function of the heart, Millikan needs to show that we can clearly distinguish between direct and indirect causation in this way. However, as an advocate of evolution, she has no designer whose intentions she can query. Therefore, she needs to show that the causal properties of a system are enough to make this distinction – but are they?

There would seem to be room for argument concerning what is and is not a direct function of something. So, in relation to the function of the heart, we can say that its function is to pump blood (a direct function). However, what if we said that its function is to provide oxygen to the brain? Is this a direct or indirect function? Undoubtedly, human beings exist because of this function of hearts, which may in turn be causally related to the existence of human beings (and hearts). Does the heart oxygenate the blood or merely pump it? Or, in fact, merely *pump*? Where do we draw a line between what is its proper function and what isn't? The problem would seem to be one of causal scope. For example, human beings exist because of their superior intelligence (which allowed our ancestors to craft tools, manipulate our environment, etc.) – is this in turn therefore a direct function of the heart (through the oxygenation of the blood which passes to the brain)? To follow this through to its absurd conclusion: is the existence of power tools a direct function of the heart? To draw a limit to the consequences of function, however, would seem to require a decision – an arbitrary declaration of limit. Otherwise, the consequences of function can be stretched and broadened in ways which make the whole notion unworkable.

Another objection in relation to a thing's function would be regarding the question of what 'it' is, or how we circumscribe its identity. For instance, in relation to the function of the brain, do we consider it as a whole, or – given its many functions – as diverse parts? Furthermore, do we draw a distinction between the brain and the nervous system, or do we consider them to form a whole? As can be seen, where parts are connected, where do we draw the line between one thing and another? Isn't this fairly arbitrary – or, at most, for the purposes of our ease of understanding? So, as with drawing distinctions between physical components of the organism (heart, brain, nervous system, etc.), so too with function: both, I argue, are human conventions which serve certain observer-relative purposes.

Ultimately, then, as Searle argues, it would seem that we must accept that

²⁷ However, even this may be problematic. Firstly, there is the question of unconscious motivation in a psychoanalytic sense (as to what someone *really* intends); secondly, there is the question of 'moral luck' (see Nagel, 1979, and Williams, 1981, where we regularly judge agents on the consequences of actions outside of their control – which suggests that the conscious intention behind an action is not necessarily the defining feature – e.g. as with negligence or neglect).

Functions are never intrinsic; they are assigned relative to the interests of users and observers.²⁸

In conclusion, we are therefore faced with the following options:

- (1) Non-agentive function is an intrinsic feature of the world, either
 - (a) because it is a naturally occurring feature, or put there by a divine designer; or,
 - (b) because its apparent purpose is an *as-if* intrinsic feature of the world;
- (2) Non-agentive function is an observer-relative feature of the world (and humans ascribe purpose and goals to processes according to their own interests).

Option (1) (a) would seem to suppose some inherent purpose in nature, and so is generally open to the sort of objections that the teleological argument for the existence of God is vulnerable to:²⁹ e.g. that we cannot know with any certainty what those purposes are (they are open to interpretation). Therefore, since we have difficulty in identifying when function is intended and when it is not – and are unable to identify necessary and sufficient criteria for intentional or intrinsic function – then function is, to all intents and purposes, imposed by human observers.

Option (1) (b) is, as I hope I have shown, incoherent. It would seem to assume that by terming such features *as-if* functions we can account for what appears to be either natural order or divine design without assuming either. However, as I have argued, it does not do this, but merely begs the question as to why such functions exist in the first place, and why we may not interpret them another way. If we can do the latter, then surely function is supplied by human purposes; if we can't, then on what grounds can we be sure that the identified purpose is the *true* purpose?

This therefore leaves us with option (2).

The significance of this conclusion for the debate concerning the definition of death is therefore that, if function cannot be considered to be an intrinsic property of biological organisms, then for someone holding a functional definition of death (whereby death is construed as the cessation of certain vital biological functions), death itself cannot be considered to be an intrinsic feature of the world. Therefore, without human beings to ascribe observer-relative function to things, there would be no such thing as death. Furthermore, it is hard to see any alternatives which could propose a strictly-biological definition of death which did not in some way rely on some notion of function. The alternative to this would be to utilize a definition of death which did not seek to be exclusively biological (i.e. a partly-biological approach,

²⁸ Searle (1995), p.19.

²⁹ See, for instance, Hume (1966), *Dialogues Concerning Natural Religion*, Part VIII, and Hume (1975), *An Enquiry Concerning Human Understanding*, section XI; and for a more general critique, Hick (1990), pp.23-6. Obviously, the main argument against (1)(a) is the idea that there are more plausible explanations – i.e. the theory of evolution – e.g. Dawkins (1986).

such as one reliant upon some notion of personal identity). However, if this latter course is adopted then, since it will necessarily involve social and cultural notions (at least in part), I can see no way in which death can be considered anything other than an observer-relative feature of reality. Therefore, whatever course is adopted, the possibility that death is an intrinsic feature of the world seems to be excluded.

Function and Technology

Whilst I agree with Bedau and Searle on this point, there may however be a further point which makes the whole issue of a functional, biological definition of death unworkable anyway. If anything may be considered a table which will serve as a table, then similarly, anything can be considered a heart (or lung, or brain-stem) which will serve as such. Therefore, since organ transplantation and artificial aids to respiration and heartbeat can replace or supplement the functioning of the original organs, then – theoretically at least – we can consider death as relative to our technological know-how. In other words, organs can be replaced and supplemented by technology, and you are only dead when there is nothing more (technologically speaking) that can be done for you.³⁰

It may be argued against this last point that not all functions of a human organism can be replaced or supplemented – for instance, the function of the brain-stem. However, as I have already argued in Chapter 2, since we are only dealing with the notion of our current state of medical technology I do not see how this possibility can be conclusively ruled out. It may be argued, as David Lamb has done,³¹ that ‘if that were possible then anything could happen’, suggesting that such science-fiction-type notions of theoretical possibility should not play a part in the definition of death. However, even if this is accepted as a *reasonable* point, it does not affect the *logical* point that the objection here is based upon *empirical improbability*, rather than *logical impossibility*.

It may be objected here that in highlighting the significance of technological issues for this debate I am not in fact invalidating the idea that death consists in the cessation of certain functions, but merely expanding the functional definition to include both biological *and* technological components. For, whether the heart that stops beating (or the brain-stem that stops functioning) is biological or mechanical, death still consists in the cessation of its function.³² I would reply to this objection that whilst I would agree that the notion of function may still play a part in certain conceptions of death, I would also argue that the significant difference between a biological functional definition and a biological/technological functional definition is that in the former case, death is seen in terms of the *irreversible* cessation of function, whereas in the latter, the very idea of irreversibility is at odds with the notions of repair and maintenance that accompany the utilization of technology. In other words, if a heart

³⁰ There is a growing amount of interest in the question of exactly what causes natural death, and whether we can in fact do something about it. In some cases, this is serious science (such as in the search for the genetic triggers that cause aging, or limit cell division); however, there is also a growing ‘crank’ fringe – see Sanders (2002).

³¹ Lamb (1985), p.62.

³² I include heart transplants as a form of technology here since they would not be possible without certain technological developments, but also because they are an example of the notion of replaceability which is central to the idea of technology.

irreversibly ceases to function, we may possibly replace it with another one, the heart of an animal, or even a mechanical substitute. Therefore, whilst the notion of function is still relevant to such a definition, the accompanying notion of irreversibility is replaced by questions of current medical knowledge, technological advancement, etc., in relation to which the definition of death is constantly shifting. So, even if death *could* still be thought of as a functional matter, the conditions which cause the complete cessation of function are always on the move.

It would seem, then, that we do not have to equate cessation of biological function with death, because the cessation of those functions may in theory be reversible. Death occurs, therefore, when, according to our current medical technology, nothing more can be done. Far from being an intrinsic feature of the world, therefore, death – according to the functional definition – becomes contingent.

Conclusion

As can be seen from the above, not only is the functional definition of death a necessarily observer-relative view, but it also allows – via consideration of the role of technology in relation to function – for a definition which is contingent upon medical knowledge and technological advances. In fact, the very idea of cryogenic suspension is testament to this idea, for what is this if not an embodiment of the very idea that ‘science will one day find a cure for what we currently think of as death’?

However, these are not the only problems with the functional definition of death. One further problem (considered earlier in chapter 2) is that we may envisage a situation where an organism functions but it would be controversial to consider such a being as alive.³³ For instance, if we could replace the functioning of the brain-stem where a human being had no head, then the organism would still function, and – according to the functional definition – be technically alive. If a headless body can therefore be made to function, then to say that the organism is dead because it lacks the function for conscious awareness is to introduce consciousness as an intrinsic feature of being a living organism. In one sense, of course, consciousness is *not* an essential feature of being a living organism; there are organisms aplenty which lack it, and which we would not consider for that reason ‘dead’ (mainly – as with plants, insects, etc. – because we never considered them to have it – though this does depend, of course, on how consciousness is defined).³⁴ However, for human beings, the case is somewhat different, for we generally presume at least the capacity of consciousness as central to our notion of what a human being is.³⁵ One interesting question which arises from such concerns, therefore, is whether the idea that death is an intrinsic feature of the world can be rescued (at least with regard to human beings) by considering consciousness an intrinsic feature of the human being, the permanent incapacity for which constitutes death. I will deal with this possibility shortly, but in doing so I think we must leave behind the possibility that the functional definition (and therefore the strictly-biological approach) can treat death as an intrinsic feature. Failure in function of one aspect of one organ in an organism need not entail death so long as the overall

³³ Gervais (1986, p.57), raises this objection.

³⁴ Which I deal with in the next chapter.

³⁵ I realise, of course, that the notion of ‘human being’ employed here is a vague one, but I am content to leave it unexplored for the moment because it exemplifies the sort of ‘institutional fact’ (to use Searle’s term) the status of which I shall be exploring in more detail in subsequent chapters.

organism can continue to function as a whole. In terms of the permanent incapacity for consciousness, therefore, we are faced with the removal of a function which – whilst we might consider it significant – need not entail the death of the organism.

We may perhaps alleviate this problem by distinguishing between ‘human organism’ and ‘person’.³⁶ So, whilst the notion of a person that can actively fulfil certain social roles is observer-relative and therefore not intrinsic, the role played by the capacity for conscious thought in enabling such activities means that the failure of this capacity is sufficient for that of being a person also. Therefore, whilst the notion of the person is observer-relative, incapacity for consciousness (it may be argued) is a purely physiological concern, and thus may be considered an intrinsic feature. Thus, in a roundabout way, human death becomes an intrinsic feature of the world.³⁷ In this way is avoided the much thornier problem of what consciousness *is*, for whilst this question is hotly disputed and undecided, there is pretty much unanimous agreement that the *capacity* for consciousness resides in certain causal interactions within the brain.

I will turn to this question in the next chapter.

³⁶ As proposed by Green and Wikler (1980).

³⁷ This is, essentially, the position proposed by Agich and Jones (1986).

5. Death and Consciousness

Introduction

In the previous chapter, I concluded that the functional definition of death cannot lead to considering death itself as an intrinsic feature of the world. However, given the prominent place held by the capacity for conscious awareness in our common concept of a human being, it is tempting to consider whether the permanent cessation for the capacity for consciousness might not, in some way, provide a basis for considering *human* death as an intrinsic feature.

Answering this question requires a number of preliminary issues to be addressed:

- (1) In terms of being an intrinsic feature of reality, is 'capacity for consciousness' open to the same objections as the notion of biological function?
- (2) If consciousness is a contingent feature of human beings (and human beings are a contingent feature of the world), then what consequences does this have for considering a definition of death based upon the capacity for consciousness?
- (3) What consequences do the competing notions of consciousness have for any purported definition of death?

I will now turn to these questions.

Capacity for Consciousness

In terms of considering the capacity for consciousness as a criterion for being alive or dead, the first question we must deal with is whether the notion of capacity for consciousness fares any better than that of function in terms of its being considered an intrinsic feature.

On a purely verbal level, there would seem to be little difference between 'capacity' and 'function' (for our purposes). It would seem to be a fairly straightforward matter to account for capacity in terms of function. So, instead of saying, 'this object has the capacity to act as a table', we can similarly say, 'the function of this object is to act as a table'. So, we can say that 'the function of the heart is to pump blood', or 'the heart has the capacity to pump blood'. Therefore, regarding consciousness, we can say, 'the brain has the capacity for consciousness', or 'a function of (a certain part of) the brain is to produce consciousness'. The only difference in the use of 'capacity' and 'function' would seem to be that 'function' has more of suggestion of *intrinsic* capacity, but this is only a vagary of usage (we might equally talk of something 'being able to function as' something else – i.e. agentive function). The point therefore is that, whether we are talking of a particular function of the brain or the capacity to produce that thing (consciousness), we are in the same boat. We can therefore assess capacity

for consciousness in the same way as we assessed the possibility of intrinsic function in the last chapter.

However, having said this, with capacity for consciousness/conscious function we have a slightly different problem. In considering the heart, we may distinguish between a description of the causal interrelation of its parts, and ascribing function and purpose to it. The former concerns ontologically-objective features, whilst the latter involves the ascription of human values. However, in considering the brain, when we attempt the same sort of distinction (between causal description and value ascription), the outcome seems less convincing. This is mainly, I think, because of our own subjective experience of being conscious: when we attribute function to the heart, we are still essentially describing objective, third-person, physical properties (though, of course, subjectively we also feel our own heartbeat); however, when we are attributing function to the brain, we have in mind – in part, at least – a *qualitative* outcome (that of phenomenal consciousness).¹ Any reluctance to accept the distinction in this context therefore stems from a feeling that one's own consciousness is not dependent upon agentive ascription. In other words, that my heart has a teleological function is something which is dependent upon human value, but that I am conscious – in as much as it concerns my subjective, phenomenal experience – is not.

For now, we may initially conclude that – in as much as the brain is a functioning organ – the capacity for consciousness may be thought of as a non-agentive ascription (and, therefore, observer relative). However, in as much as consciousness has another, subjective aspect (that of phenomenal consciousness), then the comparison with the heart (and other things to which we ascribe non-agentive function) is asymmetrical: the physical aspects are parallel, but (whilst we may be aware of our heartbeat) the primary function of the heart is adequately described by an account of its physical causal properties; however, with the brain, this is not the case.

I will return to this last point shortly, when I consider competing interpretations of the nature of consciousness and their significance for the status of human consciousness itself. However, before doing so, there is a further issue which first needs addressing. I argued earlier – in line with Searle and Bedau – that function is an observer-relative feature of the world. However, in considering here whether capacity for consciousness should also be considered an observer-relative feature, we are left with a question: can contingent features also be intrinsic?

Contingent and Observer-Relative Features

In utilising Searle's distinction between intrinsic and observer-relative features of the world, we come to the question of what status human consciousness has.² If we hold consciousness to be an intrinsic feature, and that the permanent cessation of the capacity for it may be viewed as a sufficient condition for death, then accordingly

¹ I am here following the distinction made by David Chalmers between those aspects of our mental functioning which may be described in purely causal terms (a 'psychological' description), and those which represent subjective experience ('phenomenal'), and hence (arguably) cannot be wholly captured by merely causal description. The latter, of course, are also covered by the term 'qualia'. See Chalmers (1996), pp.11-16.

² Hereafter, simply 'consciousness'.

human death might also be considered to possess intrinsic status. In answering this question, however, a number of points need first to be clarified.

To begin with we need to distinguish between those features which are observer relative, and those whose existence is merely contingent. The existence of human beings, for instance, is contingent upon such factors as biological evolution, historical accident, and the very existence of life itself on this particular planet. Yet this does not mean, seeing as human beings *do* in fact exist, that they do not possess intrinsic features. We may distinguish, therefore, between those things which exist contingently, because they are observer relative, and those which, although their existence is contingent, nonetheless possess intrinsic features.

Observer-relative features exist because of the human ability to conceptualise. For instance, without language and the ability to create and utilise ‘institutional facts’ (as Searle calls them),³ there would not exist such things as ‘money’, ‘ownership’, ‘property’, etc. These notions all require social, linguistic interaction, and could not form part of the ‘brute facts’ of a universe in which humans – or language-capable beings – did not exist.⁴ In this sense, observer-relative features are contingent merely in the sense that their existence is dependent on the existence of this type of language user. More importantly, however, these features are, in Searle’s phrase, ‘ontologically subjective’, in that they exist *only for* those language users. So, whilst ‘money’ only exists because language users exist (a contingent fact), there is a further sense in which only such language users can really *use* ‘money’ (as part of the linguistic, social institution of property and ownership). In other words, a dog may fetch your newspaper, but it cannot buy one.⁵

The other sense in which the existence of certain features is contingent is, as has already been noted, when we consider that certain things simply may never have come into being. For example, the human brain may not have evolved at all, and there might have been a situation in which the planet that we now call home did not in fact host the type of language user that such neurological capacity makes possible. However, the point to be realised here is that contingency in this simpler sense does not in itself entail observer relativity; the fact that human brains may never have existed does not mean that their contingent existence debars them from the possession of intrinsic features. A tree at the end of my street possesses mass, has a certain organic structure, chemical consistency, etc; all these features are not the less to be considered intrinsic simply because that tree may never have existed. The distinction

³ See, for example, Searle (1995), pp.27-9. The main point here is that we can talk about certain things factually only because there are human beings who can create the conceptual frameworks (‘institutions’) wherein this is possible.

⁴ Humans may not be necessary for the existence of observer-relative features, but some sort of conscious, concept-utilizing, socially-interacting, linguistically-sophisticated creature would be. So, the fact that animals have a certain sort of behavioural language would not in itself guarantee the existence of observer-relative features. In other words, chimps may communicate, socially interact, problem-solve, and so on, but it is unlikely that they have the concept of – for instance – ‘money’ (or some primitive precursor of it). I don’t think that the possibility can be ruled out *a priori*, however, but merely in terms of what it appears chimps are capable of (it may turn out, for instance, that chimp language is more sophisticated than has been supposed).

⁵ There are, of course, incidents where animals seem to partake in these institutional practices: cats and dogs become sole beneficiaries of eccentric will bequests, someone marries a horse or a goat, and so on. However, in all such cases, the animal is a passive participant, whose role is ascribed and maintained by human agency.

between observer-relative and intrinsic features is therefore not to be confused with contingent and necessary existence; contingency is not coextensive with observer relativity, and a thing may possess intrinsic features whose existence is not necessary.

Of course, this latter category includes almost every individual thing; the computer on which I am typing exists contingently (it might not have existed at all – either as a particular instance – this particular computer – or as a general type of thing), and yet it possesses the intrinsic features of mass, dimension, etc.⁶ In a similar way, the existence of each individual physical object in the universe may be said to be contingent, whilst the existence of the class of physical objects itself may not be.⁷ However, the main point here is that, simply because a thing's existence is contingent, this does not in itself make it observer relative. For Searle then, an intrinsic feature is something whose existence is independent of our conceptualisation of it. Therefore, independent of the existence of human values and language, intrinsic features still exist.

The Status of Consciousness

So, what status does consciousness have? Is it an intrinsic feature of a human being (such as their weight or height)? As I have already argued, whether a feature exists contingently or even intermittently does not affect its intrinsic status. Therefore, if consciousness can be considered to be an intrinsic (although intermittent and contingently-existing) feature of human beings, it would seem to suggest that it represents a unique class of feature.

To begin with, we may imagine a world without consciousness, whereas to imagine a world without physicality would be – arguably – not to imagine a world at all.⁸ If we therefore take it to be intrinsic, consciousness would seem to be a feature which it is possible that no particular thing embodies at any one time, in as much as we can imagine a world without any conscious beings in it.

This is different from the case with other intrinsic features – such as dimension – in that, as I have already suggested, there cannot exist a world in which dimension did not exist (for it would not be a world).⁹ If we accept this view, therefore,

⁶ Another way of thinking about this, of course, is in terms of the age-old distinction between universals and particulars. The 'particular' object possesses the 'universal' features of mass and dimension.

⁷ Is it a contingent matter that the physical universe exists? Until we know more about how the universe began – and we may never know enough – then this would seem to be an unanswerable question.

⁸ Unless, of course, we were to hold to some form of idealism, but this in itself does not entail antirealism (see the discussion of this point in 'Conflict and Resolution in the Development of Categories', in chapter 3 above).

⁹ I feel that I am at the limit of my metaphysics here, and I am prepared to admit that these points may be debateable or defeasible in a more radical direction. For instance, someone more versed in modern physics might like to argue for the possible non-existence of one intrinsic feature or other (such as mass or dimension) at different stages of the development (or demise) of the physical universe, but my ignorance in this field forbids me to do so. For example: 'up until the universe was about 1 millionth of 1 second old, the constituents of the primordial soup still did not yet directly resemble any object we commonly see around us, or have yet isolated in our laboratories. At that time, free quarks were everywhere, and *matter had no mass.*' Krauss (2001), p.52 (my emphasis). See also the observation made by Ladyman (2002), p.144, discussed in chapter 3 above, regarding quantum physics and the fast-disappearing list of primary qualities. However, my main purpose here is simply to explore the

consciousness would seem to be an anomaly among intrinsic features, for whilst it is possible for any one thing not to exist, whatever does exist commonly possesses all the intrinsic properties that there are (mass, dimension, etc.). However, as I have shown, it is possible to conceive of a world in which consciousness did not exist (or, even, for us to believe that this was once the case – i.e. before it developed). Therefore, consciousness – to use traditional technical terms – would seem to represent a universal predicate which did not require instantiation in any particular in order to be considered intrinsic; or, in plainer terms, consciousness may be an example of an independent feature of reality which there may, at certain times, be no instance of.

Let us entertain a possible objection to this argument. If we consider that sometime shortly after the Big Bang – as scientific wisdom has it – the universe consisted only of gases, then there would have been no ‘solid’ objects. In terms of the traditional tripartite division of solid, liquid, and gaseous states of things, only the latter would have existed. In such a world, ‘solidity’ would not have been a feature possessed by anything. In this sense, so we might argue, ‘solidity’ is an intrinsic feature which – like consciousness – need not exist in any particular.

The problem with this objection, however, is that the case of solidity is not an exact parallel. Certain physical features of the world – such as solidity, liquidity, etc. – can be reductively explained. The structure and interrelation of the atoms which make up the gas can also be called upon to explain how they might be structured in order to form a solid or a liquid. Therefore, instead of talking about the contingent existence of the feature of ‘solidity’, we may simply talk of the contingent interrelation of atomic particles in a certain pattern. In other words, we would say that ‘solidity’ is a *high-level* feature which *supervenes* on the *low-level* brute facts of the universe. Such supervenience is not, in itself, anything mystical, and furthermore allows of a physically reductive explanation (i.e. ‘high-level’ features are reducible to ‘low-level’ features – which are in turn brute, physical, and irreducible givens). However, the next objection is an obvious one: but can’t we similarly explain consciousness reductively, and so avoid attributing to it this special status? Answering this question, however, requires a clear definition of what is meant by ‘consciousness’.

Limits of the Enquiry

Before beginning my discussion of the nature of consciousness, and its relevance to the death debate, I must first make some general observations as to the approach I will take.

So far, I have dealt with the notion of consciousness in a very vague way. Mainly, this was to facilitate the preliminary discussion surrounding the potential problem involved in considering consciousness to be an intrinsic and yet contingent feature. However, having considered that point, it now becomes important to decide to what extent distinct and competing conceptions of consciousness have a bearing on its potentially intrinsic status. For instance, if it is possible – as some philosophers believe – to account for consciousness in purely quantitative and causal terms – then

status of consciousness in the light of Searle’s distinctions. Moreover, whether Searle’s distinction actually holds will be a question I will return to later.

there may be no barrier to considering consciousness as an intrinsic feature. However, if – as it is also sometimes argued – consciousness *cannot* solely be accounted for quantitatively, then perhaps this represents a problem for any arguments which hold that consciousness is an intrinsic feature.

Secondly, whilst I have been vague as to the nature of consciousness, I have also been somewhat anthropocentric. In other words, where I have used the term ‘consciousness’, I have generally had in mind ‘human consciousness’, and have excluded any notion of possible animal consciousness or degrees of conscious activity. I have also assumed, for the purposes of this debate, that animals are not capable of the sort of thought formation that would enable them to partake in the kind of utilization of institutional facts (e.g. money) that distinguishes human interaction. This issue will become relevant now, however, when we consider the distinction between ‘phenomenal’ and ‘psychological’ consciousness; if, for instance, it turns out that animals possess the same capacity for a particular type of consciousness, and that this particular type of consciousness is significant to the determination of death, then we may be able to establish a unified definition of death at least between humans and other creatures capable of this type of consciousness (thus avoiding a twofold, or even threefold definition of death). I shall therefore return to this issue once my conclusions have been established.

Thirdly, in the following sections, whilst I will set about clarifying my position on the two above issues, I should however point out that I will not be seeking to provide original contributions to the nature-of-consciousness debate. Rather, it will be my central concern to indicate, with regard to the various possibilities discussed, what consequences it would have for the status of death if capacity for consciousness is adopted as a criterion. Furthermore, I will be seeking to identify whether or not there are more central considerations which might allow us to clarify the relation of consciousness to the status of death in a way that does not necessarily entail the advocacy of one theory of consciousness over another.

The Nature of Consciousness

The debate as to the true nature of consciousness is a heated one. However, the camps divide roughly according to whether it is possible to explain the phenomenon of consciousness reductively. In turn, this issue gives rise to the question of whether there is a ‘real’ aspect to consciousness which reductive, third-person accounts necessarily exclude. In *The Conscious Mind*, David Chalmers sums up the position nicely. According to his division, consciousness has (arguably) two aspects¹⁰:

1. The psychological aspect. This involves the ability to calculate, judge, will, seek and achieve goals, etc. Such a view of mind admits of a reductive explanation, in that these types of action can be accounted for (again, arguably) in physiological terms (by presenting such things as ‘purpose’, ‘belief’, etc., as potentially measurable neurological activity, biological structures within the brain, potential behaviour, etc.). For example, Daniel Dennett argues that intentionality can be explained in purely causal terms, and that what we do in ascribing intention is to adopt an intentional interpretation

¹⁰ Chalmers (1996), ‘Chapter 1: Two Concepts of Mind’, pp.3-31.

(what he calls the ‘intentional stance’) of such causal interaction.¹¹ Another advocate of a reductive explanation of consciousness, Francis Crick, puts the general materialist position quite succinctly:

“You,” your joys and your sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact no more than the behaviour of a vast assembly of nerve cells and their associated molecules.¹²

Hence, the main problems here are logistical ones, such as, ‘How does the physical brain function as a biological information-processing system?’ Obviously, the philosophers who think that this question can be conclusively answered also think that consciousness can be modelled or functionally understood, and that therefore ‘strong Artificial Intelligence’ is possible.¹³ Chalmers calls theories which propose that consciousness can be explained *purely* in these terms (and that a complete reduction is therefore possible) Type-A theories.

2. The phenomenal aspect. This refers to the experience of being aware, having so-called ‘qualia’ (the subjective experiences of having ideas, sensations, etc.), and generally, the idea that there is *something it is like* to be conscious.¹⁴ Some (i.e. Type-A) theories reject the coherence of this idea because it can so easily lend itself to what they consider to be unacceptable forms of mysticism or dualism – i.e. to belief in the existence of ‘something other’, or something ‘over and above’ the physical. However, this need not be the case, and advocates of these (Type-B) theories hold that phenomenal experience is real and irreducible, whilst at the same time holding to a materialist view of the universe.¹⁵ A third approach (Type-C) would be to argue that materialism itself is false, and adopt some sort of dualism or even idealism, which holds that, not only is consciousness irreducible, but that it is a fundamental and distinct property to that of matter.¹⁶

Setting aside, for the moment, the arguments concerning the nature and existence of the phenomenal aspect (and whether it can be ultimately explained in terms of a

¹¹ See Dennett (1971) and (1991); for a general explanation of Dennett’s position, see for example, Heil (1998), pp.154-66.

¹² Crick (1994), p.3

¹³ To use Searle’s distinction (see Searle 1992, p.44 ff), advocates of ‘strong AI’ believe that a machine which had the same functional organization as a human brain would be conscious in the same sense that a human is. On the other hand, ‘weak AI’ describes the position that, whilst machines may be considered to ‘think’ (in Chalmers’s sense of performing the functions associated with the psychological aspect of consciousness), they are not actually conscious (i.e. their consciousness lacks the phenomenal, qualitative aspect).

¹⁴ The idea that there is a qualitative experience which is distinct from a description of psychological or causal interactions was famously discussed by Nagel (1974), where he points out the many problems we face in accounting, not only for the qualitative experiences of other species, but also of other human beings.

¹⁵ E.g. Searle (1992, 1997), Papineau (1993), and Tye (1986).

¹⁶ Chalmers himself advocates some blend of dualism and panpsychism. Examples of other Type-C theories would include Jackson (1982), Honderich (1981), and Sprigge (1994).

psychological/physical reduction)¹⁷, I would like to relate these two aspects to the distinctions which I have already introduced. Firstly, in parallel to the point I made in relation to the notion of intrinsic function in chapter 4, the psychological aspect would seem to represent an observer-relative feature, for whether or not some form of behaviour is thought to constitute judging, willing, goal-seeking, etc., would seem to be a matter of interpretation based upon other observer-relative concepts that we hold (such as the notions of agency, identity, etc.). As Dennett rightly holds,¹⁸ we *may* adopt an ‘intentional stance’ or interpretation of such actions, and in the ‘ascriptions to the system of beliefs and desires (and hopes, fears, intentions, hunches,...)’, explain, predict, and generally account for its behaviour. So, whether we consider consciousness in terms of brain function (e.g. functionalism),¹⁹ as capable of being cashed out behaviourally (e.g. behaviourism),²⁰ or from some other materialist perspective (e.g. eliminativism),²¹ the ascription of function or purpose (just as it was in our discussion of the function of the heart) would seem to be observer relative.

However, the case of phenomenal consciousness would seem to be different. Is it an observer-relative fact that I have conscious, subjective experiences? In relation to my own experience of myself, this seems ridiculous, for I do not rely on interpretation or analysis in order to arrive at the conclusion that I myself am conscious (i.e. it is not a process of deduction) – in fact, the *very possibility* of my being able to interpret and analyse would seem to *presuppose* that. We might demur at this point, however – in line with a point made by Nietzsche – that, in supposing that there is some baseline subjective experience called ‘consciousness’, we are behaving ‘as though knowledge here got hold of its object pure and naked’.²² Nietzsche’s point, of course, is that what appears to be ‘self-evident’ is often no more than a construction or assertion on the part of the perceiver. Wilfrid Sellars makes a similar point in arguing that knowledge is a ‘conceptual state’,²³ and that mental experiences cannot therefore represent some sort of non-inferential form of knowledge as to the nature of the mind (targets here would include such concepts as ‘immediate certainty’, and Descartes’s *Cogito* – which, incidentally, Nietzsche also criticises for similar reasons).²⁴ Sellars termed this idea that we can somehow infer epistemological or ontological truths about things from the very fact of certain experiences, the ‘myth of the given’. In this case, this would mean that our judgment that phenomenal consciousness is a non-reductive reality is no more than a dogmatic assertion that serves certain metaphysical purposes.

However, as David Chalmers points out, subjective experience itself seems to be ‘more primitive’ than knowledge, because it is a ‘nonconceptual state’.²⁵ For, even if

¹⁷ We may distinguish here between whether, for instance, such a reduction is possible (and physicalism is true), and whether we currently (or will ever) possess the necessary conceptual apparatus to facilitate such a reduction. This is pointed out by Nagel (1970), who argues that we simply do not currently possess the requisite concepts to explain consciousness (and that current physicalist attempts are therefore misguided – or at least over ambitious).

¹⁸ Dennett (1971), p.87.

¹⁹ For instance, as David Armstrong does – see, e.g., Armstrong (1973).

²⁰ See, e.g., Ryle (1949).

²¹ E.g., Churchland (1981).

²² Nietzsche (1990a), Part One, s.16, p.46.

²³ Sellars (1956).

²⁴ For Nietzsche’s criticisms of Descartes on this point, see for instance Nietzsche (1990a), sections 16 and 17, pp.45-7; for Descartes’s *Cogito* argument, see Descartes (1986), Second Meditation, p.102 ff.

²⁵ Chalmers (1996), p.380, n.6.

we admit that the *knowledge* of our being conscious is constructed step by step (and therefore reliant upon language and conceptualization), the very notion of construction assumes the existence of something *out of which we construct* (something 'more primitive'). Therefore, even where it is admitted that sophistications to do with 'I' and 'self', and concomitant arguments to do with existence and certainty, are later, observer-relative constructions, there would seem to need to be – in order to take phenomenal consciousness seriously – some sort of pre-existing base-level of 'non-conceptual' subjectivity. Margaret Archer, in *Being Human: The Problem of Agency*, puts this same point a slightly different way. For Archer, 'our sense of self, as part of our humanity, is prior and primitive to our sociality.'²⁶

Since this sensing will be seen to be wordless, and necessarily so because it is both prelinguistic and alinguistic, then it cannot rest upon any concept appropriated from society. This does not make it unconceptual: the sense of self, like all other knowledge, is conceptually formed. What it does mean, is that the theoretical work involved is performed and recorded in ways which are non-linguistic.²⁷

Obviously, this raises further issues: for Archer, this primitive, pre-linguistic sense of self provides a link with the animal kingdom, many of whom may be said to share it. Also, some may take issue with the idea of non-linguistic 'theoretical work'.²⁸ However, these issues aside, this notion of the existence of a pre-linguistic, phenomenological baseline to experience and consciousness would seem to many to be a natural and coherent one.

Treating both the psychological and phenomenal aspects of consciousness seriously, therefore, we may consider consciousness to have a dual status. Firstly, from an objective point of view, it may be said to consist of certain intrinsic features (the interaction of neurons, synaptic structures, etc.). However, whilst the existence of physical parts is, in Searle's phrase, 'ontologically objective', our perception of function and purpose is 'ontologically subjective' (i.e. they are observer-relative features, and such things exist only for humans). Secondly, from a first-person point of view, conscious experience may be considered 'ontologically subjective', in that it only exists for the subject or conscious perceiver; for instance, my qualitative experience of feeling angry, or enjoying a piece of music, is something that only exists for me. However, *that* humans have such experiences (that they are angry, enjoy music, etc.) is an intrinsic feature of their neurology; in other words, my experience of a pain (to use a well-worn example) is subjective (it has a 'hidden' qualitative aspect); however, *that I have pain experiences*, is an objective fact. Searle puts the point like this:

Although the feature of being a screwdriver is observer relative, the feature of thinking that something is a screwdriver (treating it as a screwdriver, using it as a screwdriver, etc.) is intrinsic to the thinkers

²⁶ Archer (2000), p.121.

²⁷ Ibid., p.124.

²⁸ Though, incidentally, this position need not be problematic, for we need only consider the significance of split-brain experiments for the presence of non-linguistic conceptualisation and reasoning – see, e.g., Gazzaniga, et al., (1962).

(treaters, users, etc.). Being a screwdriver is observer relative, but the features of the observers that enable them to create such observer-relative features of the world are intrinsic features of the observers.²⁹

In other words, even though ‘mental phenomena are [...] ontologically subjective’,³⁰ having such phenomena is an intrinsic feature of human brains, and the phenomena’s existence are not dependent upon the existence of language users, social practices, etc. Phenomenal consciousness would therefore seem to be an intrinsic feature of human beings.

This is a subtle point, and is perhaps worth illustrating further: the experiences I have when I watch a film, feel pain, eat an apple, and so on, are qualitatively private; other people can do these things, but they cannot do them *as me*. So, in this sense, such experiences are subjective (and in Searle’s sense, ontologically so, because those particular experiences that *I* have exist only for me). *However*, the fact that I have such experiences, that I feel qualitative states at all, reflects an ontologically-objective aspect of human beings; we are capable of such things (just as physical objects have a certain mass). In this sense, therefore, phenomenal consciousness is an intrinsic feature of human beings (unless, of course, you deny the existence of phenomenal consciousness, as certain (Type-A) reductivists do).³¹

Epistemic Objectivity

Having dealt above with the ontology of consciousness, there remains the epistemological question of what status our judgements concerning it have. Here we may utilize Searle’s other distinction involving *epistemically-objective* and *epistemically-subjective* judgements. Put crudely, epistemically-objective statements (such as, ‘The Mountain is 1,000 feet high’) involve facts, whilst epistemically-subjective statements (such as, ‘*Hamlet* is a wonderful play’) are attitudinal or value-based. However, it is not the case that we can only be epistemically objective about intrinsic (ontologically-objective) features, and epistemically subjective about observer-relative (ontologically-subjective) features. We may also be epistemically subjective about ontologically-objective features (‘The mountain is awe-inspiring’), and epistemically objective about ontologically-subjective features (‘*Hamlet* is Shakespeare’s longest play’).

Statements >>	Epistemically-Objective	Epistemically-Subjective
Features ∨∨		
Ontologically-Objective	The Mountain is 1,000 feet high	<i>Hamlet</i> is Shakespeare’s longest play
Ontologically-Subjective	The mountain is awe-inspiring’	<i>Hamlet</i> is a wonderful play

²⁹ Searle (1995), pp.10-11.

³⁰ *Ibid.*, p.12.

³¹ A good example of this attitude and its arguments can be found in Dennett (1991), chapter 12, ‘Qualia Disqualified’.

In relation to consciousness, therefore, we must deal with the phenomenal and psychological aspects separately. Firstly, it must be noted that the psychological aspect of consciousness, whilst it can arguably be reduced to talk of intrinsic features (structures possessed by the brain, causal interaction of parts, and so on), cannot itself be considered ontologically objective. As I noted in my discussion of function and capacity, accounts which ascribe goals and intentions to systems – in fact, involve the very idea of ‘system’ at all – are observer relative. Furthermore, it would seem to be impossible for any psychological account to survive *without* attributing such function and goal-orientation, or else we would be merely left with the description of intrinsic physical properties – which, as far as consciousness is concerned, would not be an account at all. In other words, the enlightening concepts for understanding of the brain’s role in consciousness are mostly observer-relative, and descriptions merely in terms of atoms and molecules would not be very informative without observer-relative concepts to relate them to.³²

In summary, I have already argued (in line with Searle), that the phenomenal account of consciousness (the qualitative experience of having mental phenomena) is observer relative. Therefore, in different ways, both the psychological and phenomenal aspects of consciousness are (in a sense) observer-relative, ontologically-subjective features of the brain: the former, because, as with a functional account, a psychological explanation must involve non-agentive ascription; and the latter because, in terms of the mental phenomena themselves (the qualia), they arise because of our position as subjective observers (we experience them qualitatively). However, there is one important exception to this categorization: whilst the qualitative experience of mental phenomena is ontologically subjective, the *fact that we have* qualitative experiences is not. In other words, that human beings have such experiences is a biological fact (i.e. brains allow such qualitative phenomenal experiences). Holding this view is, of course, to subscribe to the ‘consciousness has a real phenomenal aspect’ camp of the debate (which, at least for the purposes of this argument, I hold to be a possibility).

Epistemically speaking, therefore, we may hold objective judgements about consciousness in both its psychological and phenomenal aspects: the former, based on the observer-relative ascription of function to the interaction of intrinsic features of the brain (i.e. goal-oriented systems of neurons, etc.); the latter, based on both our own first-person, subjective experience, and factual assertions concerning the existence of phenomenal consciousness in other human beings, animals, etc.³³ Thus, whilst my own mental experience is ontologically subjective, I may still make epistemically-objective judgements concerning it, in as much as others may seek to verify its nature (such as whether I am really happy) in much the same way as they might seek to verify more overtly objective phenomena – such as the existence of objects in the world.³⁴ This, ultimately, will have an important consequence for any

³² Of course, this ultimately suggests that the study of biology is much more reliant upon observer-relative concepts than physics is. Furthermore, the extent to which the concepts involved in the hard sciences concern intrinsic/observer-relative features is the basis of the debate between realists and social-constructivists regarding science.

³³ Of course, this is not to say that such assertions are necessarily correct, but merely that the solipsist, the non-sceptic (using the ‘argument from analogy’), the advocate of animal rights (perhaps), etc., are in effect in dispute concerning a factual matter (the existence of phenomenal consciousness in others). Though, just who the solipsist is arguing against is a puzzling matter...

³⁴ It should be noted here that I am not using these observations as proofs for the non-reducibility of such qualitative experiences; for the sake of my wider argument, however, I take such a non-reducible

attempt to present an account of death based on the absence of the capacity for phenomenal consciousness – to which I will now turn.

Cessation of Consciousness as a Criterion for Determining Death

I have so far argued that death cannot be considered an intrinsic feature of the world whilst its definition rests on observer-relative features (e.g. the functional definition). Therefore, without another suitable candidate, the definition of death may join other normative debates which concern conflicting values as to whether one thing or another should happen (for instance, whether or not to be vegetarian, or to impose trade embargoes on a certain country). This is not to say, of course, that ethical and scientific judgements are on a par, but merely to point out that, in the final analysis, in determining death we would leave the realm of ‘brute facts’ and move – at least significantly – into the realm of observer-relative and (ultimately) culturally-dependent ‘institutional facts’ (and from the strictly-biological view to the partly-biological view).

In the last section, I argued that holding that (a) phenomenal consciousness exists, and is real and irreducible, leads to the conclusion that (b) the capacity to experience conscious mental phenomena is an intrinsic (yet contingent, and not always present) feature of the human brain. Therefore, since the functional definition of death must involve observer-relative, non-agentive ascription, could phenomenal consciousness, as an intrinsic feature, provide a basis for resolving the death debate?

In response to this proposal, we may identify two main questions:

- (1) In what way does adopting this criterion help resolve the practical difficulties faced in determining death?
- (2) Even if doing so *would* resolve these difficulties, why should we equate death with cessation of consciousness *at all*?

The first of these problems is an epistemic one. That is, given the debate surrounding the phenomenal aspect of consciousness, can it really be used in practical situations in order to determine the moment of death? If death may be said to occur when there is a permanent cessation in the capacity for phenomenal consciousness, those individuals may be said to be dead who, in terms of their behaviour, display no evidence of the type of behaviour associated with the presence of that capacity.

However, there are a number of difficulties here. Firstly, the situation is an instance of the so-called ‘problem of other minds’; in other words, how do we know that, behind the apparently intelligent, purposeful, conscious actions of other individuals, there exist the same qualitative, phenomenal, mental experience as accompanies our own actions? In brief, we cannot know that is the case, and may only apply an ‘argument by analogy’ with our own experience (i.e. others behave as I do, I have ‘something going on inside’, therefore so do they). Obviously, materialist attempts to dissolve this problem reductively (e.g. certain brands of behaviourism, eliminativism, and other

state to be fact. The reasons for this approach will become clear when I consider the possibility that a definition of death can be based upon the permanent cessation of the capacity for such experience (which I will come to shortly).

types of reductive materialism), do away with the very thing that would be explained. However, those who would treat phenomenal consciousness seriously (and non-reductively) – as I have argued we should, for the moment – are faced with what would seem to be this insurmountable epistemic difficulty.

We might, of course, reject this problem as a sceptical paradox, and simply adopt the common-sense approach that the argument from analogy is ultimately based on. However, even if we do this, we may still be faced with cases where even the argument from analogy does not work: for instance, how would we know whether (a) phenomenal consciousness is or is not present, at some level, in (for instance) a comatose patient, and (b) whether the capacity for phenomenal consciousness has permanently ceased? These might be said to be scientific and technological concerns; if we can determine which areas of the brain are responsible for phenomenal consciousness, then we might be able to answer both (a) and (b) via brain scans (for instance, fMRI scanning). Even if such technology is not currently sophisticated enough, and our knowledge of the brain is not advanced enough, to answer such questions conclusively, we might nonetheless argue that herein lies the ultimate answer to our problems.³⁵

Secondly, an implicit assumption in attempting to adopt the presence of phenomenal consciousness as a sign of the presence also of some sort of intelligent, rational, volitional entity, is that consciousness does not seem to be necessary (or at least, need not accompany) many of these would-be defining criteria. As psychologist Julian Jaynes argues:

[Consciousness] is not to be confused with reactivity. It is not involved in a host of perceptual phenomena. It is not involved in the performance of skills and often hinders their execution. It need not be involved in speaking, writing, listening, or reading. It does not copy down experience, as most people think. Consciousness is not at all involved in signal learning, and need not be involved in the learning of skills or solutions, which can go on without any consciousness whatever. It is not necessary

³⁵ However, this may not be the case. Furthermore, if we are to detect consciousness via brain-scan technology, then we would do so based on brain *function*. Such a method would rely on the idea that certain parts of the brain have certain roles, and the presence of electrical activity, blood flow, etc. (such as an fMRI scan would pick up) would indicate the corresponding mental activity. However, this relies on two assumptions: (1) the functional roles played by certain parts of the brain are the same in all brains, and (2) such activity is always accompanied by phenomenal consciousness. However, medical research currently indicates that – whilst there *is* some standardization – ‘brain-maps’ can differ between individuals (according to handedness, gender, and cases of abnormal development), as can the morphology of the brain itself, and that there would currently seem to be insufficient understanding of the precise relationship between function and brain areas to determine conclusively in all cases what are the nature of actual or potential mental states in the comatose. For instance, see Brett, et al. (2002), which details some of the current technological, methodological, and epistemological problems in this type of research). Furthermore, Blackmore (2003), points out that the problem of mapping states of consciousness is the difficulty in clearly defining which areas of the brain are essential to the state in question (pp.326-7); she also observes that – for instance – unconsciousness and consciousness are not accompanied by activation of different areas of the brain, but are rather accompanied by differing degrees of activity in *the whole brain* (pp.228-9). Therefore, the functional approach to brain studies, whilst undoubtedly useful, is far from an ‘easy way-in’ to understanding the nature of consciousness, and may ultimately be undermined by the non-standard nature of brain organization, and the difficulty in relating subjective states to corresponding physical events.

for making judgements or in simple thinking. It is not the seat of reason, and indeed some of the most difficult instances of creative reasoning go on without any attending consciousness.³⁶

Therefore, since these things may occur without conscious awareness (phenomenally defined), it would seem to be a futile effort to cite them as evidence of the presence of phenomenal consciousness itself.³⁷ For, as Jaynes goes on to say:

If our reasonings have been correct, it is perfectly possible that there could have existed a race of men who spoke, judged, reasoned, solved problems, indeed did most of the things that we do, but who were not conscious at all.³⁸

Therefore, even if we could confidently identify those areas of the brain that performed certain tasks, and furthermore, could ascertain when they were active (or potentially so), then there is no guarantee that we are not witnessing some sort of reflex brain activity which is ultimately empty of phenomenal content. So, even if the technological difficulties can be overcome, there does seem to remain an epistemological one.

This is not to say, of course, that phenomenal consciousness may not play some part at some stage in the development of these psychological processes,³⁹ but for all that, the objection seems an arresting one in the current context. Of course, reductivists concerning consciousness would wish to use such observations to point out that this only proves that what Chalmers calls the 'psychological aspect of consciousness' is consciousness *tout court*. However, this need not be the case, for, as I have already pointed out in response to Sellars, there is nothing self-contradictory in treating

³⁶ Jaynes (1976), p.47. These quoted observations represent a summary of Jaynes's argument in the whole first chapter of the book – see pp.21-47, 'Chapter 1: The Consciousness of Consciousness'.

³⁷ Apart from those detractors who consider phenomenal consciousness not to exist at all (e.g. Dennett), there are those who, like Jaynes, believe that it exists, whilst pointing out that phenomenal consciousness is not 'consciousness *simpliciter*', and that there must therefore exist a sort of non-phenomenal consciousness (which is what Jaynes is arguing for). For instance, see Lormand (1996), pp.242-61, who argues that such things as moods and attitudes constitute nonphenomenal conscious states (i.e. they are conscious, in as much as they display an awareness of things, and nonphenomenal in as much as there is not necessarily 'something it is like' to have them). Lormand thus distinguishes between moods and attitudes, and the phenomenal conscious states that moods and attitudes can (but need not always) give rise to (just as depression can give rise to dark thoughts, but need not itself become phenomenally conscious).

³⁸ Jaynes (1976), p.47. Jaynes's wider thesis – that, indeed, there was a time in history when this was the case – is a broad-ranging and fascinating one, covering religion, mythology, anthropology, ancient literature, the development of language, and much more. However, in utilising his initial argument here, I am not necessarily committing myself to his ultimate conclusions. However, the idea that phenomenal conscious – even when it is treated non-reductively – need not be involved in any of these activities is an important one. Of course, at bottom, Jaynes's argument is historical and would seem to be ultimately unverifiable (a sort of 'problem of other minds' at a remove of two thousand years). The presence in contemporary philosophy of mind debates of the subject of 'zombies' however, and the persistent attempt to show that so-called phenomenal consciousness can be 'explained away', shows that there is at the basis of Jaynes's theory an important observation.

³⁹ In fact, I think this is the case – as does Lormand (1996), pp.254-7, who argues that phenomenal consciousness is necessary in order for nonphenomenal consciousness to exist, and that the possibility of making sense of nonphenomenal states (e.g. arriving at conclusions concerning dispositional beliefs) is dependent upon it.

phenomenal experience as a more primitive and non-conceptual aspect of consciousness.

Given the above difficulties, therefore, there are currently serious practical and epistemological difficulties involved in adopting the cessation of the capacity for phenomenal consciousness as a sufficient criterion for determining death. However, in a sense, they are only that, and there remains the logical possibility that phenomenal consciousness is in fact the criteriological candidate that we seek (at least, for a definition of death involving *intrinsic* features). However, this brings us to the second and more difficult problem of why we should equate death with permanent cessation of the capacity for phenomenal consciousness *at all*.

Why Choose Consciousness?

Why should we hold that the cessation of the capacity for phenomenal consciousness should be equated with death? In other words, diagnostic and epistemological difficulties already noted aside, why should consciousness be considered as significant or essential?

So far, I have argued that no 'strictly-biological' definition of death is in fact possible which does not, in turn, require some degree of value ascription. Therefore, if we now apply this conclusion to the consciousness criterion, then on what grounds can we assume that consciousness is an 'essential' or 'significant' feature of a living human being? Is consciousness more vital than the other signs of life? What about cases where it is permanently absent, but other signs (e.g. heartbeat, respiration, etc.) are present? One advocate (such as Puccetti) may agree with the consciousness criterion, whilst another (such as Jonas), may find it inconclusive or dubious. As already noted in the range of views discussed in chapters 2 and 3, there would seem to be no purely logical means of determining exactly what the concept of a living human being consists of – and why, therefore, consciousness should be a defining feature of it.

Furthermore, with an attempt which seeks to identify specific physiological criteria for death, we once again face the problem of 'reversibility': if the functional state of certain areas of the brain determines the capacity for phenomenal consciousness, then there is the possibility of technological progress shifting that boundary (e.g. through the potential for repair of brain-cell damage). Are we therefore to abandon one criterion (reversibility) in favour of another (consciousness)? Therefore, even if we might argue that consciousness *is* an intrinsic feature of the world, and that thereby we might arrive at a definition of death based on intrinsic features, we are still presented with a value-driven definitional problem. Such competing criteria must therefore take their place alongside extra-biological features (such as – soon to be considered – personal identity and self-hood) in relation to an increasingly variable and shifting concept of death. Therefore, this conclusion not only suggests that the definition of death cannot be resolved in a strictly-biological (or strictly-logical) way, but furthermore that because of this, there may be a *number* of possible definitions (or *redefinitions*) of death which reflect value-based variance between individuals and cultures as to which criteria are significant.⁴⁰

⁴⁰ I shall explore what I have termed the 'potential variance' of the concept of death in chapter 8.

Conclusion

In this chapter I have explored the question of whether the permanent cessation of the capacity for consciousness can be considered a sufficient criterion for the determination of death. In doing so, I have argued that it is only the phenomenal aspect of consciousness (as defined by Chalmers) that can be considered in anyway to be an intrinsic feature of a human being.

However, whilst this initially seemed hopeful as a line of enquiry, the application of the criterion gave rise not only to epistemological and practical problems in determining death, but was also subject to many of the same problems that beset previously considered attempts. As argued in chapter 2, advocates of the strictly-biological definition, such as Lamb, Becker, and Russell, ultimately fail in their respective approaches because they do not recognise the role of *extra*-biological attributions of value, which in turn reveal the definitional dilemma itself as a conflict *between* such values. Furthermore, in chapter 3, possible resolutions of the problem according to a broader conceptual analysis were also shown to be problematic in that human concepts can be seen to be ultimately grounded in the non-rational, and such resolutions are therefore problematic for the reason that our rational concepts and categories are open to their variable influence. Lastly, the conclusions in chapter 4 have shown that any type of functional definition of death (upon which strictly-biological approaches would seem to depend) can be said to be an intrinsic feature of the world. Therefore, regarding the earlier concern respecting a possible twofold definition of death (i.e. human death and general organismic death), I think it is by now becoming apparent that this must at least be the case: in rejecting the strictly biological definition (and the contention that function is intrinsic), we are faced not only with the fact that human death cannot be the same as that shared by other creatures and natural entities, but by the possibility that death *in any sense* is itself merely observer relative. Ironically, then, some form of consciousness or personal identity approach (if successful) may be the *only* candidate for an account of death as an intrinsic feature; in this case, only human death could possess this status.

So, as we have seen in the course of this chapter, we have at least, in identifying a significant and – some might argue – essential feature of human beings (i.e. the capacity for phenomenal consciousness), isolated what is perhaps the only intrinsic feature which may play a part in defining death. However, as we have also seen, this does not in fact provide a resolution to the problem, for whilst some form of phenomenal consciousness may be considered an intrinsic feature of human beings, there would seem to be no logically compelling reason for us to identify this feature as essential – compelling in some *other* sense, possibly (which I shall return to later), but in terms of providing us with an essential feature which is independent of value-based choices, then no. A human being without the potential for consciousness may still be considered a human being, and a living one at that. Furthermore, in more practical terms, it may also be seen that a definition of death based on the absence of phenomenal consciousness would be extremely difficult to implement (though this in itself is no argument against the sufficiency of the criterion on a theoretical level).

However, that the use of phenomenal consciousness as a criterion for defining death subsequently proves problematic, or else cannot be linked to the concept of death in a way that makes death itself an intrinsic phenomenon (and thus resolve the debate),

does not in fact mean that phenomenal consciousness could not play an important role in the definition of death as an observer-relative phenomenon. I shall argue shortly that a number of the criteria discussed so far *can* play a part in defining death as a non-intrinsic (observer-relative) feature. However, before turning to these issues, it is first necessary to consider one last candidate for a non-observer-relative definition of *human* death: that of personal identity.

6. Death and Personal Identity

Introduction

Given the difficulties involved in defining death in what I have been calling a 'strictly-biological sense', it has occurred to some that these problems may be circumvented if we view death as a failure to fulfil those criteria necessary for personal identity (thereby adopting a 'partly-biological approach', as I term it). The best example of this sort of approach is that presented by Karen Gervais in her book *Redefining Death*, where she argues for a personal-identity based, neocortical definition of death.¹ In line with Green and Wikler,² Gervais argues that death in human terms (i.e. the death of humans) involves death of the person (as opposed to the mere death of the organism – which, however, she argues is sufficient for animals and other 'non-persons'). So, when the biological capacity which enables such personhood permanently ceases, the person is 'dead'. Since what is meant by 'person' here is the nexus of social relationships formed around the conscious entity that thinks of himself (for instance) as 'John Jones', a sufficient condition for the death of 'Jones' is the permanent cessation of the capacity to maintain those very roles, personal history, etc., that together make up 'Jones'. So, since these capacities generally reside in the neocortex or 'higher brain', neocortical death thus becomes – for humans (or at least, human *persons*) – death *per se*.

In this chapter I will therefore examine a range of personal-identity-based approaches to the definition of death, and conclude by underlining what I think are the general problems with such views.

Green and Wikler's Ontological Argument

Since Gervais is essentially adapting Green and Wikler's position in response to the criticisms of Agich and Jones,³ it is perhaps best, before considering the merits of such an approach in itself, to set out the genesis and development of this theory via the debate between its respective proponents and critics. Firstly, Green and Wikler characterise previous philosophical attempts to resolve the problems concerning the definition of death as taking either a moral or a biological slant.⁴ However, rejecting both these perspectives, they put forward what they term an 'ontological argument', based on the concept of 'being the same individual'. In these terms

Jones' death occurs *either* at the time that the patient dies, if the patient has remained Jones; *or* at the time the patient ceased to be Jones, whichever comes first. If, as we contend, the patient ceased to be Jones at

¹ Gervais (1986). Gervais's analysis of Green and Wikler, her response to the criticisms of Agich and Jones (1986), and presentation of her own version of the personal identity/neocortical definition of death, can be found in Chapter 5 of *Redefining Death*, pp.111-58.

² Green and Wikler (1980).

³ Agich and Jones (1986).

⁴ I will not consider Green and Wikler's criticisms of these respective approaches here as I have already extensively criticised the biological approach in previous chapters; the idea that death can be defined by reference to moral concerns, since it involves adopting an observer-relative definition of death, is a subject I will consider in subsequent chapters.

the time of brain death, then Jones' brain death is Jones' death. Thus, if the loss of capacity for mental activity which occurs at brain death constitutes death, it is not for moral reasons, nor for biological reasons, but for *ontological* reasons.⁵

Thus, for Green and Wikler, since 'brain death strips the body of its psychological traits',⁶ the person's identity cannot 'survive the kinds of changes which brain death involves.'⁷ In simple terms, death is seen as the death of the person, and brain death is therefore just a simple biological criterion for no longer being that same person.

However, whilst acknowledging many of the problems associated with providing necessary and sufficient criteria for personhood, Green and Wikler instead concentrate on the issue of whether a particular entity is the same *individual*. In doing so, the authors claim to circumvent these difficulties and provide a basis for establishing the continued existence – or not – of the individual concerned.

Thus, we can afford to remain uncommitted on whether persons are essentially beings with psychological properties and on whether Jones is essentially a person. We demonstrate instead that Jones, whatever kind of entity he is, is essentially an entity with psychological properties. Thus, when brain death strips the patient's body of all its psychological traits, Jones ceases to exist.⁸

Regarding the establishment of the identity of the individual 'Jones', Green and Wikler favour an account based on psychological continuity. However, since such continuity necessarily involves a functioning neocortex, Jones lives and dies – not with the fate of the body or organism as a whole – but with the capacity of the brain to sustain those psychological traits which constitute 'Jones'. Thus, 'Jones' may live on in a comatose body (which yet retains the neocortex intact), as a disembodied brain, a transplanted brain in another body, or – supposedly (though Green and Wikler do not use this example) – in another brain onto which the neurological organisation of Jones's 'old' brain has been 'mapped'. The point therefore is that 'Jones' lives whilst that which enables psychological continuity between one 'Jones-stage' and another also persists.⁹

Objections to Green and Wikler

There are various problems with this view. Firstly, one of the more controversial consequences of the approach is that the incapacity to sustain individual identity can frequently be found in an otherwise 'living' body – not only in cases of persistent vegetative state, but also where brain-damaged infants are born with functioning brainstems (in which latter case – arguably – no personality develops at all). However, Green and Wikler anticipate this objection, and are prompt to distinguish between the

⁵ Green and Wikler (1980), p.118.

⁶ Ibid.

⁷ Ibid., p.119.

⁸ Ibid., p.121.

⁹ That some of the science-fiction-type thought experiments considered by Parfitt (1984) may in fact cause problems for this view, I will consider below.

use of their criteria in order to determine death, and its utilisation as a ‘moral divide’ (to use Laurence Becker’s term).¹⁰ For the

argument, requiring no moral premises, is not sufficient to yield moral conclusions. We show that the brain dead are dead, but it does not follow that brain death is the appropriate moment to turn off ventilators or to remove organs.¹¹

Thus, Green and Wikler’s ontological argument is at most a foundation for further moral discussion concerning what we should do with those who are ‘dead’ in this way.

However, whilst they may have headed off this particular concern, others remain. One important difficulty is that, as Agich and Jones point out, simply ceasing to meet criteria necessary to maintain a certain model of personal identity is not always in itself sufficient for death. For example, it may be that only a persona – such as ‘Jones’ – is disrupted or irrevocably lost so that, whilst ‘Jones’ is ‘dead’, some ‘other’ being is decidedly alive. So,

The brain death issue is not whether that person is that person, Jones, or that person, Smith, as Green and Wikler insist; the issue is whether the patient is dead when his or her brain is dead *whoever* he or she is.¹²

A further criticism made by Agich and Jones is that Green and Wikler’s personal identity criteria for determining death would seem to represent ‘a rather onerous test’, for, as they somewhat humorously point out: ‘Given this requirement, it is surprising that anyone is ever declared dead.’¹³ For

when we want to know whether an individual is dead, we do not have to inquire into the identity of the individual. We only have to inquire whether the individual is dead. Green and Wikler’s excursus into the theory of personal identity confuses the central point at issue.¹⁴

Indeed, the absurd consequences of Green and Wikler’s view can be further illustrated if we consider that it is a test for death that can be readily performed on any living, conscious being – and prove positive! Successful candidates may include amnesiacs, stroke victims, those suffering from certain types of mental illness, and those suffering from senility. Green and Wikler’s motivation for the adoption of this type of personal identity criterion is to avoid both the problems that beset the moral and biological arguments for death, but also to sidestep the more problematic search for sufficient criteria for establishing the existence of personhood. Thus, they reject ‘kind-essentialism’ (what essentially constitutes a human being) in favour of an ‘individual-essentialist account’, and replace the question of whether *a* person is still alive, with the more specific question of whether ‘Jones’ is alive.¹⁵ However, as already argued, the fact that ‘Jones’ may die does not necessarily entail the death of the individual

¹⁰ Becker (1975), pp.334-359.

¹¹ Green and Wikler (1980), p.128.

¹² Agich and Jones (1986), p.269.

¹³ *Ibid.*, p.269, n.6.

¹⁴ *Ibid.*, p.270.

¹⁵ *Ibid.*, p.271.

(who, whilst he may not be 'Jones', may yet be a person), which would suggest ultimately that an individual-essentialist imposes, in a sense, too harsh a criterion in that it allows death to be pronounced when it would seem extremely controversial or even ridiculous to do so.

Green and Wikler may counter here that this type of criticism is unwarranted; after all, they are not (as they point out) setting a 'moral divide', but merely an ontological one (i.e. under certain circumstances, 'Jones' is dead). But, then, what force do their arguments have? 'Jones' is dead, but his body lives, and one may even conduct a conversation with him (albeit it as some nameless person, or 'Neo-Jones'). If they were to argue, for instance, that pronouncement of Jones's death allows us to treat Jones's social obligations, legal status, etc., as null and void, in some cases this would surely be a situation better handled by existing attitudes and legislation relevant to individuals who are mentally ill, senile, or otherwise incapable of making rational decisions (and would certainly represent a less 'jarring' option ontologically speaking). In other words, it is more common (and, perhaps, more in keeping with prevalent notions of personhood) to think of the man Jones as suffering in some way (e.g. from senility, or brain trauma), rather than considering the person 'Jones' to be dead.

On the other hand, a failure of the test as to whether an individual is 'Jones' which leaves a never-again-to-be-conscious – yet living – individual, would seem to create a new category of individual (or organism) devoid of personhood.¹⁶ Yet, surely, in making this distinction, Green and Wikler have already prejudiced any decision of what is to be done with this being by the mere fact of considering it 'not Jones anymore'; for indeed, the moment my grandfather ceases to be my grandfather, and becomes a semi-self-regulating organism (or whatever), is the moment many of my decisions are made for me. In this sense, therefore, Green and Wikler's divide is not, ultimately, a morally neutral one.¹⁷ On the other hand, however, if capacity for consciousness is maintained, then surely I would be reluctant to consider the individual as not really my grandfather anymore, even if he did not meet the criteria for individual personal identity. In such a situation, I would probably rather think merely that, 'My grandfather does not know who he is', or some such thing. In this sense, Green and Wikler's criteria serve to dehumanise the individual, and disrupt many of the common, natural responses we might have to them, merely for the purpose of solving our definitional dilemmas.

One last criticism of the Green and Wikler position is that, although they argue that 'Jones' is essentially an entity possessing certain psychological properties, they provide no real argument or justification of this viewpoint. For, as Karen Gervais points out,¹⁸ 'People may disagree on what Jones essentially is', and just assuming – as Green and Wikler do – that it is possible to make a clear 'patient/Jones' distinction, is not sufficient to prove that they are ultimately separable in this way. Of course, people can and do adopt such a view in relation to death, but this does not alter the fact that the adoption of such a view is a value-based decision in relation to what is important to the people adopting such a view. In other words, it is possible to have a wider concept of what 'Jones' is, and therefore when he ceases to be, and so even if it

¹⁶ Which, as Gervais (1986, p.114) argues, is an unnecessary and confusing dualism.

¹⁷ In fact, I argue that no such divide *can* be value-free in this sense.

¹⁸ Gervais (1986), pp.115-6.

can be demonstrated to me that, according to Green and Wikler's strict personal identity criteria, 'Jones' no longer exists, there is no reason to adopt this particular model of personal identity. I may, for instance, hold that the mental aspect of 'Jones' is only a facet of the person, and that, in an important sense (as with the example of my grandfather), 'Jones' still exists for me.¹⁹

Finally, it remains to be seen whether the account of personal identity involved here is itself a coherent one for Green and Wikler's purposes. It may prove, for example, that such an approach has its own inherent problems which may undermine its use as a criterion for the definition of death. That is, even if the death of 'Jones' is sufficient for death, the approach to personal identity employed here (i.e. psychological continuity) may give rise to other unforeseen problems – though this is a more general point, and one which I shall consider near the end of this chapter.

Karen Gervais and *Redefining Death*

Like Green and Wikler, Karen Gervais defends a neocortically based view of human death as death of the person.

Since human death is the death of the person, and the death of the person occurs with permanent loss of consciousness, neocortical death is an adequate criterion for declaring death.²⁰

However, unlike Green and Wikler, Gervais retreats from their individual-essentialism and proposes a 'more modest mentalist view'.²¹ This, she argues, avoids the type of problems that face a more restrictive view, allowing us simply to argue that the permanent absence of the capacity for consciousness should be a sufficient criterion for declaring an individual 'dead' (whether 'Jones' or not), and that the 'individual's essence consists in the possession of a conscious, yet not necessarily continuous, mental life'.²²

In doing this, Gervais argues that what is centrally involved in any definition of human death is what she terms a 'decision of significance'.²³ In other words, she argues,

Our concept of death is not a fact awaiting our discovery; rather, it must be chosen on the basis of ethical reflection.²⁴

¹⁹ Gervais provides a more detailed critique of Green and Wikler than I have supplied here, and I have also supplemented her position with points and criticisms of my own. She also identifies certain inconsistencies in their argument (such as treating anencephalic infants and 'neocortically dead' adults differently – thus requiring an additional, biological definition of death for the former). However, it is not my purpose here to revisit Gervais's every point and observation with a view to agreeing or dissenting, but merely to show that there are good reasons for rejecting the individual-essentialism of Green and Wikler. Ultimately, though, I will argue that there are also good reasons for rejecting any sort of personal-identity based definition of death.

²⁰ Gervais (1986), pp.150-1.

²¹ *Ibid.*, p. x.

²² *Ibid.*, pp.157-8.

²³ *Ibid.*, p.156.

²⁴ *Ibid.*, p.155.

This conclusion – a consideration which seems unrecognised among most other commentators²⁵ – is one that I have great sympathy with, and in many respects represents a key note of what I have been arguing to date: ‘death’ is not something which exists in the world (intrinsically), independently of human value systems and conceptualisations (though I would not say necessarily that the value-based aspect is purely ethical in nature, nor that the decision which ‘resolves’ the situation purely a moral one, but I shall return to this point in following chapters).

Death is not simply “a matter of scientific fact,” as Lamb would have it. We must make an ethical decision about how best to regard the human for the purpose of declaring death.²⁶

Gervais, however, does not go into the reasons as to *why* death has this status (such as I have attempted in chapter 3, and will return to later), nor does she attempt to provide a theoretical framework within which she can distinguish between concepts which may be said to represent intrinsic features, and those which carry non-agentive ascription of value and purpose (such as I have also attempted to do in utilising Searle). As a result, Gervais seems to be left with a twofold concept of death: on the one hand, what we might call ‘organismic’ or ‘natural’ death (which human organisms might share with animal organisms, and other natural entities); and on the other hand, human, ‘person-centred’ death.

The death of a person is different from that of a dog or a cat; every death of a person unravels and reconstitutes a complex net of rights and obligations that usually involve many people. Death is not simply an event in the life of the deceased. We must think of human death not in terms of what it has in common with canine death or feline death, but in terms of what it represents for human relationships: abandonment of all roles, the end of all interactions, and the reconstituting of rights and obligations. Permanent unconsciousness, whatever its basis, represents these changes.²⁷

This discussion is similar in nature to the one at the other end of the life continuum (i.e. that concerning abortion). Michael Tooley,²⁸ for instance, argues that we may distinguish between the concept of a person and the concept of a human organism (a representative of the species *Homo sapiens*). This distinction, Tooley argues (parallel to Gervais), allows us to treat a non-person (in this case, the foetus), differently to a person (a conscious entity): the latter, by virtue of its consciousness, is a rights-bearer, whereas the former, by virtue of its lack of this feature, is not. This distinction is also defended by others, such as Peter Singer²⁹ and Judith Jarvis Thomson,³⁰ but there are also those who disagree that such a distinction is possible – or, more precisely, that criticise this approach for misconstruing the *nature* of this distinction. For instance, Bernard Williams criticises Tooley for glibly ignoring the social consequences of his

²⁵ With the possible exception of Jonas (1978) and Evans (1990, 1994), discussed in chapter 2.

²⁶ Gervais (1986), p.152.

²⁷ Ibid.

²⁸ Tooley (1992).

²⁹ Singer (1993).

³⁰ Thomson (1992).

arguments: what sort of world would this be if such arguments held?³¹ Singer, Thomson, and Tooley therefore seem to assume that such a distinction is significant *in itself*, but surely, as I have argued, it is not that easy or simple: as Williams points out, it is the wider social significance of such questions which have a bearing on the meaning of the distinction, and not the other way around. How we treat a non-person is determined by our general views on life, the value of existence, and other non- or semi-rational attitudes, and not by the mere fact (if that is so) that they do not possess personhood.

A different criticism is presented by Eric Olson.³² Tooley, Singer, and Thomson are all representatives of what Olson calls the ‘Standard View’ of personal identity, whereby organism and person are clearly distinguished. However, as Olson argues, this leaves us in the bizarre position that – according to the Standard View – ‘I’ am not identical with the foetus that my body grew out of. This is because a foetus possesses no psychological properties, and the defining features of personhood are psychological (rationality and self-consciousness). Olson highlights the difficulties and absurdities inherent in this view well: the Standard View’s inability to account for the relationship between the organism and the person results in a situation where we are almost led to talk of the existence of two entities.

This view entails that, although we are material beings, we are not human animals: we are not members of the species *Homo sapiens*. Apparently we are not organisms at all, in spite of appearances – even though we are alive and are composed entirely of living tissues arranged in just the way that the tissues of a living organism are arranged.³³

This, he points out, is the reason for many of the strange situations that crop up in the personal identity debate (such as, in fact, Parfit highlights – which I shall deal with shortly).³⁴

In relation to the definition of death, there is therefore the similar problem that human beings embody two aspects: they are both organisms and persons. Gervais presents the definition of death problem as a choice between these two at-times-divergent features of a human being, and it’s resolution as the provision of justification as to the adoption of one or the other definition (in her case, death defined as death of the

³¹ Williams (1995).

³² Olson (1997).

³³ *Ibid.*, p.101.

³⁴ Olson’s own solution is to consider that identity only entails between biological entities (i.e. between the foetus and the adult organism), what he calls the ‘Biological View’. Thus, for instance, this would avoid the well-known problem of brain transplantation scenarios: if some other entity possesses all or part of your brain, then they simply possess part of you (the mental aspect), just as they might possess any other non-essential feature (such as a limb or an organ). I think, however, that whilst Olson’s attack on the Standard View is well conceived, his solution may have its own flaws – not least, the observation (already considered) that deciding between the persistence of mental qualities and organismic functioning and integrity is arguably something that already implies metaphysical commitment and possession of certain values. Furthermore, it would seem odd to deny that the mental element is in some way inessential in the way that Olson thinks it is. My own position is therefore to reject both these views in favour of the possibility that many conceptions of self and death are in fact possible (multiple ‘redefinitions’).

person).³⁵ In her eyes, this justification must take the shape of a moral argument. In other words, we must investigate what moral consequences there are for adopting organismic or person-based definitions of death in certain problematic circumstances. Of course, given that most of these problematic situations involve the permanent absence of the patient's capacity for consciousness, the moral considerations tend to centre on the consequences for the patient's relatives, the strains put upon the health service, the needs of the organ transplant programme, etc., the needs of the patients themselves, whilst not ignored (we should respect living wills, religious beliefs, the dignity of the dying, etc.), are outweighed by these other considerations. Furthermore, since these external concerns are better acknowledged by adopting a person-centred concept of death (and merely hampered or ignored by an organismic concept), then these moral considerations persuade us that the person-centred definition is the one to adopt.

Human death should be regarded as a person-centred occurrence. Since permanent unconsciousness is the sign of personal death, the brain-death and neocortical-death criteria are consistent with this concept. Treating such individuals as living human beings must be disallowed on the grounds that there is no moral cost to them in our considering them dead, and high moral cost to others in our considering them alive.³⁶

The adoption of this definition of death therefore takes a contractual form, wherein parties of differing views must arrive at an 'outcome that each person realises in advance is optimal for both'.³⁷ This is then put forward as a template for public policy, arguing that we must avoid foisting a highly technical, perhaps counter-intuitive definition of death (which Gervais admits neocortical death might be, considering our emotional reactions to patients in a persistent vegetative state) upon a somewhat sceptical and under-informed public. Therefore, she argues, so that they can play their part in the contractual agreement as to what death is, a thorough educational programme must take place to inform the public of the relevant issues involved.

Responses to Gervais

Firstly, it should be noted that in some respects I agree with Gervais's approach. On the one hand, for instance, she identifies that the adoption of one or other definition of death implies what she calls a 'decision of significance' as to which features of the individual – spontaneous function, consciousness, personal identity – are to be considered significant or essential features of the human being, so that their absence might be a firm indication that an individual is 'dead'. However, as I have already argued, such 'decisions of significance' in turn represent value-motivated attempts to draw the borderline in a particular way. In other words, what is important to one person may not be so important to another; one person may be content to consider insignificant the fact that a patient may have unaided respiratory and cardiac function, because the potential for consciousness has permanently disappeared; another may

³⁵ As should now be apparent, this is not the way I see the problem; for me, both the concepts of 'functioning organism' and 'person' are non-intrinsic and observer-relative. Therefore, they *both already* represent value-based ascriptions. I shall unfold the consequences of this view in a later section in this chapter.

³⁶ Gervais (1986), pp.175-176.

³⁷ Ibid., p.167.

point to the presence of these functions as indications that the being in question is still alive. Deciding between these two viewpoints is not something that can be done by strict appeal to 'objective facts', since many of these facts still leave the question open. So, even if we say, 'the patient will never again be conscious', the further questions which arise are, 'How do we value consciousness?' and 'Do we value it over and to the exclusion of spontaneous function of the organism?' Obviously, I am not saying that we think exclusively or even explicitly in terms of value, and furthermore it would be wrong to consider such questions simply as concerning value; more precisely, we might say that our ontological judgements (What is a person? When is he dead?) are informed, or motivated, or tied-in with value-based assertions (that consciousness matters, that spontaneous function is significant, etc.) which – as I have suggested in chapter 3 – it may turn out cannot be cashed out rationally, but which at the very least represent ineradicable and irreducible aspects of such concepts.

Given that our judgements at even a fundamental level (such as the non-agentive ascription of a certain purpose or function to a living organism) are 'double sided' in this way (containing fact and value), the genesis of the problem concerning the definition of death can be seen in a whole new light. As I have so far argued, the problem as it stands exists because the development of technology has allowed the possibility of situations where values and concepts which previously existed in harmony were put into conflict with one another. So, in creating the possibility of a 'beating-heart cadaver', in invalidating the cardio-respiratory criteria for death, advancements in medical technology have created situations in which particular aspects of certain concepts – e.g. 'a living person' – have been set against one another.

I have also argued that the reason this very division of concepts is possible has something to do with the necessary opacity of concepts in general. As Quine and Wittgenstein were among the first to point out (in their different ways), there must come a point where a term is referentially opaque, and there can be no more analysis of the term in yet more fundamental ways. Of course, such limits are relative, and – as Quine further argues – it is always possible to more fundamentally explain one 'conceptual scheme' within another (but which *itself* is only relatively analysable). Furthermore, the way a term is used and what it means may be influenced by empirical matters – such as, for our purposes, advancements in medical technology. Thus, I argue, what we are faced with in the problem of defining death stems from a previously opaque term (e.g. 'living person') being divided, and its parts exposed and seen – under these new conditions – to be potentially contradictory.

I will deal with what form possible resolutions to this problem may take in subsequent chapters, but I now wish to show how the picture of the problem as I have just set it out can help us evaluate Gervais's position – or, in fact, any personal identity argument.

The Twofold Definition of Death

As we have seen, Gervais's conclusions lead her to consider death in a twofold manner: human (personal-identity based) death and animal (organismic) death. Whilst, initially, she presents these two views as competing candidates for human death, she ultimately decides upon the former (personal identity) on moral grounds. However, whilst I agree with her that this problem cannot be resolved by a strictly-biological

analysis of organismic functioning, and that any attempt to decide between organismic and personal-identity based definitions of death requires a 'decision of significance', to adopt two definitions of death (one for humans and another for other living entities) would seem to leave us with an awkward and somewhat arbitrary dualism. Furthermore, it would mean that a human could be 'alive' in organismic terms whilst 'dead' in terms of being a person. However, since this is the very problem that we started with, is it ultimately satisfactory to adopt a model that allows this?

Gervais's reasons for adopting this twofold definition can be summarised as follows:

1. Any strictly-biological, organismic definition of death is in itself insufficient (in that it requires a 'decision of significance' to work – e.g. that the continuing integrated functioning of the organism is more important than capacity for consciousness).
2. Considering human death as the death of the person (a partly-biological approach) also requires a decision of significance, but it is preferable on moral grounds.
3. The current dilemma (as to how death should be defined and when it should be declared) therefore requires a redefinition in terms of what physiological functions are 'essential to the existence of the person',³⁸ so that the permanent cessation of these functions constitutes the person's death.

Firstly, therefore, Gervais would seem to assume that an organismic definition of death, whilst insufficient for human purposes, is nonetheless sufficient for the purpose of defining non-human organismic death (e.g. that of animals and other organisms). However, the organismic (or functional) definition is insufficient not only for humans, but in general. By 'insufficient' here, of course, I don't mean that it is completely, under all circumstances, unworkable, but merely that it cannot be cashed out via a 'strictly-biological' analysis involving purely intrinsic features. If Gervais's twofold definition is based (as it seems to be) on being able to strictly distinguish between organism- and person-centred death in this way, then it is consequently flawed: the functional/organismic definition of death is as much reliant upon our own (human) ascription of value and purpose as any definition based on the incapacity for consciousness (as I have argued in chapter 4 in my rejection of the notion of intrinsic function). In an important sense, therefore, both definitions embody a concern with value: an organism has spontaneous, integrated function, therefore it is still alive; an individual retains the capacity for consciousness, therefore it is still a living person.

Gervais's rejection of the organismic definition stems from the fact that, for her, the capacity for consciousness (however minimal) is the *sine qua non* of personhood. A functioning organism without the capacity for consciousness is 'dead', whilst an even-minimally-conscious person, whose organismic function is artificially maintained – or even one whose brain waves have been 'downloaded' onto a computer – is still alive. However, given my above assertion that both these concepts ('organism' and 'person') are value-based and normative, in what sense is a twofold definition of death warranted?

³⁸ Gervais (1986), p.177.

Regarding her second point, Gervais argues that an analysis of what is meant by the term 'human being' does not provide us with a direct means of resolving the death debate. Fundamentally, this is because certain technologically-instigated situations allow for important aspects of the whole human being to be absent (e.g. the 'organism' may be alive whilst the 'person' is 'dead'). I agree with this, and have so far been at pains to show how I think the nature of human concepts and necessary limitations in their analysis, in conjunction with advancements in medical technology, have given rise to this situation. However, having arrived at this point, Gervais concludes not only that we must redefine death, and that an ontological analysis alone cannot help us, but furthermore that this redefinition must have a moral basis. In other words, since we cannot by ontological analysis alone determine sufficient criteria for whether the human being as a whole has ceased to exist, we must decide (via its moral consequences) whether death should be thought of in terms of the 'organism' or the 'person'.

The problem with this, of course, is that just because it can be shown that a problem is not ultimately solvable in *definitional* terms (by essentialist-type appeals to what something 'really' is), does not mean that our one necessary recourse is to a *moral* justification of one path or the other. Merely because we cannot decide whether a certain creature is a mammal or not, or whether a certain astral body should be classed as a planet, does not mean that such things are therefore to be decided by appeal to the moral consequences of adopting one definition or another.

The final step of Gervais's argument involves the identification of those physiological criteria whose continued functioning is essential to the continued existence of the person. Whilst at first this may seem to be a reasonable position – provided that the personal-identity basis for death is accepted, of course – it does seem to involve the same type of problem concerning 'kind essentialism' that Green and Wikler were seeking to avoid in employing their notion of 'individual essentialism'. For Gervais's argument to succeed she must be able to clearly distinguish between those features which may persist in integrated function of the organism, and those which are essential for the persistence of personhood – and here we are again faced with the same type of classificatory dilemma which, as I have argued, constitutes the essence of the definition of death problem itself. Where does the person³⁹ stop and the organism begin? Can the two be exclusively defined? The answer to these two questions, I argue, is that, for reasons stemming from the way in which the concept of 'person' has over time and through communal linguistic usage been formed, there is no compelling way in which this task may be accomplished (without, at least, being open to the charge of conceptual gerrymandering – which, actually, I think most attempts at resolving this conflict to date *are*). To illustrate this in more detail, I will now consider what I have termed the 'classificatory dilemma' in relation to the general problem of personal identity, and thereby show why it cannot be used as a means of resolving the problem of the definition of death.

³⁹ Here, and in the following sections, I use 'person' in its widest sense – i.e. not merely in Green and Wikler's sense of a conscious and psychologically intact entity (i.e. 'Jones'); a 'person', in my sense, may include some of the very features that Green and Wikler wish to exclude (i.e. physical processes, physical and bodily continuity, and so on). I believe this wider concept is more reflective of common usage, and hence – where we consider the definitional problem of death to stem from the conflict of observer-relative values (as I believe it does) – more relevant to this discussion.

The General Problem with Personal Identity Arguments

In the preface to the first volume of his autobiography, the writer Anthony Burgess considers the nature of producing such an account of one's life:

The delineation of oneself at such length must look like egocentricity, but an autobiography has to be egocentric. On the other hand, what do we mean by the ego? It is an existential concept, I believe, and the ego I examine is multiple and somewhat different from the ego that is doing the examining. Even the ego that began the book in September 1985 is not the one that has completed it in 1986. In other words, the book is about somebody else, connected by the ligature of a common track in time and space to the writer of this last segment of it, which cheats and looks like the first.⁴⁰

This somewhat sceptical picture presents the problem of personal identity in a nutshell: how, without 'cheating', can we establish the 'ligature of a common track in time and space' between all the different physical stages and mental states of the same person's life? In terms of the philosophical history of the problem, the two main approaches attempt to establish continuity either via psychological or physiological criteria (or, perhaps, some combination of the two). Thus, an individual is the same (has the same identity) as a particular individual in the past if we can establish continuity between either the existence of the physical bodies of those individuals (or significant parts thereof), or continuity between some aspect of the mental functioning of these two individuals. In terms of physiological criteria, I might argue that I am the same person as a certain 10-year-old that I see in a photograph, because it can be demonstrated – through the resemblance of our physical features, identifying marks, etc. – that in a significant sense that body is connected to mine. Similarly, I might argue that I *am* that 10-year-old, because I can *remember* doing and experiencing certain things that happened to that child (and, perhaps, remember looking in the mirror and seeing the same facial features, hairstyle, etc., in the photograph).

There are many objections and consequent refinements relating to both these approaches. The standard objection to the physiological approach is called the example of 'Theseus' ship',⁴¹ whereby each plank in a wooden ship is replaced by a new one over a period of time, until eventually the ship consists of totally new planks; therefore, is it the 'same' ship? In relation to the psychological argument, common objections include the difficulty of whether a memory-based criterion for psychological continuity can account for such things as the memory gaps caused by brain damage, etc. However, it is not my purpose here to revisit the whole sequence of arguments, objections, and counter-arguments which make up the vast personal identity debate, or to pass judgement upon the various candidates for one or another of the approaches to solving the problem. Rather, I am merely concerned to show in what way I think the problem of personal identity reflects a more fundamental problem; one that is inherent in the application of a number of types of concept, and stems from limitations to do with the nature of categorization and concept formation.

⁴⁰ Burgess (1987), p.viii.

⁴¹ See, for example, Maslin (2001), pp.249-50.

In *Reasons and Persons*,⁴² Derek Parfit argues – successfully, I think – that there are (at least logically) plausible situations in which we are faced with a choice as to whether an individual can be said to possess personal identity or not. In a now famous thought experiment, Parfit envisages that a ‘Teletransporter’ exists which can copy a person’s exact physical constitution, send the information as a radio signal to Mars, and reproduce that individual there (whilst destroying the individual’s body on Earth).⁴³ However, in the scenario, the Teletransporter malfunctions, and the original’s body, instead of being destroyed, is merely fatally damaged. The original person is then informed of the malfunction, told that they will die from cardiac failure over the next few days, and invited to come to terms with their temporary ‘dual existence’ and imminent ‘death’.

Parfit ultimately concludes that there is no personal identity criterion which can be used to decide between such a case (or even similar, modified cases), and that the choice which faces us is therefore a real one. Furthermore, he concludes that we must accept that in such cases we have what he terms ‘survival without identity’, in that there is a sense in which the other ‘me’ shows that ‘I’ will survive, but not in such a unique way as to ensure the integrity of personal identity. Thus, for Parfit, there is no possibility of a concept of personal identity which would not allow this sort of aberration to happen. Therefore, we would be better off in such a situation to wish our Martian counterpart well, and accept his continued existence in light of our own demise as the best of a bad lot.

If we now view this sort of problem in light of the already discussed classificatory dilemma, we shall see that it is ultimately analysable in the same form. Whether the scenario concerns a dual existence occasioned by a malfunctioning Teletransporter, one where an individual’s brain is split and put into two different bodies, or any other similarly problematic scenario, the central difficulty lies with maintaining the dogma of binary opposition (that a person is either ‘Jones’ or ‘not-Jones’) whilst establishing criteria whereby the status of cases may be unambiguously pronounced upon. Similarly, therefore, Parfit’s science-fiction type examples allow scenarios in which the concept of ‘person’ can be split, and reveal conflicts between its constituent elements – or, also, conflicts between our assumptions as to what a person is and our use of the term – *much in the same way* as technological advancements have allowed the development of similar scenarios and quandaries in the definition of death debate.

Gervais on Personal Identity

In terms of its specific application to Gervais’s position, therefore, it can be shown that her position ignores the sorts of problem which Parfit highlights. Her notion of personal identity can therefore be shown to be vulnerable to two types of objection – both of which ultimately spring from the fundamental problems inherent in categorization and concept formation which I have identified.

Firstly, as I have already noted, Gervais considers consciousness as the *sine qua non* of personhood. However, not only does she not define what she means by the vague term ‘consciousness’, but she is therefore unaware of the problems faced by anyone

⁴² Parfit (1984).

⁴³ *Ibid.*, pp.199-201.

seeking to differentiate between conscious and non-conscious, volitional and instinctive/automatic functioning, etc.⁴⁴ Furthermore, Gervais seems to conflate 'consciousness' with 'possessing psychological properties', yet there is by no means an automatic equivalence between these two things. For instance, it is common to consider behavioural traits and emotional proclivities as belonging to the personality, or even as an indication of someone being a particular person (e.g. 'That doesn't sound like something he would do'). Yet these are not exclusively the properties of consciousness, and most importantly there are a whole host of qualities which we may associate with a person which have a non-rational, non-conscious – even purely physical – basis. In this latter case, a person's physical appearance, some distinct physical quality that they possess, etc., may arguably be as much of the 'person' as any other 'conscious' quality. In this way, Gervais's position ignores the wider sense of 'person' that this picture suggests, and therefore the problem of decidedly divorcing and identifying aspects of it as essential.

The second form of objection considers the other side of the coin in pointing out that such science-fiction scenarios as Parfit and others have envisaged – and Gervais herself considers at one point – can be used to undermine the possibility of unique personal identity. In considering whether it might be possible to consider a person as still existing when their memory had been 'transferred to a computer tape with an input-output mechanism that would allow ongoing interaction between the tape and the rest of the world'⁴⁵ (a possibility first envisaged by Robert Veatch),⁴⁶ she concludes:

It is difficult in principle to deny that the tape is the same person. Any claim that the tape is not embodied in a human body begs the question. This suggests, contrary to the writings of Veatch, that *personal identity is preserved as long as a personal history is preserved*, and that embodiment in a human frame is not a necessary condition for the retention of personal identity.⁴⁷

Having accepted the possibility of such survival, however (or at least its cogency as an objection to be used against those who argue for the important role of the physical brain in view of its irreplaceability and uniqueness), it is surprising that she does not follow through the computer analogy to its logical conclusions. In other words, an obvious consequence of being able to reduce an individual to a finite set of data (by transferring their 'brain waves' to 'computer tape', or recording the state of the cells in their body for radio teletransportation to Mars), is that such information is *duplicable* – which, if we follow Parfit's reasoning, undermines the possibility of personal identity criteria which guarantee uniqueness, and allows for what Parfit terms 'branching' cases.⁴⁸

⁴⁴ Discussed in the previous chapter.

⁴⁵ Gervais (1986), p.129.

⁴⁶ Veatch (1986), p.69. Veatch holds that the essential feature of human beings (that makes them alive) is the capacity for social interaction. However, whilst considerations of personal identity would seem to be relevant to this argument, Veatch does not go beyond this assertion; therefore, I have not considered his arguments here, but briefly later (see chapter 8).

⁴⁷ Gervais (1986), p.131 (the italics are mine).

⁴⁸ Parfit does appear in the bibliography to Gervais's book, but is not anywhere mentioned in the text.

Furthermore, the brain-splitting experiments such as Parfit considers present a parallel case to the determination of death debate: if we can house the two halves of someone's brain in two separate bodies, which is the 'real' person? Making a decision itself requires a Gervaisian 'decision of significance' as to which half is more important, how much is 'enough' of someone's brain, etc.

However, we need not in fact resort to Parfit's science-fiction scenarios in order to consider examples which are problematic to the personal-identity-based approach. Thomas Nagel, for instance, considers the problems presented by cerebral commissurotomy (brain-bisection), whereby the *corpus callosum* (the bridge which connects the two cerebral hemispheres) is severed.⁴⁹ As a result, it has been found that the left and right hemispheres can both receive and process information independently, can work simultaneously, and even at cross purposes. Such findings, Nagel argues, naturally beg the question as to how best to explain this situation: how many minds do we have?

There is a powerful inclination to feel that there must be *some* whole number of minds in those heads, but the data prevent us from deciding how many.⁵⁰

This inconclusiveness, Nagel argues, must lead us to revise the assumptions we have about *normal* mental functioning, and that 'these very unusual cases should cause us to be skeptical about the concept of a single subject of consciousness as it applies to ourselves.'⁵¹ Perhaps, he concludes, we may even one day come to think of the concept of a single person as 'quaint'.⁵² Roland Puccetti similarly argues that we may credit different activities to these different 'selves':

I cannot believe my right-sided cerebral companion contributed much to the writing of this paper, for example, or understood it as he saw me writing it out. He leaves the philosophising to me.⁵³

Therefore, the brain-splitting experiments such as Parfit, Puccetti, and Nagel consider present parallel cases to the determination of death debate: if we can house the two halves of someone's brain in two separate bodies, or cause the two halves of the same brain to behave independently, then which is the 'real' person? How many 'persons' are there?

Some have sought to resist this sceptical approach. For instance, John Eccles has argued that we may think of the activities of the non-verbal right-hemisphere as unconscious and automaton-like.⁵⁴ However, as Puccetti points out, we have as much reason for attributing conscious willed action to the right, non-verbal brain as we do

⁴⁹ Nagel (1979c). For an account of the original experiments, see Gazzaniga, Bogen, and Sperry (1962), and for a wider account and more general discussion, see Carter (1998), pp.42-3, and Blackmore (2003), pp.103-8.

⁵⁰ Nagel (1979c), p.161.

⁵¹ Ibid., p.163.

⁵² Ibid., p.164.

⁵³ Puccetti (1973a), p.354.

⁵⁴ Eccles (1967, 1970).

for the left, verbal brain (for, after all, speech may be unconscious).⁵⁵ Another objection considered by Puccetti is raised by Donald MacKay,⁵⁶ who argues that we may consider the apparently separate actions of split brain patients as under the control of a ‘metaorganising system’, or unified self. However, Puccetti responds,⁵⁷ apart from the fact that there is too much evidence that the two hemispheres can often act at complete variance to each other in split-brain patients, there would seem to be no reason to postulate such an overarching entity at all. They seem to have no awareness of one another, and MacKay’s attempt would seem to be simply a metaphysically-motivated attempt to retain the old-fashioned notion of self.

However, this said, could I make a decision as to which aspect of my mental functioning was in fact the real ‘me’? As Puccetti implies, philosophers and writers (those favouring the verbal hemisphere) would choose the left hemisphere (in most people); however, this is not automatically the case, and artistic and creative types would perhaps make the opposite choice (again, in most cases). Making such a decision itself would therefore seem to require not only a Gervaisian ‘decision of significance’ as to which half is more important, how much is ‘enough’ of someone’s brain, etc., but also a metaphysical commitment to the existence of a single (Cartesian?) self – or that, faced with competing options, we must choose *one*.⁵⁸

Conclusion

In this chapter, I have argued that personal identity approaches to redefining death fare no better than other approaches already considered – and for similar reasons. Gervais’s notion of personal identity would therefore seem to be problematic in that it relies on assumptions which are themselves highly controversial and questionable. The notion of personal identity, therefore, rather than supply a means of resolving the definition of death, merely mirrors the same type of definitional problem (as I have outlined it) on another level. Personal identity approaches in general, therefore, in as much as they seek to provide sufficient criteria for the survival of a unified self, rely on a number of assumptions already criticised: that the essentialist assertion that a concept or entity must embody a specific set of criteria so as to represent a distinct ‘kind’; that such features can be reduced to non-observer-relative features of a human being (i.e. the capacity of the brain to produce consciousness, and thereby personhood).

Furthermore, however, Gervais’s attempt to resolve the debate morally (via a contractual method), would seem to admit that the whole debate is actually a cultural one. In as much as morality may be seen as an aspect of our social discourse, and therefore not an intrinsic feature of the world (that is, of course, if we reject certain forms of moral realism), then moral values must take their place alongside other observer-relative features of the world. This overall view, I shall argue, in turn opens

⁵⁵ Puccetti (1973a), pp.341-2. It is only a generalization that the right-brain is non-verbal, and experiments have shown that it may possess some basic linguistic ability. Furthermore, it is also a generalization that it is the right brain which is the non-verbal one – this may be reversed in left-handed people (to a certain extent), and brain dominance (in terms of verbal/non-verbal processing) can differ from individual to individual. See, for instance, Carter (1998), p.46-7.

⁵⁶ MacKay (1969a, 1969b).

⁵⁷ Puccetti (1973a), pp.342-3.

⁵⁸ Of course, where verbalisation/spatial awareness is not lateralised to such an extreme, then the choice becomes more complicated.

up the definition of death debate to the influence of a wide array of competing value-driven influences – to a world, in fact, where concepts such as death are more potentially variable than we would perhaps like to think, or that Gervais recognises.

What these values are, and what this means for the debate, I will now consider.

7. An Observer-Relative Death

Introduction

So far I have considered the case for death as an intrinsic, ontologically-objective feature of the world. This has taken the form of the ‘strictly-biological’ views of such as Laurence C. Becker,¹ David Lamb,² Christopher Pallis,³ and Tom Russell,⁴ who favoured a functional definition of death, and such ‘partly-biological’ views of Green and Wikler,⁵ and Karen Gervais,⁶ which attempted to show that the permanent incapacity of the neocortex was a sufficient condition for considering the ‘person’ dead. However, as I hope I have shown, both these approaches fail in resolving the death debate: firstly, they fail to show that (in Searle’s sense) death is an intrinsic feature of the world; secondly, as a consequence, both approaches fail to account for the fact that, as death becomes an observer-relative feature of the world, so the putative solutions considered merely beg the question, ‘What is it that we consider to be essential to the concept of a living person?’ Heartbeat and spontaneous respiration are significant to some,⁷ whilst the absence of the capacity for conscious thought is decisive for others;⁸ that the individual (‘Jones’) is no longer with us may be sufficient for his death (some argue), whilst for others there is a wider concept of a person of whom the integrity of psychological personal identity is merely an aspect. Because there are competing answers to this question, and because the observer-relative concepts that it involves are (to an extent) changeable and culturally dependent, then the above approaches merely represent attempts to ‘gerrymander’ the constituency of death. Having reached this point, I therefore wish to argue that the concept of death *cannot* be determined by reference to brute facts or intrinsic features, and it thus has far more in common with other ‘epistemically-thick’ concepts in ethics, aesthetics, or other cultural phenomena (even sport). It is to comparisons with questions in such fields that I will therefore now turn.

A Note on Physicalism

However, before I do so, it is perhaps only fair to anticipate one last objection: it may be counter-argued here that, in dismissing the functional, consciousness-based, and ontological (personal-identity based) definitions of death, I have not thereby dismissed the possibility that death is an intrinsic feature of the world *in some other sense*. In reply to this I can only point out what I take to be a decisive point: in categorizing something or someone as dead, we do not merely describe causal properties of physical objects (i.e. their purely intrinsic features), but ascribe significance based on the value that we perceive those properties to have in our own application of related concepts (i.e. ‘person’, ‘alive’, ‘quality of life’, and so on). Descriptions of causal relationships are therefore in a sense neutral in a way that

¹ Becker (1975).

² Lamb (1978, 1985).

³ Pallis (1983a, 1983b, 1985).

⁴ Tom Russell (2000).

⁵ Green and Wikler (1980).

⁶ Gervais (1986).

⁷ E.g. Jonas (1978) and Martyn Evans (1990, 1994).

⁸ E.g. Puccetti (1976).

ascriptions of life are not. To illustrate this, consider the following scenario: a robot is built with the identical causal properties as those possessed by a human being. Setting aside the question of whether this would in fact make it 'conscious', can we consider it alive? On what basis do we ascribe life to it? I argue that those who would reject ascribing living status to this robot do so for reasons over and above an analysis of such a thing's causal properties, and that therefore, the ascription of life to it entails something more than that; furthermore, these ascriptions are observer-relative, and so cannot be said to be 'entailed' by the causal properties themselves; as for life, so for death.

But, it may be objected, wouldn't the causal properties of a purely physical universe *necessarily* entail certain social facts? Or, wouldn't it be necessary to posit the existence of non-physical facts, or non-determinism, or some other means of avoiding fixing the form of our social concepts? Therefore, is the notion of physicalism an objection to the possibility of potential variance within socially constructed concepts?

In her article, 'Conceivability, Possibility, and the Mind-Body Problem', Katalin Balog sums up the physicalist position as follows:

Any world which is a *minimal* physical duplicate of our world is a duplicate *simpliciter* of our world.⁹

In other words, 'Two worlds are physical duplicates if and only if they agree on all the true statements expressed in the language of physics.'¹⁰ However, as noted above in the discussion of Searle, the brute physical facts of the world do not determine its observer-relative features (i.e. social facts); more specifically, the laws of physics do not determine our notions of self and death (any more than they determine our notions of money or the rules of football). As a consequence, it may be said that physically identical worlds are not identical worlds in all senses, and that it may be possible for otherwise identical worlds to differ in terms of the makeup of their social facts.

A determinist might interject here that physical facts will necessarily determine how individuals behave, and therefore constrain the evolution of social facts along certain lines; so, it may be argued, our physically-determined brain states will eventually lead us to have particular, physically-determined social concepts. However, I would maintain that, even if this were so, it would be logically possible for the social facts to have evolved a different way; in other words, even if the world is determined, it is logically possible that it might have been determined another way (just as we might imagine a universe with different physical laws, or where different events unfolded). Therefore, in such a world, the social facts would be different. Furthermore, since I am proposing an increased possible scope of concept formation and change, we must assume that we are physically determined to appreciate choice! Perhaps this is crudely put, and perhaps if physicalism were true then some sort of compatibilist position would need to be adopted, but the main point is that we seem to be aware of choice, and also aware that the facts of the situation do not entail the form of the socially constructed concept. So, adopting a two-level, compatibilist position, we might yet say that on one level we appear to have choice (regardless of the actual state of affairs

⁹ Balog (1999), p.503 (italics in the original).

¹⁰ Ibid.

on another, more primary level). More simply, however, we might just point out that – given the findings of quantum physics and the possibility of *indeterminism* – physicalism is not necessarily the only option.

In arguing that there is an element of choice in our concepts, I would therefore argue that the formation and change of these concepts is independent of the truth of physicalism. This is not to say, therefore, that (for instance) notions of personal identity or death *supervene* on physical facts, for this would – it seems to me – suggest that the concepts involved were destined to take a certain definite shape. Rather, I would argue that these concepts spring from a sort of ‘debate’ between the significance of certain biological criteria, moral concerns, epistemic needs, social forces, and so on. In this sense, it can be seen that the latitude which I have attributed to the development and expansion of concepts of self and death in this chapter is untouched by the truth of the physicalist position.

The fact of death is therefore in a certain sense contingent: it is dependent on technological innovation, culturally-relative concepts of person and self, and even moral considerations. However, in adopting this view, I am saying more than that death is a changeable feature; rather, I am trying to point out that ‘person X weighs 13 stone’ and ‘person X is dead’ are different *types* of statement. The former is a statement about an intrinsic feature of the world which is subject to change; the latter is a statement implying values and concepts which are socially derived, and therefore open to change in a different way. The former is subject to physical laws; the latter – at least in part – subject to social conventions and concepts. A more exact parallel of the type of statement that ‘person x is dead’ would therefore be something like ‘person x is guilty’ (in a legal sense). In both situations, factual matters play a part (e.g. the intrinsic features of physical objects), and yet also in both instances there is the presence of what Searle calls ‘institutional facts’. In other words, for the social facts of moral and legal culpability to exist, there must first be established the social institution of morality, theories of agency and identity, and other related concepts. Similarly, then, decisions of death rely, at least in part, on the existence of an institution of social values to do with the value of life, the nature of identity, etc. I would now like to follow the consequences of this proposition for the definition of death in more detail.

An Observer-Relative Definition of Death

It would seem then that, if the above arguments are accepted, the definition of death and (for instance) the definition of a ‘goal’ in football have at least this in common: they are (are least partly) epistemically-objective statements about observer-relative and ontologically-subjective events. However, as I have already pointed out, according to Searle’s scheme, this does not mean that judgements about such things cannot be objective – in fact, he says, they can be either subjective or objective. This simply means that whilst the things we make judgements about may only have a subjective, observer-relative existence, the judgements we make can themselves be evaluative (subjective) or factual (objective).

For instance, Searle argues that the terms we use to describe objects have a factual character. If I call something a screwdriver, this assertion can be tested; it is a fact which can be ascertained as being true or false (either the object concerned is a

screwdriver or it isn't). However, to say that a screwdriver has an aesthetically pleasing design is to introduce a subjective valuation of it.

Death, Fact, and Value – Problems with Searle's View?

Searle's view of the physical world, as is probably apparent by now, is based upon the established distinction involving the categorization of judgements into 'fact' and 'value' in as much as he assumes the existence of features of a mind-independent reality which are non-evaluative (intrinsic) features. His utilisation of epistemically-subjective and -objective statements is therefore beholden to this established tradition of distinguishing between things which are objectively-true (and can be verified), and things which are merely expressions of value or preference. To be fair to Searle, however, he does admit that there is a continuum which exists between these two terms, and that they should merely be considered as representing two extremes of that continuum (and not a dichotomy).¹¹ Also, a further virtue of Searle's view is that – as I have just shown – objectivity does not begin and end with brute facts, and we may also make objective statements about things whose existence is contingent upon human value (i.e. are observer-relative). In this sense, he avoids the sort of primitive dichotomy that was prevalent with, for instance, the logical positivists.¹²

However, I want to argue that, in relation to the definition of death, a lack of recognition of the problems involved in this type of fact/value distinction has itself clouded the overall problem of how death is to be defined. This is because of two main assumptions:

- (1) It is possible to clearly distinguish between matters of fact and evaluative judgements.
- (2) The determination and definition of death involves purely intrinsic and brute facts.

In my arguments so far, I have therefore attempted to show such 'facts' – or empirical evidence – are not, in and of themselves, always conclusive. Furthermore, what makes them conclusive is not necessarily the presence of further facts, but rather clear and unambiguous guidelines or rules. These rules help us determine not only which facts are relevant, but also, what the facts 'mean'. It is these rules which allow the grouping of facts to be more than simply arbitrary collections of observed truths (I shall return to this shortly when I look at Searle's notion of the roles of constitutive and regulative rules in social practice).

To illustrate this point, we can think of any situation where the total facts of the situation leave the conclusion unspecified or open. In moral philosophy, therefore, a parallel situation would obviously be Hume's 'is-ought' gap,¹³ whereby no factual statements ('Jim robbed Bill') can entail a moral conclusion ('Jim ought not to have robbed Bill' or 'Stealing is wrong'). In a parallel way, therefore, the concept of

¹¹ Searle's belief in intrinsic features of the world would seem to be distinct from his moral realism (e.g. Searle, 1964). For my purposes, therefore, I treat Searle as a realist who denies the notion of intrinsic value or purpose (teleology).

¹² See, e.g., Ayer (1946).

¹³ See Hume (1967), Book III, Part I, Section i.

'death' can be seen to apply to factual statements concerning the function of aspects of a human being in much the same way as the concepts of 'right' and 'wrong' apply to the actions of human agents. Therefore, just as some moral philosophers might argue that certain actions are inherently right or wrong, or at least imply some moral obligation (moral realism),¹⁴ so certain philosophers in the definition of death debate have assumed that 'death' must be a feature of the world (metaphysical realism). However, as I hope I have shown, at least in relation to concepts such as 'death', this idea is flawed. Furthermore, the problem also extends to the more basic notions on which a concept of death might be built (namely, that of 'function'), and which it is therefore also difficult to consider intrinsic.

Some philosophers, such as Quine, have therefore argued that the separation of judgements into 'fact' and 'value' is an artificial exercise which ultimately ignores the necessarily entwined nature of the two concepts. To quote Quine's well-known illustration of this point:

The lore of our fathers is a fabric of sentences [...] It is a pale grey lore, black with fact and white with convention. But I have found no substantial reasons for concluding that there are any quite black threads in it, or any white ones.¹⁵

The 'lore' that we inherit – the concepts and possibilities of our language with which we describe and interpret the world – has neither purely 'black' (factual) nor 'white' (evaluative/conventional) aspects, but is rather 'grey' (an intermingling of both). Rather than a dichotomy, therefore, a better picture would be a continuum: some concepts are mostly 'black' (factual), with some 'white' (evaluative), and vice versa.

Another way of talking about this situation is to say that there is, as Hilary Putnam puts it, 'an entanglement of fact and value'.¹⁶ Putnam, in criticising the use of a metaphysical fact/value distinction made by some philosophers, argues that one key problem for this distinction is the existence of what are called 'thick ethical concepts'. Since one purpose of drawing any type of fact/value distinction is to distinguish between 'science' (factual statements) and 'ethics' (value judgements), the existence of terms which seem to span the divide would make any sharp distinction untenable. For instance, if I call someone 'generous', it may be argued that I am making *both* a statement of fact (the person's action is generous, having certain characteristics) and evaluating that person's behaviour (that action is 'good'). Therefore, the concept is 'thick' in the sense that it is not 'thin' enough to be merely classified as either a 'fact' or a 'value'.

A moral philosopher wishing to preserve some version of the fact/value distinction might argue (as Hume and Ayer¹⁷ have done) that such terms actually have no factual content (i.e. the non-cognitive position that they are merely conventional expressions of non-rational desires – e.g. emotivism and expressivism), or even that we may

¹⁴ Such as, in fact, Searle himself holds.

¹⁵ Quoted in Putnam (2002), p. 12. Incidentally, by considering no concepts to be purely factual in nature, Quine would seem to be in disagreement with Searle as to the existence of purely 'brute facts' (intrinsic features of the world).

¹⁶ Ibid., Chapter 2: "The Entanglement of Fact and Value", pp. 28-45.

¹⁷ Ayer (1946), chapter 6.

distinguish between the different aspects of 'thick' concepts (between, for instance, evaluative aspects and descriptive aspects of a concept – for instance, G. E. Moore's intuitionism, where 'good' is a non-natural property which we apply to factual situations).¹⁸ I do not have space to fully assess such positions here, but it should be clear that in proposing that death is an epistemically thick concept, I am ultimately criticising any strict application of the fact-value distinction. So, because, for the purposes of my argument, I have relied quite heavily on the coherence of Searle's distinctions, it might be natural to assume that any problem for Searle's position would also create a problem for me. However, for two reasons, I do not think this is so.

Firstly, I am quite content that I have proven that, given Searle's definition of 'intrinsic features', all attempts to consider such concepts as personal identity and function as intrinsic fail. Furthermore, any attempt to rescue these attempts would seem to rely upon a reintroduction of teleological or metaphysical features into a mind-independent reality, which would in turn represent a problematic and controversial move (and one which I have already rejected).¹⁹ I therefore consider this aspect of the application of Searle's ideas to be unproblematic. So, even if Searle is wrong, he is not wrong by being insufficiently metaphysically conservative.

Secondly, if we consider Searle's assertion that there are at least some intrinsic features to be problematic (because no 'facts' are ever value-free), then the criticism begins to run the other way. In other words, this is a criticism that might be levelled at him by antirealists or social constructivists,²⁰ and so any notion of death that would arise from such alternative perspectives would, by their very nature, *already* be 'observer relative'.²¹

Therefore, there is a sense in which – for my purposes – the antirealist attack on Searle's position is immaterial to the death debate: if Searle is right, then death is an observer-relative feature; if social constructivism is right, then death is *still* an observer-relative feature (or a socially-constructed equivalent thereof). The only problem for me would be a realist that wished to argue that some feature of mind-independent reality (such as biological function) provided a means to treat death as an intrinsic feature of that reality. However, since I have analysed these arguments (in chapter 4), and found them wanting, I consider my *use* of Searle to be justified and unproblematic (even if his position as a whole is not).

Death as an Epistemically-Thick Concept: the Partly-Biological View

In relation to the problem of the definition of death, I want to argue that 'epistemically thick' concepts do exist, and that death is one such example. Since this is not a term Searle uses, we might also translate this concept into his schema: certain observer-

¹⁸ Moore (1993).

¹⁹ See, for instance, the views of Wright (1973) and Milikan (1984, 1989) concerning the notion of intrinsic function, which I discussed in chapter 4, and also the notion of primitive natural kinds, as advocated by Wiggins (1980) and Doepke (1992) in chapter 3.

²⁰ See, e.g., Page (2004), and Rorty (1997).

²¹ Searle devotes Chapter 7 of *The Construction of Social Reality*, 'Does the Real World Exist? Part I: Attacks on Realism', to a discussion of the antirealist approach. For instance, a Nietzsche-type approach would focus on Searle's assumptions that cause and effect, mass, the notion of a distinct object, etc., were mind-independent.

relative concepts (such as in ethics) can have epistemically-subjective and epistemically-objective aspects. For instance, it is a question of epistemically-objective fact as to whether an individual has stolen a car; however, it is a question of epistemically-subjective fact as to whether this is right or wrong. The notions of 'right' and 'wrong' are therefore institutional facts which admit of a value-based application. However, in the following discussion, I will simply refer to 'epistemically-thick' concepts in what I take to be a roughly equivalent way to Searle's position.²² Death, therefore, is an epistemically-thick concept. However, in terming it thus, I would not necessarily wish to call death an epistemically-thick *ethical* concept, because that may be to presume that the evaluative aspect of the concept had a prescriptive purpose (or however we interpret the character of ethical statements). However, whilst I would certainly not rule this out, neither would I wish to say that this is the only role played by the evaluative aspect of the concept. I will therefore argue more generally that death is a thick *epistemic* concept, spanning both objective and subjective statements in the epistemic sense. So, in this sense, instead of the 'is-ought' gap, we may refer to what might be termed the 'is-what' gap. In other words, there are relatively independent factual aspects of situations involving the definition of death, and there are evaluative aspects: certain relatively uncontroversial 'facts' (*is*) and what they mean (*what*). Furthermore, we may say this, more subtly, about concepts much closer to the 'fact' end of the continuum than where I take death to be situated. For instance, it may be argued that science itself employs certain values in its search for 'objective truth'. I am thinking of such things as the *coherence* of a theory, its *simplicity*, its *stability*, etc. However, the concept of death has much stronger and more evident evaluative aspects than this.

It is in this sense, therefore, that we may conclude that any definition of death must adopt some version of a 'partly-biological' view. In effect, by calling death a 'thick epistemic concept', this is all we mean: there are facts to do with death (which are *relevant* to it) which take a primarily biological form;²³ however, these facts are not themselves sufficient for a definition of death, and death must therefore be considered an observer-relative concept which is relative to human values and culture. In a way, then, all observer-relative concepts are epistemically thick: they involve some intrinsic features, but the ontologically-subjective features are linked to human value and are therefore supplied by conscious human agents.

This can be seen if we consider, once more, the occurrence of certain types of so-called irreversible coma. An individual may retain the capacity for independent and spontaneous respiration whilst the capacity for all cognitive function has gone. Now, certain facts in this situation (construed in a certain way) are not really in conflict; however, certain *values* are. For instance, in saying that a person is alive because they

²² At the end of the day, it does not really matter to my purpose if Searle would not agree with my equivalence of this interpretation of his concepts with Putnam's concept of epistemic thickness. My main use of Searle has been in criticising the possibility of considering death as an intrinsic feature; should the opposite be true, therefore, and lead to attacks upon the nature of the intrinsic itself, then – as I have already stated – this is fine for the purposes of my argument (which will consider death as an observer-relative feature).

²³ Of course, if we were to consider a science-fiction scenario where there are no biological aspects to survival (e.g. survival of brain patterns on a computer hard drive), then we would have to alter 'partly-biological' to 'partly-physical'. However, the same principle would apply: there are observer-relative ascriptions which would mean that no 'strictly-physical' (and therefore possibly intrinsic) definition of death would be possible.

can breathe unaided, one of the things we are valuing here is the notion of *spontaneous function*. Since this is one of the signs of life, we are reluctant therefore to consider an organism which possesses this feature 'dead'. However, as human beings, we also value consciousness, or the capacity for *awareness* and *interaction*. In certain instances of coma, the absence of one, together with the presence of the other, therefore causes a conflict (and subsequently the type of ambiguity which I outlined initially). As I have already suggested, the resolution of such conflicts does not necessarily come via the discovery of further facts, but rather through addressing the 'rules' which give such facts their significance. Now, since these rules – as I shall shortly argue through the parallel with other social practices (i.e. football) – are not factually derived (or at least, not solely so), but stem instead from a complex of valuations and conventions, the similarity between the definition of a goal and the definition of death is therefore greater than many would first suppose. However, if this is admitted, this leaves us with a problem: sporting rules may be altered arbitrarily, but can we treat death in such a way?

Social Reality: Constitutive and Regulative Rules, and Conventions

I would like now to embark on a treatment of death as an observer-relative feature of the world. To do this, I will once again use Searle's form of realism as a basis for discussion.²⁴ In *The Construction of Social Reality*, Searle argues that social reality exists because, through collective human intentionality, we create the possibility of social (or *institutional*) facts. Therefore, for such a thing as money to exist, there must exist a collective intention that certain-shaped and –coloured bits of metal, pieces of paper, and so on, should have a certain social significance. Without this, none of these things *in themselves* would mean anything at all (they would have no *intrinsic value*).

The second main ingredient of Searle's social ontology is the identification of rules which govern these practices.

Firstly, a social practice must have constitutive rules; these consist of rules which make the practice the thing it is. So, in terms of football, for instance, the constitutive rules would be such things as: no outfield players can handle the ball during play; kicking another player to obtain the ball is a foul; if a ball crosses a team's goal-line, then a goal is scored to the other team; and numerous other rules. Without these rules, it would not be football: if players pick up and run with the ball (as they do in rugby), or assault other players to get the ball (as some might say they also do in rugby...), or score points if the ball flies over the crossbar of the goal, then it would not really be a game of football.

Secondly, practices are also accompanied by regulative rules; these consist of rules which 'regulate antecedently existing activities.'

²⁴ See Searle (1995), pp.27-29 for a brief overview of his basic idea; also, Page (2004), pp.264-6, for a succinct account. Many have criticised Searle's distinction between constitutive and regulative rules – as he admits himself – Searle (1995), p.29, for instance: Giddens (1986), Ruben (1997). However, whilst I am not myself completely happy with the distinction, the criticism is not essentially significant for my purposes, because I am primarily interested – as previously – in having a realist framework to discuss these ideas. Whether such a view turns out to be problematic, therefore, is incidental, for if I can express my view that death is socially constructed and problematic *within* such a perspective, then it goes without saying that this will also be the case for certain antirealist/social constructivist positions.

For example, the rule “drive on the right-hand side of the road” regulates driving; but driving can exist prior to the existence of that rule.²⁵

So, if we break a regulative rule, then we can still be said to be taking part in the practice which the rule regulates: to drive on the wrong side of the road is still to be driving, though it may be punishable by law; to occasionally handle the ball or foul another player is still to play football, though such things are punishable by specific disciplinary sanctions (such as a yellow or red card). Regulative rules, then, do not exist to create the practice, but regulate it according to other principles: in the case of driving, regulative rules exist for the sake of safety (to protect pedestrians and other drivers); in football, regulative rules exist to ensure fair play, or to safeguard against injury.

Finally, Searle distinguishes between constitutive and regulative rules, and conventions: it is a convention that drivers acknowledge the courteous actions of other drivers (such as letting another car join a line of traffic in front of you) with a wave of the hand, but it is not a constitutive or regulative rule. Similarly, in football, teams will conventionally put the ball out of play when a player of the opposing team is injured; however, there is no rule of either sort that commands them to do this.

The Game of Death

Given these distinctions, it is interesting to apply these concepts to the determination of death. Firstly, since death is an observer-relative feature, then it must be considered an institutional fact which exists because of the collective intentionality of human beings. So, death exists because human beings credit certain brute facts of the world with certain significance: so, those who do not possess a range of specific physiological features (respiration, heartbeat, capacity for conscious awareness, and so on), are considered dead, whilst those who *do* possess them, may be considered alive.

Translating talk of death into Searle’s terminology for the explanation of how social facts come to exist, we find that – much like Wittgenstein’s explanation of linguistic practices²⁶ – we are tempted to talk of social practices as if they are a ‘game’. So, what are the ‘rules’ in the game of determining death? It seems very strange to talk in this way, but as this is arguably a consequence of moving away from the traditional idea of death as an intrinsic feature of the world, and towards a more controversial observer-relative view of death, then perhaps it is inevitable that it should feel odd.

Firstly, what are the constitutive rules of death? Most important, it would seem, would be the idea that someone is either dead or alive – in other words, the notion of binary opposition. Another rule might be the notion of irreversibility: if someone is dead, and then comes back to life, then he or she was not really dead in the first place. Other constitutive rules might involve the possession or absence of certain significant features: a person who lacks all spontaneous function, but is conscious, should not be

²⁵ Searle (1995), p.27.

²⁶ See, for instance, Wittgenstein (1967), p.5: ‘I shall also call the whole, consisting of language and the actions into which it is woven, the “language game”.’ This approach to language seems to form the main theme of sections 1-64.

considered dead; on the other hand, the presence of spontaneous function in non-essential features (such as the continued nervous and muscular activity in a headless chicken or certain other animals), is not considered sufficient for the individual to be considered still alive.

In relation to the above, regulative rules would seek to refine and legislate activities within the 'game'; these mostly reflect the growth of medical knowledge through history, examples of which would be: a person is not dead merely because his heart stops, or his breathing has ceased (such cases may be revived); the failure of an organ is not necessarily fatal (it may be replaced or artificially maintained); certain states may imitate death (such as hypothermia, cataleptic narcolepsy, 'locked-in' syndrome, and so on), and carefully specific tests are required to differentiate between genuine and 'faux' death.

Therefore, in the problems concerning the definition of death, we can see that the conflicts concerned involve the application of problematic or ambiguous constitutive rules. For instance, prior to the advent of technology that would keep a comatose person alive, the criteria concerning spontaneous function and capacity for consciousness had not been brought into conflict. However, subsequently, these constitutive rules can be seen, in certain controversial cases, to provide alternative and conflicting outcomes. To use an analogy, it is as if, in a game of football, on trying to decide whether the ball had crossed the goal-line (i.e. a goal had been scored), it were suddenly realised that *no one knew what that meant*.

A Sporting Analogy: Football and Death

Sport is full of controversy. Frequently, this controversy surrounds matters of fact. In such cases, there may be debate – for instance – as to whether a ball has crossed a particular line. In football, where that line is the goal-line, there is no doubt when either the ball hits the back of the net, or is cleared before it gets anywhere near the goal (these are simple, binary cases of 'goal' and 'no goal', where the distance between the ball and the line is readily apparent). However, in the famous case of the 1966 World Cup final, England's third goal was a disputed case: the ball hit the bar, bounced down near the goal line, and was cleared. Did it cross the line, and was it a goal? The rule states that, 'the whole of the ball must cross the whole of the line' for it to count as a goal. Since then, television replays and computer projections of the event suggest that it did not; however, this illustrates my point: any difficulty there is in deciding whether the ball crossed the line or not is purely an empirical one; in other words, there is no ambiguity in the definitions involved, and the only problem involves clarifying the actual issue of where the ball was in relation to the line.

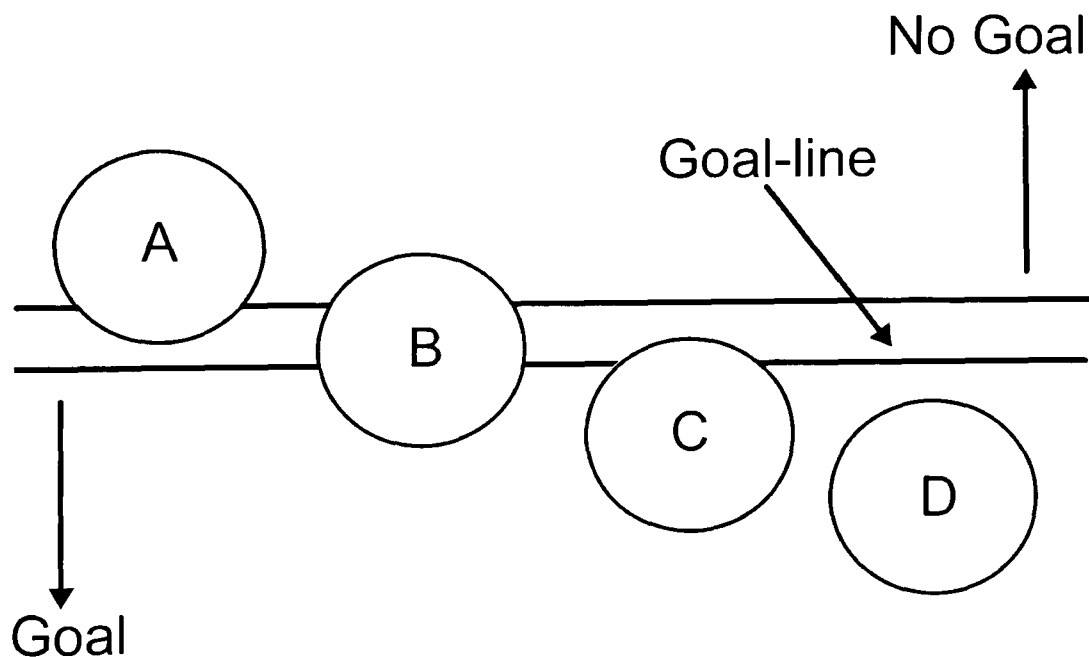
However, whilst I do not know the exact details, there must have been a time in the development of the game of football where controversy regarding marginal cases had to be directly addressed. It is therefore easy to imagine that there was a time when the need for greater precision regarding goal-line decisions became apparent. Now, if we are to imagine the sort of controversy that took place in the 1966 World Cup final taking place at a time when such lack of precision had not yet been addressed, then it will be clearly seen what I mean by definitional ambiguity.

So: Geoff Hurst's shot hits the bar and bounces down near the line, before being cleared away; the English team claims a goal, and the Germans protest. On one level, we are presented with the same empirical problems as there are after the rule change: did the ball cross the line? However, on closer inspection, because we are imagining the event taking place at a time when certain laws lacked a certain precision, we are also presented with a definitional ambiguity. Actually, the first official rules established in 1863 state that:

A goal shall be won when the ball passes between the goal posts or over the space between the goal-posts (at whatever height), not being thrown, knocked on, or carried.²⁷

Now, here the ambiguity consists in a number of factors: there is no specific reference to the goal line, but let us suppose that passing 'over the space between the goal posts' means 'crossing over the line'; but what would 'crossing the line' consist of here? For instance, is it (see associated diagram):

- Case A: For the front part of the ball to cross part of the line?
- Case B: For the front part of the ball to cross the whole of the line?
- Case C: For the whole of the ball to cross most of the line?
- Case D: For the whole of the ball to cross the whole of the line?



The important point here is that, since the law concerned is vague or ambiguous, any of the above interpretations may be valid (though Case D is obviously a clear goal). Furthermore, as a consequence, no empirical evidence will resolve the issue whilst such an ambiguity is present. What does it matter that we can conclusively obtain the exact position of the ball (i.e. through television footage or computer projections), when the law does not specify (unlike the modern one) how far the ball must be over

²⁷ See Ken Aston's excellent history of football – Aston (2008).

the line?²⁸ Whilst there is a possible ambiguity in the definition of ‘goal’ involving the terms ‘ball’ and ‘goal-line’, the problem is not (or not only) an empirical one, but a definitional one.²⁹

The resolution of such definitional problems must therefore involve, fundamentally, a clarification of the definitions used. In the definition of a goal in football, this has now been achieved by the somewhat arbitrary amendment of the law (or, as Searle would say, the ‘regulative rules’ of the game). In other words, the definitions of ‘in’ and ‘out’ of play, the interaction of players, the ball, lines, etc, could all have turned out differently – as they have in different sports (in rugby, for instance, if the ball touches a line, then the ball is considered ‘over’; this is almost the opposite to the case in football). Furthermore, and most importantly, there is no sense in which such an arbitrary decision can be incorrect. If a goal is defined in a particular way, but the regulative rule is subsequently changed, the previous rule was not incorrect; rather, it was decided that the interests of the game would be better served under the new rule. However, in suggesting that the act of changing such sporting rules can be somewhat arbitrary, this is not to say that a number of different sorts of factor could not play a part in such a decision: for instance, in deciding how a goal should be defined, it may have been a deciding factor that it is easier to tell whether a ball has crossed the line completely than it is to tell if part of the ball has crossed part of the line. Consequently, in saying that occasionally certain rules in sport may involve definitional ambiguity, I am not saying that no empirical or other factors are relevant to its resolution, but merely that such factors do not, of themselves, necessarily provide a resolution.

Conclusion

These points, whilst relatively uncontroversial in relation to the definition of regulative rules within football, can be seen to be far from so in relation to the definition of death. Firstly, it should be noted that the analogy is not exactly parallel: the conflict in the laws of football can be seen as involving *regulative* rules; the conflict in the definition of death involves *constitutive* rules. So, the ambiguity concerning the definition of death is *deeper*; it concerns what death *is*. Ultimately, I will argue, on one level this makes no difference in as much as either set of rules can be changed; however, what this observation does explain is the greater difficulties that we face in the latter case. Constitutive rules are more fundamental to the practice concerned, because they define what that practice *is*; however, regulative rules are more changeable because they respond to surface-level problems, and therefore changes to these rules do not amend the *nature* of the practice, but merely refine its expression.

Furthermore, however, the key difference between the ‘games’ of death and football is that the former represents a more deeply embedded social practice than football, and thus its rules are more difficult to change (because of the greater importance

²⁸ Actually, there may be some ambiguity even with the modern law. For instance, the rule does not specify from what perspective the ball should appear to have crossed the line – I am assuming a bird’s eye view (i.e. directly above), but this is not explicit. In fact, this is the very sort of definitional ambiguity that sometimes results in the sort of divergent applications and controversy that I wish to illustrate.

²⁹ In fact, this ambiguity would be a direct parallel of my discussion of Quine in chapter 3 concerning the inscrutability of reference – for ‘rabbit part’ read ‘ball part’!

placed upon its outcomes).³⁰ However, I shall argue that the difference here – whilst great – is not one of kind, but one of *degree*. This is an important point, I think, because it underlines the central conclusion of my thesis: death is not an intrinsic feature of the world, and in abandoning that possibility, we also abandon the possibility of treating it as a *privileged* social practice. That it is more important than football, despite Bill Shankly's famous quip,³¹ is therefore due to the greater existential consequences attached to it, and the more fundamental values that it concerns, and not to a difference in the nature of the practice itself: football and death are therefore points on a continuum, where the more embedded the practice in the social fabric, the more intransigent its constitutive rules seem. However, this intransigence is only relative to human ingenuity and desire, and there is no ultimate reason why very different notions of death and personal identity might not exist.

This topic will be the subject of my next chapter.

³⁰ I speak, of course, to those people who live outside of Liverpool.

³¹ The onetime manager of Liverpool FC is reported to have said: 'Some people think football is a matter of life and death ... I can assure you it is much more serious than that.' Metcalf (1986), p.237.

8. Death and Immortality

Introduction

In this chapter I will consider certain applications of my arguments concerning the concept of death. So far, I have been at pains to point out what I think are compelling reasons for rejecting the idea that death is in some way an intrinsic feature of the world. I have also tried to show that, whilst death is an observer-relative feature (and we must therefore adopt a 'partly-biological' view), certain criteria relating to the traditional concept of death have been brought into conflict with one another by advances in medical technology. Therefore, whilst different conceptions of death are possible (in the same way that we may adopt different constitutive rules in changing the nature of a game), there is a general, understandable reluctance to do so.

Therefore, in the following sections, I first consider the question of what limits there are to the adoption of new conceptions of death. After this, I set out a range of views on death from differing perspectives in science-fiction and religion – which I take to be the two areas which provide the most far-reaching and contentious alternatives to the traditional conception of death.

Before commencing, however, I should perhaps just point out that the various topics in this chapter are touched upon quite lightly – and, some might argue, too briefly. However, I don't have space to consider the many possible responses to the positions advanced, and I would not want it to be thought that I considered the arguments set down as conclusive treatments of the issues concerned; rather, I would ask that they be considered possible avenues down which the directions opened up by the conclusions of the previous chapters may be pursued.

Death and Choice: Morality

In the search for ontologically-objective criteria for the disambiguation of controversial cases in the determination of death, proponents of various positions have rejected the relevance of moral criteria to the debate (e.g. as Becker, Lamb, and Russell have argued, for instance). However, as this rejection was linked to an attempt to treat death as an intrinsic feature of the world (or as based on intrinsic features), it seems – now that I have concluded that death is observer-relative – that this issue must be revisited. What part, then, do moral concerns play in our adoption of an *observer-relative* definition of death?

There have already been attempts to resolve this problem from a moral perspective. Hans Jonas has argued that, whilst the brain-dead¹ should *not* be declared dead, we have no duty to prolong the life of such individuals, because the inherent value of the patient's life to the patient himself is practically non-existent.² Howard Brody, on the

¹ By 'brain-dead' and 'brain-death' here and in the following paragraphs, I mean *whole-brain* death, referring to the essential functions of the brain-stem *in addition to* other areas of the brain (including the neocortex).

² In this sense, he argues, 'the Report [of the Ad Hoc Committee of the Harvard Medical School to Examine the Definition of Brain Death] has strictly speaking defined not death, the ultimate state, itself,

other hand, whilst he rejects the neocortical definition of death, argues that brain-death is synonymous with death itself; this is because the person without capacity for consciousness is no longer a person (and therefore not capable of bearing moral rights), though they might nevertheless still retain spontaneous function (and are therefore, according to Brody, still alive).³ Robert Veatch takes a minimalist approach, arguing that the essential feature of a human being is the embodied capacity for experience and social interaction; since consciousness is a prerequisite for this, neocortical death is synonymous with the death of the human being.⁴ Paul Ramsey⁵ and Robert Schwager⁶ both consider spontaneous respiration and heartbeat as a sufficient condition for life, and our moral obligations therefore spring from our inherent duty to the still-living human being. John C. Hoffman⁷ defends Veatch, but argues that the question of what death is can only be answered by identifying which features hold an intrinsic moral value, and therefore entail moral obligations (deciding, along with Veatch, that the capacity for experience and social interaction is the defining feature). Finally, Karen Gervais⁸ argues that human death is the death of the person (which either neocortical or whole-brain death are sufficient to bring about), but that choosing to instate this definition is a public policy decision that requires a contractarian approach.⁹

However, all of the above approaches rely to some extent on essentialist assertions: Gervais, Veatch, and Hoffman argue that consciousness is essential to personhood, which is in turn the essence of a human being; Jonas takes a rights-based approach, considering the essence of a human being as the capacity to be a moral agent; and Ramsey and Schwager view spontaneous function as a defining feature of being alive. However, to take Hoffman's point, each of the competing views of what essentially constitutes a human being entails an alternative assertion of 'intrinsic value'.¹⁰ But can what has arguably been traditionally thought of as an ontological concern be resolved by purely evaluative considerations?

This is a tricky question, I feel, and one that threatens to blur the distinction between ontological and moral questions. For instance, returning to Searle's terminology, moral concerns may force change upon the constitutive and regulative rules of many social practices: this may range from the alteration of somewhat arbitrary customs, such as the tightening of sporting rules concerning physical contact in rugby union so as to protect players; to deeper changes, such as the outlawing of the Hindu practice of *suttee*, where widows followed their deceased husbands onto the funeral pyre. In both these instances, conflict between values leads to a revision of the practices in what we like to think of as 'moral progress' (rugby becomes more considerate of

but a criterion for permitting it to take place unopposed – e.g. by turning off the respirator.' Jonas (1978), pp.51-2.

³ Brody (1981). As Gervais points out (1986, p.88), this seems to commit Brody to the problematic position of considering neocortically dead individuals as living non-persons.

⁴ Veatch (1975).

⁵ Ramsey (1970).

⁶ Schwager (1978).

⁷ Hoffman (1979).

⁸ Gervais (1986).

⁹ *Ibid.*, pp.166-77. Gervais is not a moralist *per se*, but her conclusions suggest that, the ontological considerations once established, death should be defined in the process of a public policy debate (which, by definition, will take on a primarily moral character).

¹⁰ Hoffman (1979), p.436.

safety; funeral practices become more ‘enlightened’ and humane); in such cases, the deeper, morally-driven constitutive rules can be seen to outweigh the dictates of tradition and social expectation.

However, if we consider the moral approach to be relevant to the definition of death, we must admit that it is only one possible influence in a range of other factors; many social practices are also shaped by aesthetic principles, evolutionary drives, epistemic concerns, religious and spiritual aspirations, and so on. As a result, we cannot privilege moral discourse over all these other concerns – unless, of course, we want to say that a concern for morality is the fundamental and defining feature of humanity (which, apart from being a highly debateable assumption, is difficult to see as anything other than a bald metaphysical assertion). Therefore whilst, in the above examples I considered moral considerations to be ‘deeper’ than other concerns (e.g. the desire for sporting spectacle in rugby), this is not to say that it is the *deepest*, or that it is somehow a reliable and independent arbiter of what should and should not be the case as regards the rules of social practice.

What we therefore now require is a picture of how social practices interrelate, before considering how they may change.

Wittgenstein’s River

Various philosophical approaches argue that we cannot completely distinguish between ontology and epistemology; in other words, between what there is, and how we come to know it. So, as a consequence, we may argue that our ontological concepts are always ‘tinged’ in some way by limits that we, as humans, are forced to accept. The nature of these limits, of course, is open to debate: a Kantian explanation would see such limits as built into the possibilities of experience; a Nietzschean approach, on the other hand, would point out that our rational understanding of the world is motivated by irrational forces that we are seldom (and can never be wholly) aware of; Searle has argued that many features of our social practices are observer-relative and only make sense in relation to social institutions and collective intentionality. However, for whatever reason, it may be argued that the concepts that we utilize in our social practices (and even, some argue, in our fundamental epistemological and scientific ones) are ultimately without absolute rational grounding.¹¹

However, this is not to say that, as a result, our practices and concepts become arbitrary. Searle has already given us one picture of how social practices may be embedded by human collective intentionality, and Wittgenstein has similarly argued that human practices reach down into a non-rational basis or ‘form of life’ within

¹¹ I take this to be a relatively uncontroversial position (though obviously there will be some who oppose it). I have highlighted this aspect of the philosophies of Wittgenstein (1967, 1974), Quine (1969a), and Nietzsche (1990a), but there are many other examples I could choose. However, perhaps the seminal and definitive statement on this subject is Kurt Gödel’s ‘incompleteness theorem’, which denies (generally) that ‘all truths or laws pertaining to their subjects are provable within them.’ (Detlefsen, 2000, p.319). Obviously, this has wide-reaching consequences, but most significantly suggests the failure of attempts to prove the ‘completeness’ of first-order logic (which was the goal of A. N. Whitehead and Bertrand Russell’s *Principia Mathematica*). In the current context, it simply means that such concepts as that of ‘self’ and ‘death’ cannot (arguably) be defined without the assumption of some arbitrary, conventional, or non-rational element.

which such practices make sense. For this reason, therefore, even though we may lack a fixed set of ontologically-independent facts which determine our social practices (atoms do not tell us how death should be determined, because both living and dead things are made up of them), we may nonetheless still picture a way in which such practices are not *completely* arbitrary; if there is logic and consistency within games, and some games are 'deeper' than others, then this interplay is perfectly capable of explaining social interaction – including the practices surrounding death. This therefore provides us with a conceptualist framework which – to an extent – constrains what is and is not conceivable (though to what extent this is *mutable*, I shall shortly explore).

I take this picture to be similar to the conclusion that Wittgenstein arrives at in his later philosophy: the meaning of certain propositions is relative to other, 'deeper' propositions; so, whilst we may not ultimately be able to find a rational ground for our practices, there are limits to the choices that we can make as to what exactly those 'fundamental' propositions are.

95. The propositions describing this world-picture might be part of a kind of mythology. And their role is like that of rules of a game; and the game can be learned purely practically, without learning any explicit rules.

96. It might be imagined that some propositions, of the form of empirical propositions, were hardened and functioned as channels for such empirical propositions as were not hardened but fluid; and that this relation altered with time, in that fluid propositions hardened, and hard ones became fluid.

97. The mythology may change back into a state of flux, the river-bed of thoughts may shift. But I distinguish between the movement of the waters on the river-bed and the shift of the bed itself; though there is not a sharp division of the one from the other.¹²

As can be seen, the metaphor of a river is a good one: the fast flowing current represents the quickly changing world of facts, and our knowledge of things *within* certain contexts; just as we might seek to ascertain the state of play in a game. However, as we move deeper, the propositions become more fundamental and less changeable – until, Wittgenstein argues, we reach a point where certain empirical propositions become 'hardened' and function as 'channels' for other, shallower, 'fluid propositions'.

Here, in fact, we have a parallel to Quine's notions of ontological relativity and relative conceptual schemes:¹³ the 'hardened' propositions are not absolute, and their relation with more fluid propositions is 'altered with time'; however, there is always a sense in which the deeper propositions are necessary for the shallower ones to have a reference point or 'anchor'.

¹² Wittgenstein (1974), p.15.

¹³ Quine (1969a).

If we relate this picture to social practices, and especially to death practices, we can see that advances in medical technology have resulted in a ‘shift of the [river] bed itself’; assertions which were considered fundamental and unalterable (such as the binary definition of death, the co-incidence of cessation of spontaneous function with death itself) have loosened from the bed-rock and have begun to float more freely (although, to continue the metaphor, still at a deep, slow-moving level – that is, the features of what we consider essential or important to a human being are still deeply resistant to change).

However, if we accept this picture of deep and shallow structures in relation to social, linguistic, and epistemic practices, then we are left with the question: How much is it possible to revise our fundamental conceptions of life, death, and self? Obviously, this is not a question that can be easily answered, but a further problem lies in the assumption that we can set limits: if fundamental concepts are potentially open to change, then how can we possibly say what it is and is not possible for them to evolve into? This highlights the dynamic nature of this particular view of human discourse, and shows that the desire to set limits to it – to say (using Wittgenstein’s metaphor) that *this* part of the river bed is fixed and unchanging – is to use concepts themselves which are part of this process, and therefore *themselves* potentially open to change. In other words, any metaphysical position is merely one among others, and therefore open to change.¹⁴

I will now go on to consider how this type of perspective (where no elements of social discourse are ultimately fixed and unchangeable) may allow us to understand historical developments in the notions of life, death, and self. In doing so, I will look at various conceptions of the self – as body, as soul, as mind – and attempt to show how it is possible to find (or imagine) variations of these concepts which stretch our notion of what death is.

Death as the Departure of the Soul

Before going on to talk about the concept of the soul in relation to our notions of the self and death, I would like firstly to make some observations as to its possible significance for determining and defining death.

It might be thought that the debate concerning the definition of death would be immeasurably simplified – if not resolved – if we could assume the existence of the soul. Of course, there would still be the attendant problem of ascertaining exactly *when* the soul leaves the body (or whether it is still present), but we could at least entertain the following:

- (1) that the soul is a human being’s vehicle of consciousness;
- (2) that it is separate from the body;
- (3) that it survives and possibly pre-exists the body;
- (4) that it is the essence of what we consider to be the human individual.

¹⁴ Which, of course, is precisely the point made by Gödel.

Obviously, even if we accepted these statements as true, we would still be faced with problems: does the fact that the body is still, to some extent, 'alive' mean that the soul has not yet left, or that those features which persist are non-essential to the soul? If the soul has left, can it be brought back (as in the Biblical case of Lazarus)? Does the ongoing presence of the soul entail a duty to maintain life, or might we argue that, under certain circumstances, it might be right to 'release' the soul from its earthly prison (where, perhaps, it can never again lead a full earthly life)?

As these questions show, the problem of *determining* death does not go away if we assume the existence of the soul, but (it may be argued) the problem of *defining* it does; the problem of when someone is dead can be seen to be an empirical one, but the fact of death would become (dare I say it) an *intrinsic* one. In this sense, problems of whether to 'release' souls or not could become a legitimate moral (or theological) concern. However, the above said, there are good reasons why this approach – even if we could prove the soul's existence – is no more fruitful than any other.

Gareth B. Matthews¹⁵ argues for a revival of the idea that death takes place when the soul leaves the body. In association with this, he argues a return to the notion of 'quickening' (as a sign of when the soul enters the body), and, in relation to death, its opposite (which Matthews does not give a term, but which we might call 'hastening'). For Matthews, the determining sign of when quickening takes place in the foetus relates to when it can be decided that it possesses psychological qualities; so, when it responds to stimuli with purposive action, or shows signs of expressing preference. However, the one example he gives – that of a foetus which responded to an injection of a sweetened substance (saccharine) into the amniotic fluid by 'swallowing larger than usual amounts'¹⁶ of liquid – could also, arguably, be interpreted as a bodily reflex, and not entail consciousness at all. In fact, Matthews seems to have a very skeletal idea of what is entailed by the term 'psychological features', and the idea that there might be a problem in distinguishing between unconscious and conscious processes (or psychological and phenomenal consciousness) seems not to have occurred to him at all.

Another problem is why we should associate quickening with the arrival of the soul. Traditionally, as with Descartes, the soul has been considered a purely human entity not granted to animals (as Matthews himself points out).¹⁷ However, the sort of psychological quickening that Matthews identifies is not, surely, that different from that which is displayed by animals – in fact, as Matthews argues, quickening can be thought of as the point at which the foetus first becomes 'a little animal'.¹⁸ It is debateable, therefore, whether such 'psychological' qualities are in fact the defining features that Matthews takes them to be: if an animal foetus (e.g. a lamb) were to display such responses, it would seem strange to decide that we cannot abort the human foetus, but that we can later kill the lamb *even when it displays signs of greater development*. Furthermore, why should we therefore consider quickening to be an event that is specifically related to the arrival/genesis of a metaphysical entity (the soul)? It would seem that either Matthews must admit that animals have souls too (which I doubt he wants to), or that he has in mind some looser notion of 'soul' which

¹⁵ Matthews (1979).

¹⁶ Ibid., p.156.

¹⁷ Ibid., p.153, n.8.

¹⁸ Ibid., p.155.

is meant to serve for some essentialist notion of what defines the human being. Either way, there seems to be a lack of clarity and precision in his argument, and it is left ambiguous as to what exactly is intended by this notion of the soul.

Turning to the other extremity of the continuum, Matthews does not give much space to the notion of 'hastening'. However, the same problems as just outlined in relation to quickening also apply here: there is no distinction between psychological features and phenomenal consciousness, and Matthews seems to have no awareness of the difference between these aspects or how it might have a bearing on his definitions. Therefore, in both cases, we are left with a bald essentialist assertion that the human being consists in certain psychological features (which are also shared with animals), and that their presence or absence constitutes life or death respectively. However, in terms of determining death, this leaves us no closer to a clear procedure than previous essentialist attempts – and, what is more, with the added metaphysical baggage of a non-corporeal entity (that is, if this is the way that Matthews intends his concept of soul to be understood – which, as I say, is unclear). Lastly, we might also consider that, if it is possible that a soul may be returned or 'called back' to the body, then death may become 'reversible' – we might even call back souls to other bodies.¹⁹

As I suggested initially, therefore, the problems that we face in determining death are not avoided by the introduction of the idea of a soul. *If* we could ascertain the definite existence of the soul, and which functions it is responsible for maintaining, then perhaps a less problematic procedure for determining death could be arrived at. However, since medical science seems either to ignore or dismiss the relevance of the concept of the soul, and it has not been conclusively shown that such an entity does in fact exist, then its relevance for the determination of death (if it has one) seems to lie somewhere in the future of metaphysical/scientific research.

However, if the notion of the soul does not help us in *determining* death, what about the significance of the concept for the *definition* of death? We might point out that, even though it is difficult at the moment for us to decide the moment of death via reference to the departure of the soul, this does not rule out the possibility that, in the future, its existence and nature might be established. However, it might be argued, whilst it cannot currently help us resolve all problematic cases in the determination of death debate, its existence would at least provide us with a clear sense of what death *is* (i.e. the departure of the soul).

But would it? Having argued that the notion of the soul does not (as least, currently) help us in determining death, I would now like to set down what I think are good reasons why it cannot ultimately help us to establish what death *is*. I will also later discuss what I think are the reasons for this, and why, in my view, religion represents the sort of 'debate with death' that can also be found in science-fiction contexts.

The Deconstruction of the Soul²⁰

¹⁹ There are many treatments of this theme in film and literature. My favourite is the film *All of Me* (Reiner, 1984), where Steve Martin portrays a lawyer who ends up sharing his body with the soul of a recently deceased female client (Lily Tomlin). The client had planned to transfer her soul to the body of a young woman (whose soul would then 'pass on'), but the transfer goes wrong.

²⁰ Incidentally, it should not be thought that my use of 'deconstruction' here or subsequently has any intended relation to the post-modernist position associated with Jaques Derrida, etc. As should now be

Descartes and Plato both seemed to hold a similar conception of the soul. Plato, in *The Republic*²¹ and *The Phaedo*,²² argued that not only is the soul the seat of understanding and reason, but that it is also the essence of the individual. Similarly, Descartes argued that man's essential activity is thinking, and that the soul is the vehicle of this process; that 'I' is the subject of the predicate 'think'.²³

This view has been extremely influential, and it may be debated whether in fact Plato and Descartes have merely given voice to a prevalent 'folk belief' in the identity of the individual. For instance, even in Berger and Luckman's famous work on social construction, *The Social Construction of Reality*, where one might expect to find the notion of the individual socially 'deconstructed', the authors seem to retain an almost Cartesian concept of the individual. On the one hand, they argue that

Identity is formed by social processes. Once crystallized, it is maintained, modified, or even reshaped by social relations. The social processes involved in both the formation and the maintenance of identity are determined by the social structure. Conversely, the identities produced by the interplay of organism, individual consciousness and social structure react upon the given social structure, maintaining it, modifying it, or even reshaping it.²⁴

Whilst I agree with the broad sweep of the above, it should be pointed out that Berger and Luckman already seem to utilize a 'ready-made' concept of the individual. As they state shortly after:

Identity is a phenomenon that emerges from the dialectic between individual and society.²⁵

And the concept of the individual itself? How does that emerge? There may be some sense in which the authors consider the notion of an individual self with set limits as being defined by biology: 'the organism posits limits to what is socially possible.'²⁶ However, as I hope I have already shown, biological criteria do not always provide incontrovertible criteria for identity or self. In the following sections, therefore, I will consider various different ways in which this Cartesian/Platonic notion of the individual can be broken down and replaced with other, broader concepts.

apparent, my approach stems mainly from the Anglo-American analytic tradition, though is not strictly confined to that.

²¹ Plato (1987), Part XI: 'The Immortality of the Soul and the Rewards of Goodness', pp.378-93.

²² Plato (1969), pp.99-183.

²³ See, for instance, Descartes (1968), Second Meditation, pp.102-12.

²⁴ Berger and Luckman (1966), p.194. This is not to say, of course, that social construction has *not* considered the notion of individuality to be socially constructed – it has become a standard approach. See, for instance, Anderson (1997). Obviously, what I am proposing here is a variation of this type of approach for the related concepts of 'self' and 'death'.

²⁵ *Ibid.*, p.195.

²⁶ *Ibid.*, p.201. However, they do then go on to talk about these biological factors limiting 'the range of social possibilities open to any individual' (pp.201-2), again suggesting a pre-formed notion of individual.

Returning to the religious perspective, whilst it is easy to find traditional religious doctrine which backs up this Cartesian/Platonic view, it is also possible to find divergent ones. The most obvious example of a more radical concept of the self is to be found in Buddhism. In the doctrine of *anatta* ('no-self'), Buddha argues that the idea of the self is a mistake, and that such a 'misconception was a symptom of that ignorance which kept us bound to the cycle of suffering.'²⁷ As Karen Armstrong explains:

Where the seventeenth-century French philosopher René Descartes would declare 'I think, therefore I am,' the Buddha came to the opposite conclusion. The more he thought, in the mindful, yogic way he had developed, the clearer it seemed that what we call the 'self' is a delusion. In his view, the more closely we examine ourselves, the harder it becomes to find anything that we can pinpoint as a fixed entity. The human personality was not a static being to which things happened. Put under the microscope of yogic analysis, each person was a process.²⁸

So, in the metaphors Buddha used to illustrate this view, the self is compared to 'a blazing fire or a rushing stream';²⁹ and, he argued, just as fire or water is different from moment to moment, so is the self. For Buddhism, therefore, *personalism*,³⁰ or the idea that Buddha's teaching may be interpreted as leaving room for the concept of a distinct self, is generally considered a misinterpretation of his doctrine, and a metaphysical mistake.

This brings to mind, in different ways, Gilbert Ryle's concept of the Cartesian self as a 'category mistake'³¹ and David Hume's picture of self as a 'bundle or collection of different perceptions'.³² However, it also suggests Nietzsche in the way that 'soul atomism',³³ as he calls it, is seen as a doctrine which springs from a psychological need to attribute a centre to things – and specifically, to the self. However, for Nietzsche, this is simply another example of a philosophical prejudice, born of Platonist Christianity's need to guarantee the personal identity, moral culpability, etc., of individual agents. However, in an almost Buddhist-like passage, Nietzsche takes apart Descartes's famous *cogito* argument:

a thought comes when 'it' wants, not when 'I' want; so that it is a *falsification* of the facts to say: the subject 'I' is the condition of the predicate 'think'. *It* thinks: but that this 'it' is precisely that famous old 'I' is, to put it mildly, only an assumption, an assertion, above all not an 'immediate certainty'. For even with this 'it thinks' one has already gone too far: this 'it' already contains an *interpretation* of the event and does not belong to the event itself.³⁴

²⁷ Armstrong (2000), p.102.

²⁸ *Ibid.*, p.101.

²⁹ *Ibid.*

³⁰ See Conze (1959), 'The Personalist Controversy', pp.192-7.

³¹ Ryle (1949). Similar approaches can also be found in other analytic literature – e.g. Geach (1972).

³² Hume (1967), p.180.

³³ Nietzsche (1990a), Part One, section 12, p.43. See also sections 16-21, and 32.

³⁴ *Ibid.*, section 17, p.47. The italics are in the original.

What we are faced with here, therefore, is the notion of the self (or soul) as a socially constructed entity (or, in this case, a *deconstructed* entity). Initially, this might seem like a negative view; Buddhism and Nietzsche have both been accused of nihilism. However, once we can see that it is possible for the self to be taken apart in this way, this leaves room for other concepts of self. But what might these possibilities involve?

The Expanded Self

Divergent notions of self are common in religious literature. In many cases, they take the form of what I will term an 'expanded' notion of self, which frequently has an experiential (as opposed to merely theoretical) basis.

So, for instance, the American poet Walt Whitman provides perhaps a typical expression of this view of self. One of the best examples of this is his long poem, 'Song of Myself', where he undertakes a vast, sprawling celebration of the panorama of human existence. The 'self' of the poem, however, is not simply that of the poet – or at least, not only that – but an embodiment of the whole range of human experience. In this sense, therefore, the poem approaches a mystico-religious vision of the 'expanded self'.

The poem begins:

I celebrate myself,
And what I assume you shall assume,
For every atom belonging to me as good as belongs to you.³⁵

From then on, the poet sets out a vision of a cosmic self that both is and is not the poet's own:

I am the mate and companion of all people, all just as immortal and fathomless
as myself;
They do not know how immortal, but I know.³⁶

In all people I see myself, none more and not one a barleycorn less,
And the good or bad I say of myself I say of them.³⁷

I am Walt Whitman, an American, one of the roughs, a kosmos³⁸

Furthermore, the self of the poem is therefore not one with a definite limit or moral perspective; it is one that includes the diverse range of all people, allowing the poet to identify with the totality of selves, their attitudes, desires, and perspectives.

Through me many long dumb voices,
Voices of the interminable generations of slaves,
Voices of prostitutes and of deformed persons,
Voices of the diseased and despairing, and of thieves and dwarfs,

³⁵ Whitman (1973), p.32 (all subsequent references are to the page numbers of this edition).

³⁶ Ibid., p.38.

³⁷ Ibid., p.51.

³⁸ Ibid., p.56.

[...]
Through me forbidden voices,
Voices of sexes and lusts....voices veiled, and I remove the veil,
Voices indecent by me clarified and transfigured.³⁹

Anticipating that the reader will have difficulty in seeing unity in such diversity, Whitman concludes:

Do I contradict myself?
Very well then....I contradict myself;
I am large....I contain multitudes.⁴⁰

One of the interesting consequences of the presentation of this self is the consequent attitude to death that it entails:

Has any one supposed it lucky to be born?
I hasten to inform him or her it is just as lucky to die, and I know it.

I pass death with the dying, and birth with the new-washed babe....and am not
contained between my hat and my boots⁴¹

Therefore, in his identification with all life, the poet becomes immortal:

And I know that I am deathless
[...]
I laugh at what you call dissolution⁴²

And since each of us is a part of the other, and of the overall process of life, death is simply a process of assimilation; therefore, he says,

Who need be afraid of the merge?⁴³

We may of course view this as simply an extended poetic conceit: an 'as if' identification for metaphorical purposes. However, whatever the particular case with Whitman's poem,⁴⁴ it can be seen that the notions of self and death are inextricably intertwined: with a 'narrow' notion of self, one where the self is 'contained between my hat and my boots', then death is subject to the processes of natural dissolution; however, with the 'expanded' self, such dissolution is defeated – or at least postponed.

In viewing the self as socially constructed, therefore, we may make the link between a 'narrow' self and the needs and dictates of society. It is tied in to the need for moral agency, education, economy, and the diverse natural and cultural influences that shape society. As a consequence, the man of expanded self – the 'transcendentalist', to use

³⁹ Ibid., p.57.

⁴⁰ Ibid., p.96.

⁴¹ Ibid., p.38.

⁴² Ibid., p.52.

⁴³ Ibid., p.38.

⁴⁴ It has been argued, however, that there *is* some pivotal mystical experience underlying Whitman's early poetry – see Andó (1970), pp.111-12.

the phrase of Ralph Waldo Emerson, Whitman's contemporary – frequently finds himself out of step, and socially surplus.

New, we confess, and by no means happy, is our condition: if you want the aid of our labor, we ourselves stand in great want of the labour. We are miserable with inaction. We perish of rest and rust: but we do not like your work.⁴⁵

Emerson's point is therefore that society is aimed at social utility, and therefore, when embodiments of divergent notions of self occur, they find themselves to be 'useless'.

A more traditional expression of this view can be found in the Hindu scriptures, notably the *Bhagavad Gita*, where Krishna, the divine incarnation, consoles his disciple Arjuna:

The wise grieve not for those who live; and they grieve not for those who die – for life and death shall pass away.

Because we all have been for all time: I, thou, and those kings of men. And we all shall be for all time, we all for ever and ever.⁴⁶

In this whole universe there is nothing higher than I. All the worlds have their rest in me, as pearls upon a string.

I am the taste of living waters and the light of the sun and the moon. I am OM, the sacred word of the Vedas, sound in silence, heroism in men.⁴⁷

And I am from everlasting the seed of eternal life. I am the intelligence of the intelligent. I am the beauty of the beautiful.⁴⁸

The self outlined here is identified almost entirely with universals, and not particulars. In this sense, the individual is seen as a transitory and passing embodiment of a greater self (just as we saw in the Buddhist metaphors of fire and running water). The notion of the expanded self is therefore – if not common – then widespread enough to occur in different cultures. Buddha or Christ might also say, with Whitman, that the expanded self is 'not contained between my hat and my boots'. Furthermore, one of the central themes which link these depictions of self is the consequent attitude to death:

If any man thinks he slays, and if another thinks he is slain, neither knows the ways of truth. The Eternal in man cannot kill: the Eternal in man cannot die.⁴⁹

⁴⁵ Emerson (1981), pp.105-5.

⁴⁶ Mascaro (1962), p.10.

⁴⁷ Ibid., p.36.

⁴⁸ Ibid., p.37.

⁴⁹ Ibid., p.11.

Therefore, in identifying with the processes of life as a whole, the expanded self escapes death – as the Elizabethan poet John Donne put it, since death of the individual self is just a ‘short sleepe’:

One short sleepe past, wee wake eternally,
And death shall be no more, Death thou shalt die.⁵⁰

But is it possible to ‘kill’ death merely by consciously revising our concept of self? As I have argued, divergent notions of self – especially that of the expanded self – are common themes of religious (Buddhist and Hindu) and semi-religious literature (Whitman and Emerson) from various cultures and periods. I have also suggested that, since the form of the self is largely dictated by social and cultural factors, that it may be possible to legitimately (i.e. actually) alter our notions of self (e.g. by education, social change, etc.). I will now consider the question of this legitimacy directly.

Death and Immortality

In the above sections, I have tried to outline the prevalence of an alternative concept of death, and also to show that it is one where the notion of self is expanded to a point where death, as we know it, is transcended. However, in presenting this view, I am not arguing for a religious point of view (although this *is* an example of a religious point of view); rather, I am trying to show that such religious perspectives exist, are not necessarily irrational, and represent what I have earlier termed a ‘debate with death’. However, having done this, I would like now to consider the deeper question of whether such positions are *coherent*. Atheists, among others, would argue that they are not, but – as I will show – the opposition here between accepting that such views are coherent and rejecting them as nonsense (literally) is not so straight forward as may be supposed (and, as many atheists suppose). Before returning to consideration of the significance of the notion of the expanded self, therefore, I must first consider some common objections (and possible solutions) to the problems faced by religious assertions concerning immortality and certain features of religious language in general.

In *Death and Immortality*,⁵¹ D.Z. Phillips defends a picture of religious language that seeks to avoid what he considers to be certain common misconceptions. Phillips argues that if we interpret religious language in a literal, propositional way, then it becomes – quite rightly – open to a host of rational objections.⁵² For instance, consider the statement that ‘God loves us’. Critics of religion – such as A.G.N. Flew – would argue that, in order for such a belief to be meaningful, it must at least be possible to say what might potentially falsify it.⁵³ Since, Flew argues, nothing would

⁵⁰ John Donne, ‘Holy Sonnets: Divine Meditations’, sonnet 6, in Gardner (1972), p.85.

⁵¹ Phillips (1970). For a concise statement of Phillips’s views on religion and their relation to Wittgenstein’s thought, see also Phillips (1971).

⁵² For example, see MacKinnon (1955), who – whilst having sympathy with some theological interpretation of immortality – rejects the common misconceptions which ignore ‘the logic of our language’ (p.262); and also, of course, Flew himself in the same collection (Flew, 1955), who further suggests that such positions as MacKinnon’s (and Phillips’s) are distortions of theological doctrine. Incidentally, Flew seems to have quite recently revised his atheistic position and now seems open to certain forms of theism. However, for my purposes, I use his ‘old’ views simply as representative of a certain common form of attack upon religious meaning (which many still hold).

⁵³ See, for instance, Flew’s famous ‘Theology and Falsification’, in Hick (1964), pp.225-8.

seem to count against the idea that 'God loves us' (not even the existence of natural evil such as earthquakes and hurricanes), then such language is meaningless.

Phillips's objection to this is that it misconstrues the meaning of religious language. For instance, the belief that we shall be reunited with our loved ones beyond the grave is not to be interpreted as a proposition to be verified or falsified by actual events, but rather as an expression of the religious believer's attitude to his family.

For Flew, the belief is shown to be confused by treating it as a prediction and by demonstrating the logical contradictoriness of what is predicted. I am suggesting that this procedure ignores the religious and ethical significance of the belief. A man may actually visualise his family, most of whom are dead, embracing each other in a reunion after death. Of course, awkward questions can be asked about whether they are really embracing each other, and if so, what is one to make of the identity of the buried or cremated bodies. But these questions are not simply awkward, they are also trivial. The picture of the family reunion may have an ethical and religious role in the person's life [...] The man may look on his relationship to his family as going beyond death, and his obligations to them as something which death cannot erase. The picture of the family reunion after death is not a prediction for which he has evidence, but a vision in terms of which much of his own life is lived out. The picture is not assessed by appeal to the evidence. On the contrary, *the picture, for this man, is the measure of assessment.*⁵⁴

Phillips, then, rejects Flew's objections because – for Phillips – religious language is non-propositional. As such, the picture which forms a part of a person's religious beliefs is to be understood not as consisting of propositional statements concerning a metaphysical reality, but rather as expressing certain moral and spiritual attitudes. If we then apply this understanding to survival of death, and the notion of immortality, we can see that the common objections (concerning personal identity, dualism, the existence of God, and so on), do not have any force.

If we ask ourselves when we would consider whether a man has a soul or not, we see that this has nothing to do with any kind of empirical question. It is not like asking whether he has a larynx or not.⁵⁵

Therefore:

Talk about the soul, then, is not talk about some strange sort of 'thing'. On the contrary, it is a kind of talk bound up with certain moral or religious reflections a man may make on the life he is leading.⁵⁶

So, to lose one's soul, to sell it, or to swear upon it, does not involve – for Phillips – assertions about the existence of some metaphysical entity, but rather concerns ways

⁵⁴ Phillips (1970), pp.68-9 (my italics).

⁵⁵ Ibid., p.44.

⁵⁶ Ibid. It is easy to see what such a position has in common with the general analytic critique of Cartesian dualism – e.g. Ryle (1949) and Geach (1972).

of speaking about one's life, attitudes, morality, etc., which form part of the wider discourse of religious belief.

In fact, Phillips argues, the religious conception of immortality is based on the fact that one's own personal self will *not* survive. Quoting Simone Weil:

The principal claim which we think we have on the universe is that our personality should continue. This claim implies all the others. The instinct of self-preservation makes us feel this continuation to be a necessity, and we believe that a necessity is a right.⁵⁷

The frustration of this desire, therefore (both Phillips and Weil argue), is the basis of the religious attitude of 'dying to the self':

Belief in [personality-based] immortality is harmful because it is not in our power to conceive of the soul as really incorporeal. So this belief is in fact a belief in the prolongation of life, and it robs death of its purpose.⁵⁸

This is, I think, a point worthy of any existentialist.⁵⁹ To imagine a divine reward, or to act out of the belief that there will be an extension of our earthly life in a post-mortem state, is to live in 'bad faith' (as Sartre would say). For Weil and Phillips, therefore, immortality consists in the attitude to life embodied by the individual in their religious practices. As Phillips concludes:

I am suggesting then, that eternal life for the believer is participation in the life of God, and that this life has to do with dying to the self, seeing that all things are a gift from God, that nothing is ours by right of necessity.⁶⁰

My purpose in presenting Phillips's views here is primarily to show that it is possible to entertain a picture of religious language which is not at odds with conventional scientific views of the nature of the human being. In other words, we need not presuppose the existence of a metaphysical soul, or even of God (conceived in a way which Phillips objects to), in order for religious language to make sense. Of course, Phillips has his critics – from both sides of the religious debate – but the strength of his position is that it allows religious language to be meaningful without either completely distorting its nature, or challenging accepted scientific dogma.

However, in light of my conclusions concerning the nature and status of the concept of death, I would now like to propose that, whilst Phillips's approach goes some way towards establishing how we can legitimately possess an alternative view of death, it does not actually go far enough. This is, I shall shortly argue, because there is an underlying conservatism to the position. Therefore, since Phillips's position is itself based on that of Wittgenstein, I shall argue that such conservatism is an aspect of Wittgenstein's own philosophy. Perceiving this, and taking account of the greater conceptual freedom which this reveals will, I shall argue, help us to understand a wider spectrum of social and religious discourse.

⁵⁷ Weil (1959), p.174.

⁵⁸ Weil (1951), p.33.

⁵⁹ See, for instance, Cox (1999).

⁶⁰ Phillips (1970), pp.54-5.

Wittgenstein's Conservatism

As a number of commentators have noted, there is a discernible conservatism to Wittgensteinian thought which is at odds with other trends in philosophy. As J. C. Nyíri notes, some of Wittgenstein's traits are typical of the figure of the conservative, who is

devoted to the familiar and mistrustful of all novelties; [...] affirms instinctively the durable, the constant, the traditional; he is sceptical of every radicalism, of utopias, and of promises in regard to the future⁶¹

Russell B. Goodman (in his study of the influence of William James upon Wittgenstein) observes that:

In the *Investigations*, Wittgenstein wants to return, to bring words home from their metaphysical to their everyday use. As Wittgenstein put it late in his life, “[where] others go on ahead, I stay in one place.”⁶²

Thus, when faced with philosophical problems, Wittgenstein's recourse to a consideration of how we use language in everyday life is seen by some to represent a conservative (as opposed to a radical) move; it considers the problems as aberrations caused by ‘language going on holiday’, and seeks to re-ground the debate by returning it to a consideration of the practices that gave birth to the concepts. As Ernest Gellner notes, however, there is a circularity involved here:

To ‘observe how we use words’ is to make statements, in ordinary language, about the role, function, effects and context of expressions. But in doing this, the concepts and presuppositions of that ordinary language are taken for granted and insinuated as the only possible view. We smile at scholasticism which *had* to prove what Revelation supplied anyway. But Linguistic Philosophy *has* to prove, ‘make acceptable’, what common sense or linguistic custom provides anyway. (Linguistic Philosophy is scholastic in this sense as well as in its pedantry.)⁶³

So, in relation to Phillips's view of religious immortality already considered, the proposition that religious language is true because of as-yet-undiscovered metaphysical truths is rejected, because it would involve a radical revision of science (and also ‘common sense’ linguistic practices). However, the reinterpretation of religious language so as to conform to scientific knowledge is seen as a return to the ‘original’ meaning of the practice. In some cases, this can involve an almost counterintuitive interpretation of the material under consideration. For instance, in

⁶¹ Nyíri (1982), p.46, quoting Gerd-Klaus Kaltenbrunner.

⁶² Goodman (2002), p.167. The quote is from Wittgenstein (1980), p.66. For other interpretations of Wittgenstein as a conservative (or ‘neo-conservative’), see for instance, Janik (1985), Norris (2003), and Wheeler (1988). It should not be thought, incidentally, that Goodman sees James as providing a conservative influence; rather, he sees it as a general point of contrast between Wittgenstein and the American pragmatists.

⁶³ Gellner (1959), p.217. Gellner's attack in this book is directed at linguistic philosophy as a general movement, but since this springs mainly from Wittgenstein – who is also one of Gellner's main targets – then his criticisms are pertinent here.

'Remarks on Frazer's Golden Bough',⁶⁴ Wittgenstein criticizes the anthropologist Sir James Frazer for what he considers to be a misinterpretation of the religious practices of primitives peoples. Instead of attributing to them a false scientific world view (where rain dances affect the weather) as Frazer does, we should rather – Wittgenstein argues – observe that such practices can be made sense of in other ways (such as a celebration of the anticipated coming of the rains).⁶⁵ However, such an interpretation is arguably at odds with much of our understanding of how such practices are understood by the practitioners themselves; and, furthermore, may be said to be applicable only to a limited range of those practices. For instance, consider a shaman who enters the spirit world to find a familiar spirit, or performs a ritual which is intended to injure an enemy at a distance, or to heal someone: should we consider the shaman as 'mistaken' in considering these practices to have the purpose he intends them to have? Interpreting such practices in this 'expressivist' way would therefore seem to argue for a privileged account of what is true about religion, identifying only certain practices and attitudes as 'genuine'. Of course, if the objective is simply to rescue a certain type of religious practice from the criticisms of science and rationalism, then this is less problematic; however, as an interpretation of religious and magical practice as a whole, it is less successful.

Needless to say, the interpretation of Wittgenstein's philosophy as fundamentally conservative has been strongly resisted by many modern-day Wittgensteinians.⁶⁶ As Denis McManus points out, we have to be careful how we interpret Wittgenstein; since, like Nietzsche, his later philosophy is non-systematic, and occasionally cryptic, our criticisms may provide no more than a form of straw man argument. Furthermore, however – if he is not conservative – we should also not assume the opposite – i.e. that Wittgenstein is actually always proposing a radical position:

The fact remains that when thinking about criticism of established practices, it IS easy to slip into ways of speaking which suggest that such criticism does indeed rest on our exposing 'the true meaning of democracy' or such like. Thus to defend Wittgenstein here one must examine these philosophical temptations with the same sensitivity that Wittgenstein himself displays, rather than simply saying that he never defends conservative theses.⁶⁷

The apology for Wittgenstein's apparent conservatism (if it is there at all) is therefore that 'it IS easy to slip into ways of speaking' which suggest a conservative approach.

⁶⁴ Wittgenstein (1979), pp.61-81.

⁶⁵ This might be termed the *expressivist* interpretation.

⁶⁶ See, for example, McManus (1995), and Diamond (1991), pp.13-38. Diamond argues that 'The idea of Wittgenstein's philosophy as inherently *conservative* is nutty. At the heart of his discussion of that most profoundly rule-governed of all our intellectual activities, mathematics, is his account of interesting mathematical questions and mathematical developments that are initiated by them as involving kinds of new move that can illuminatingly be compared with riddles. The interpretation of his philosophy as conservative results in part from lack of attention to what he says about mathematics; but it is supported also by what I earlier described as the imposition on philosophy itself of the philosophical requirement that it lay down requirements. And that is the refusal of the kind of liberation that he hoped philosophy might bring.' (p.34). However, this aside, it would seem to be easier to interpret Wittgenstein's comments on language in line with a form of conservatism than in any other way.

⁶⁷ McManus (1995).

However, in the hands of Wittgensteinians, this ‘natural tendency’ often becomes a point of dogma, used to resist not only radical possibilities for language and linguistic practice, but even any proposed novelty or potential mutability (such as, in fact, I am proposing).

However, without getting drawn into a battle for Wittgenstein’s philosophical estate (and over what Wittgenstein did and did not mean – which I have neither the space nor inclination to undertake), I would like to propose that a modified understanding of Wittgenstein’s picture of social and epistemic practices (which I considered above in his river analogy) would give us a better framework for considering the possible scope of redefinition of the concepts of self and death.

For instance, in my discussion of what I termed the ‘expanded self’, I argued that a certain concept of the self prevalent in literature and religious texts envisaged the conquering of death by identification with the wider processes of life itself. However, if we were to apply a Phillipsian/Wittgensteinian approach to this, we would find such identification would take on a purely ‘poetic’ or ‘expressivistic’ sense, and any divergent interpretation would then be contrasted with the everyday sense in which terms such as ‘I’, ‘death’, and ‘survival’ make sense. However, in exposing Wittgenstein’s conservatism (or the conservative interpretation of Wittgenstein – whichever you prefer), it is possible to advance a more dynamic relationship between these notions. In this way, ‘death’ and ‘self’ become more fluid concepts; still determined by human practices (and the way in which values and epistemic concerns are intertwined at a deep level), but more open to change and innovation. In this way, we avoid what I think is the somewhat artificial division that arguably arises out of an application of Wittgenstein’s ideas to religion. Phillips has sometimes been accused of ‘Wittgensteinian fideism’, where, because we cannot criticise religious practices from an external rational perspective (but only understand them on their own terms), they would seem to represent a form of rationally unfounded, unimpeachable belief. However, by removing the conservative element, we open up the whole debate, connecting these fundamental concepts at deep levels. In other words, no longer is religious language ‘playing a game’ that rational discourse cannot join in (or ruin); rather, ontological and epistemic assertions can form part of the same discourse.

To attempt to illustrate this, consider this notion of the expanded self. From an ‘everyday’ linguistic perspective, we might consider this to be no more than a metaphor, and that many things argue against its interpretation in a literal sense: the fact that our bodies are separate entities (and that our notion of personal identity is dependent upon this); the fact that we can only see things from our own perspective (and not from others’); the fact that we cannot feel what it is like to be other than ourselves. However, as I hope I have shown, these criteria do not possess the inviolable status that some philosophers seem to assume: for instance, if we were to adopt some of Parfit’s conclusions regarding personal identity, take into account a Nietzschean critique of the Cartesian self, and entertain a Buddhist account of causation, then the inviolable nature of these concepts begins to unravel: all living things share the same substratum of matter; consciousness may not necessarily be the *sine qua non* of personal identity, and qualia (as indicating an individual, experiential self) may be considered an illusion. Furthermore, if we attempt to appeal to ‘deeper’ notions, then these can in turn be questioned, thus throwing the whole nature of self and death open to change. In other words, if Nietzsche and Buddhism are right, and

the notion of the 'atomic' self is a metaphysical assumption which requires investment, then why can't such an investment be made *elsewhere* in order to produce a different idea of self? As I have argued in my earlier discussion of Emerson's point, it is social and cultural factors (in combination with other contingent factors) which determine and maintain the rigid need for an individual self, and these are certainly open to change.

I would not want it to be thought that I am proposing a 'shop-window' view of concepts here, where I may pick and choose what I want to consider as myself, or arbitrarily decide how or when I might die. The force of (for example) Wittgenstein's and Searle's arguments considering the importance of the social community as a whole (be it in terms of 'collective intentionality' or 'forms of life') is hard to deny: we need a linguistic community in order to have meaning at all. However, just what this linguistic community believes, or what its concepts of 'self' and 'death' are (I argue) are considerably broader than it seems either of these philosophers would concede.

Also – and I think this is an important consideration in the current context – it should be borne in mind that the type of 'expanded' concepts that I have been considering here do traditionally arise from specific linguistic communities or individuals in isolation (e.g. monasteries, religious groups and institutions, etc.). Furthermore, where the practice is especially determined and 'serious', the secluded nature of these religious communities, the physical and mental disciplines involved, etc., further suggest that the focus of effort is on changing the individual mind-set in line with goals prescribed by the linguistic community as a whole (or else the individual's mental and linguistic attachment to the 'old' community). In this way, there is possibly more chance of these concepts coming 'alive' for the practitioner, and leaving the land of mere metaphor.

Death and the Future: Science-Fiction

In my discussion of religion, I have attempted to show that it is possible to consider certain religious attitudes as presenting us with possible alternative conceptions of the self and death. In doing this, I have also tried to argue that a non-conservative application of a Wittgensteinian approach to language and social practices can give us a framework for understanding how these changes are possible.

However, just as religious concepts are influenced by accepted dogmas and traditions, different conceptions of the self and what it means to die will be found in areas which are influenced by different assumptions and values. Therefore, I would now like to consider the concept of death in relation to certain scenarios in the realms of scientific hypothesis and science-fiction.

In a sense, there is no clear distinction between science fiction and science fact. For, whilst there are certain areas of the sci-fi genre which present us with outlandish and unlikely possible futures, many of the most effective plots and envisionings involve scenarios which are only marginally advanced beyond our own time. Therefore, because technological advance in modern societies seems to progress at an exponential rate, the idea that such future realities are potentially just around the corner lends a greater credibility to the ideas and situations portrayed.

In relation to death, however, science fiction's notion of immortality traditionally focuses on the way in which existing life might be prolonged and improved: so, if religion can be said to focus on 'dying to the self' and the immortality of the 'soul', then science fiction considers the ways in which the life of the mundane personality and physical self might be extended. In the following sections, therefore, I will focus on certain suggestive scenarios from scientific research, film, and literature which propose different ways in which the concepts of 'self' and 'death' may be revised.⁶⁸

The Immortal Body

The flip-side of spiritual immortality is, of course, bodily survival. However, whilst different possibilities as to what we mean by death and survival in a 'spiritual' sense may be said to rely on the potential mutability of the concept of self, the equivalent possibilities as related to bodily immortality are shaped by technology. As James J. Hughes succinctly notes: 'Technology is problematizing death'.⁶⁹ Of course, this is itself something which I have already stressed in earlier chapters; advances in medical technology have resulted in novel situations in which traditional notions of 'person', 'dead', 'alive', etc., have been split apart, and the criteria which were thought to constitute them exposed and set at variance with each other. However, if we now consider the possibility of physical immortality (or, more humbly, just the extension of life), then technology (and its projection into the future: science fiction) is the main vehicle for our doing so.

First of all, let us consider the possible means of extending life:⁷⁰ we have cryonics, or the freezing of all or part of the body, in the hope that future scientists will be able to revive and treat patients with conditions that we currently think of as irremediable; calorie restriction, which notes that the lives of many animals are extended by restricting food intake; the use of nutritional supplements as antioxidants; the possibility of genetic manipulation to overcome biologically programmed cell death; the use of cloning to replace aging body parts; and so-called 'mind uploading', where it is imagined that one day we will be able to upload the information contained in an individual's brain so that they can continue to exist on a more durable physical vehicle (e.g. a computer hard drive).

Obviously, some of the above represent more immediate possibilities than others. However, whilst all of them share the premise that individual survival consists in the prolongation of physical aspects of the human being,⁷¹ there are differences as to what exactly constitutes physical survival. Interestingly, therefore, the different approaches also reveal different possibilities of selfhood and personal identity. For instance, James J. Hughes argues that whilst the problems we face in determining death were revealed by technological advances,

⁶⁸ Obviously, this is only a general characterization, and the opposite of this is also true in some respects: some forms of Christianity believe in physical, bodily resurrection, whilst some science-fiction scenarios consider the possibility of disembodied consciousness.

⁶⁹ Hughes (2001), p.8.

⁷⁰ What follows is a sample of the range of approaches currently favoured by life extensionists – see Saunders (2002).

⁷¹ In the 'mind upload' scenario, where we are thinking of mind as an organisation of physical neurons, it is the physical organisation we are preserving, not some immaterial entity.

Technology did not really create this gray area, but extended it and made it manifest. Death has always been a process, rather than a clearly binary state. In the Buddhist or Parfitian (1984) view to which I am partial, there is no essential or real identity in things. The boundaries we draw around “life” and the “self” are arbitrary, motivated by specific interests and purposes. Life and the self have no essential reality which can be definitively discerned, or boundaries which can be definitively marked. Rather there are a variety of processes involved in being born or dying, processes involved in the illusion of continuous self-identity. The lines that get drawn have mostly to do with the politics, economics, culture and technology of those doing the drawing.⁷²

It can be seen from the above that this is a position that I have a great deal of sympathy with: as I have argued, even if we adopt a realist view of the world, we cannot consider death to be an intrinsic feature of it; as for death, so for self. Hughes also argues that the death boundary will be constantly challenged as science and technology find new ways of preserving life.

Eventually victims of devastating neurological damage, who would previously have been declared hopeless or even dead, will be seen as still potentially living patients deserving of a trial of reparative therapy, unless portions of the brain holding identity-critical structures are demonstrably and thoroughly destroyed. If the restoration is unsuccessful, the patient may then be allowed to completely die.⁷³

Death therefore occurs when we can do nothing more – *at this time*. If we accept such views, then we are therefore faced with a redefinition of death in relation to two aspects: (1) in accordance with technological advances (such as the ability to treat previously irremediable conditions, repair cell death, etc.); and (2) in relation to shifting notions of self and personal identity.

Point (1) is obviously something that we already face, and will arguably continue to face in the future; it seems silly to deny the possibility of technological advances whose application will rescue those whom we now consider to be beyond treatment. Point (2), however, represents a more controversial claim, and involves accepting the argument that I have been advancing – i.e. that death is not an intrinsic feature, that it is tied to equally shifting notions of personal identity and selfhood, and that we therefore face multiple possible ‘redefinitions’ of death and self. Since it is this latter point that I am primarily interested in here, I will therefore consider what the possibility of such science-fiction type futures mean for the concepts of death and self.

Problems and Solutions

It may be argued that technological advance and scientific knowledge are mixed blessings: for every benefit, there is also a corresponding problem, and boon and bane seem to go hand in hand. Obviously, this twofold character is best illustrated by the

⁷² Hughes (2001), p.8.

⁷³ Ibid., p.13.

application of scientific knowledge to medicine and military defence respectively; atom bomb and the MRI scanner spring from the same root.⁷⁴

In terms of the potential expansion and revision of the cluster of concepts with which we are concerned in understanding death, science-fiction scenarios therefore present us with both solutions and problems. For instance, in the ‘mind upload’ scenario we are faced with the possibility of almost indefinite survival of the personality (and possibly consciousness – depending on the theory of philosophy of mind that is held). However, we are also simultaneously presented with the problem of loss of identity: if it is possible to codify the personality and brain waves of an individual using digital technology, then – because the information is digital – it can also be copied. We are therefore presented with a typical Parfitian-style example of ‘survival without identity’.⁷⁵

However, a similar scenario is also presented by physical cloning. In the film *The Sixth Day*,⁷⁶ Arnold Schwarzenegger plays Adam Gibson, a helicopter pilot, who is unknowingly cloned. The plan is for Gibson to be killed, and for the clone to take his place (and thereby fulfil some nefarious purpose that the bad guys have in mind); however, by mistake, someone else is killed, Gibson survives, and we are left with two Gibsons. So, the question is, who is the real one? As Mark Rowlands points out,

Suppose we ask the question, ‘Who is the real Adam Gibson?’ Chances are we’ll come up with the answer: ‘The one who was around first.’ The fact that one Adam Gibson was around before the other strongly suggests that the first one is kosher, and the second one a mere copy. So, let’s amend the story slightly. Let’s suppose, as the cloning baddies assumed, that the body of the original Adam Gibson is destroyed, and his memories are implanted into a custom-made clone. Does Adam Gibson survive? Or, rather, if the [memory-based theory of personal identity] is true, does this mean that Adam Gibson survives?⁷⁷

Obviously, if we argue that some version of the psychological-continuity theory of personal identity must be subscribed to, then this clone is indeed Adam Gibson. Of course, we might seek to get around this by arguing that psychological continuity isn’t enough; we must also have physical continuity. However, as Rowlands also points out,⁷⁸ we can further amend the scenario so that the clones are created at the same time, and both of them receive a ‘download’ of Adam Gibson’s memories. Which, then, is the ‘real’ Adam Gibson? As Rowlands notes – in line with Parfit before him – we have no criteria for distinguishing between the two (in terms of status), and in addition we are left with another situation where there is survival, but no personal identity.

James J. Hughes takes this type of scenario as conclusive, and argues that, in the development of technological means of stretching the boundaries of death, we are

⁷⁴ Magnetic Resonance Imaging is used to produce cross-sections of living tissue using low-level nuclear radiation.

⁷⁵ Parfit (1984).

⁷⁶ Spottiswoode (2000).

⁷⁷ Rowlands (2003), p.111.

⁷⁸ Ibid., p.113.

faced with a profusion of possible definitions of death and self, each driven by individual choice:

Once technology has fully teased out the constituent processes and structures of memory, cognition and personality, and given us control over them; once we are able to share or sell our skills, personality traits and memories; once some individuals begin to abandon individuality for new forms of collective identity; then the edifice of Western ethical thought since the Enlightenment will be in terminal crisis.⁷⁹

Hughes sees the sort of conceptual freedom I have suggested exists as the basis of a number of what he terms 'technological threats to liberal individualistic assumptions', leading to such things as:

Identity Malleability: Parental, social and personal control of memory, identity and personality

Posthuman Persons: Radically enhanced minds

Identity Sharing: Memories, thoughts and skills, sold or shared

Identity Cloning: Persons multiply copied into new media

Distributed Identity: Distinct persons distributed over, or sharing, a set of bodies and machines

Group Identity: Multiple bodies and machines integrated into a collective identity, without clear personal identity, e.g. Borg or "hive minds"⁸⁰

Of course, many of these scenarios have already been explored in science-fiction literature and film: in *Star Trek: The Next Generation*,⁸¹ the 'Borg' are a collective consciousness into which individuals are 'assimilated' (Group Identity); in *The Lawnmower Man*,⁸² a young, mentally-retarded gardener is subjected to a computer training programme which increases his intelligence to an almost god-like level (Posthuman Persons); in Phillip K. Dick's short story, *We Can Remember it for You Wholesale*⁸³ (later filmed as *Total Recall*),⁸⁴ memories are bought and sold, and the experiences are seamlessly interwoven into the subject's own personal history.

A similar scope for conceptual expansion is proposed by 1960s LSD guru Timothy Leary in his final book, *Design for Dying*,⁸⁵ which was written while the author was dying of prostate cancer. Leary's perspective is, as can be imagined, that of challenging conventional viewpoints and exploring possibilities. For example, he

⁷⁹ Hughes (2001), p.23.

⁸⁰ Ibid., pp.21-2

⁸¹ Roddenberry (1987-94).

⁸² Leonard (1992). The original title of the film was *Cyber God*, but since the production company also owned the rights to Stephen King's short story of the same name, the title became *Stephen King's Lawnmower Man*. However, since the film bears almost no resemblance to King's story, the author successfully sued to have his name removed.

⁸³ Dick (2000), pp.157-74.

⁸⁴ Verhoeven (1990).

⁸⁵ Leary and Sirius (1997).

considers how the advent of the internet and related cyber technologies have impacted upon – and will continue to expand – our mundane perceptions of self and identity; ‘delocalizing the self’,⁸⁶ as he puts it. Imagine, he says, that we can build a ‘simulacrum’ of ourselves; a robot surrogate that can go anywhere we want, and to which we are remotely connected; seeing through its ‘eyes’, having its experiences relayed to us as we sit in the operator’s room. However, now imagine that, instead of sending it out to some distant place, we go in search of ourself.

You walk your simulacrum down the street, which looks familiar. Hey, it’s the street that the apartment with your operator’s room is on. You walk down the familiar corridor often travelled by your meat machine and open the door to the room where your body hangs in suspension. You peer into the suspension chamber and you see...what?⁸⁷

Leary suggests that this ‘feedback loop’ will cause you to disassociate from your real body:

the physical body operating the machine would seem alien, as distinct from self as from other artefacts in the world. The sensation of self would be in the simulacrum, where the viewpoint is.⁸⁸

Therefore, for Leary, the conclusion is that, ‘You exist where your viewpoint is’. This means that, potentially, ‘you’, removed from the ‘traditional boundaries of the body’, could exist anywhere that could support your awareness. This is ultimately because the self is an arbitrary concept.

Of course, where one chooses to draw the line distinguishing inside from outside, self from other-than-self is formally arbitrary, a matter of operational convenience. In certain mystical states, self expands to fill the universe. In death, phenomenologically speaking, the universe encroaches to diminish self to nothingness. *The effect of removing the distinction of self in both instances is identical in the limit.* There is no difference and so we might presume that in fact the experiences are equivalent.⁸⁹

What Leary is pointing to here is an almost exact parallel to the religious practice of ‘dying to the self’: the extinction of self is effectively the same as the identification of the self with the universe. The only difference lies in the means adopted (i.e. cyber technology).

It is perhaps pertinent here to consider a common objection to such ‘delocalization’, which is also a potential criticism of the ‘expanded self’ hypothesis considered earlier. It is all very well, you might say, to consider delocalization whilst the body is alive to provide the vehicle for consciousness, but that itself suggests the limits to the experiment: the potential for consciousness is located in the body; therefore, the self

⁸⁶ Ibid., p.27.

⁸⁷ Ibid., p.27-8.

⁸⁸ Ibid., p.28.

⁸⁹ Ibid., p.29-30 (my italics).

cannot itself exist outside the body in anything but a metaphorical sense. Is there anything the advocate of delocalization can say in response to this?

There are a number of possible responses, and they would differ according to the approach favoured: one approach is obviously to consider the possible existence of an alternative vehicle for consciousness. In religious terms, this would obviously be a disembodied spirit; in sci-fi terms, an artificial body or digital brain. However, stemming from my earlier discussion of the competing criteria that we are faced with in determining what exactly the essential nature of our concept of a 'person' is in relation to the definition of death debate, we might argue that consciousness is not necessarily a key ingredient. It might be argued, for example, that the biological links between one human being and another are more significant than the existence of personal consciousness. As Leary argues:

The DNA code is a three-billion-year-old time capsule of consciousness. The DNA code is the miniaturized, invisible essence of wisdom of life. Most of the characteristics formerly attributed to "soul" describe the functions of DNA.⁹⁰

Therefore, if DNA can be considered the essence of who we are (our 'soul'), and our living has played a part in its evolution, then does the suggestion that we live on in its continued existence grant us a form of 'immortality'?

The point to be grasped here is that, even if we consider the capacity for phenomenal consciousness as a *sine qua non* of any definition of the self, it must be realised what sort of move this is. It is not, as it is tempting to think, an analytical observation, as if we were simply analysing the concept of self; it is rather, as Gervais has argued, a 'decision of significance'.⁹¹

Therefore, whilst it might seem odd – and counterintuitive in the extreme – to do so, there is nothing logically unsound in Leary's proposition that the self can be expanded (and survival thought possible) in ways which exclude personal consciousness.

What's it like, phenomenologically, to be dead? [...] Let's presume for the sake of argument that you're not in Algiers. Fine, then *you* are, right now, dead in Algiers. You experience no input from there, you affect nothing. People and things are unaware of you. You don't exist there.

This argument can be extended. You are dead most places in the universe at this moment. Dead in Paris. Dead in New York. (Assuming you don't happen to be reading this in either of those places.)⁹²

So, if we entertain the converse of this, then it is also possible to be 'alive' in other senses:

⁹⁰ Ibid., p.20.

⁹¹ Gervais (1986), p.2.

⁹² Leary and Sirius (1997), pp.33-4.

Of course, this “liveness” has degrees. Timothy Leary is on television right now in Paris. Somebody is reading an R. U. Sirius essay in Algiers. Phone a friend in China. You are alive there – a little bit.⁹³

Therefore, if we accept this continuum of life/death, self/not-self, then we might consider that survival may lie in a point at which a certain aspect of you lives on (a photograph, a memory held by someone else, a book you wrote, and so on), whilst consciousness does not. However, because ‘liveness’ (and ‘selfness’) is a continuum, then there would seem to be an element of choice as to where we choose to draw the line – if we draw it at all.

What’s more, Leary sees technology – especially telecommunications and the internet – as playing a key role in this expansion of self and death:

The beauty of information/communications technologies is their ability to extend the boundaries of self, to diminish distance and other physical limitations, and permit an individual to reach out nearly undiminished across time and space at the speed of light.⁹⁴

The above position is similar, in some ways, to that advanced by the French Jesuit Priest and palaeontologist, Pierre Teilhard de Chardin. Teilhard observed, in what some have taken as a remarkably prescient prediction of modern developments (such as the internet), that mankind was evolving toward greater unity. So, whilst evolution in the animal kingdom seemed to continue to embody divergence, human evolution seemed to progress along lines of *convergence*.⁹⁵ Furthermore, as we become aware of each other as part of the same global community, mankind begins to think of itself more and more as a single entity. As N. M. Wildiers puts it in his introduction to Teilhard:

The great body of mankind must be built up out of the many cells which now lead their separate existences. Just as the human brain consists of millions of neurons which have been composed into a unity through innumerable connections and combinations and thus make possible the unitary consciousness in man, so must men combine with one another to form a kind of super-organism in which a communal consciousness, a suprapersonal unity, would be manifested – with this difference: that the cells which constitute our brain no longer have an individual existence, whereas man, in virtue of his reflective consciousness, retains his individual freedom and separate existence, even within the larger organization.⁹⁶

However, within this global organization of individuals, there is a unifying point of reference: what Teilhard calls the *omega point*. It is this focus which therefore acts as the true centre of global unity, forming ‘a suprapersonal centre on which the whole of

⁹³ Ibid., p.34.

⁹⁴ Ibid.

⁹⁵ See Wildiers (1968), pp.83-108.

⁹⁶ Ibid., pp.100-1.

evolution converges.⁹⁷ In this sense, Teilhard's vision is almost an actualisation of the poet John Donne's famous picture of humanity as a continent, where

No man is an island, entire of itself; every man is a piece of the continent, a part of the main. If a clod be washed away by the sea, Europe is the less, as well as if a promontory were, as well as if a manor of thy friend's or of thine own were. Any man's death diminishes me, because I am involved in mankind; and therefore never send to know for whom the bell tolls; it tolls for thee.⁹⁸

Some of the above ideas may seem far fetched – and they are. There is nothing mundane about the notions of self and death involved in such scenarios. However, if my earlier arguments concerning the status of these concepts are accepted, then these possibilities would seem to be real (albeit 'far-fetched') possibilities.

In one sense, we might argue that such 'survival' is not really immortality at all: it is like, to take a literary example, assuming that the young man to whom Shakespeare promised immortality in his sonnets actually *did* live on:

But thy eternal Sommer shall not fade,
Nor loose possession of that faire thou ow'st,
Nor shall death brag thou wandr'st in his shade,
When in eternal lines to time thou grow'st,
So long as men can breath or eyes can see,
So long live this, and this gives life to thee.⁹⁹

Here, the poet promises that the poems themselves shall give the subject a certain immortality; elsewhere, a similar argument is made for siring children (so that the subject can 'live on' in their likeness).¹⁰⁰ It might be tempting to dismiss this as a form of vicarious immortality; as A. G. N. Flew might say, this sort of immortality would be of the same interest to me as 'the news that my appendix would be preserved eternally in a bottle'¹⁰¹ (i.e. none). However, even so, this would be to ignore the central point: the question of immortality, and the nature of self and death, *are* determined by what is important to us. So, whilst Flew might be disappointed by such a 'get-out clause' from the traditional notions of immortality and survival (and what he considers to be their inherent confusions), this does not impugn the validity of the arguments themselves.

Techno-Utopia and Group Minds

There would therefore seem to be a range of potential ways in which the traditional notion of self can be expanded, and thereby – in some cases – 'overcome' death. However, the prevalence or ingenuity of these ideas does not necessarily argue for their coherence, and we need not follow Leary and Hughes's enthusiastic acceptance of the way that such potentialities will shape our collective future.

⁹⁷ Ibid., p.108.

⁹⁸ John Donne, 'Meditation XVII', from Donne (1999).

⁹⁹ William Shakespeare, 'Sonnet 18', in Seymour-Smith (1963), p.49.

¹⁰⁰ This is largely the subject of the first seventeen sonnets.

¹⁰¹ Flew (1955), p.270.

Having so far argued that the concept of death is a more potentially variable one than is commonly thought, and that the notions of self and personal identity are equally open to change, I want now therefore to point out what I think are limits to this exercise. Hughes's 'techno-utopia',¹⁰² whilst logically – and maybe even technologically – possible, is nonetheless driven by a human fear of death. As Rodney A. Brooks perspicaciously observes:

my former postdoctoral research student Pattie Maes gave a talk titled "Why Immortality is a Dead Idea." She took as many people as she could find who had publicly predicted downloading of consciousness into silicon, and plotted the dates of their predictions, along with when they themselves would turn seventy years old. Not too surprisingly, the years matched up for each of them. Three score and ten years from their individual births, technology would be ripe for them to download their consciousness into a computer. Just in the nick of time! They were each, in their own minds, going to be remarkably lucky, to be in just the right place at the right time.¹⁰³

This overestimation of the speed of scientific progress (and the psychological motivations for it) aside, such envisionings also place a limitless faith in the potential of scientific advance – which may, ultimately, turn out to be misplaced:

This strong version of salvation seems plausible in principle. However, we may yet be hundreds of years off in figuring out just how to do it. It takes computational chauvinism to new heights. It neglects the primary role played by the bath of neurotransmitters and hormones in which our neuronal cells swim. It neglects the role of our body in placing constraints and providing noncomputational aspects to our existence. And it may be completely missing the *juice*.¹⁰⁴

Therefore, whilst we cannot dismiss such possibilities out of hand, we must consider that it is just as likely that science may find out that such things may prove – for some as-yet-unknown reason (e.g. the inability to replicate that which makes something consciously alive – the 'juice') – technologically impossible. For these reasons, I would wish to distinguish between the possibilities that science-fiction scenarios present us with, and the conclusions that we may draw from them; it may never prove possible (for purely technical reasons) to 'upload' minds into a computer, but this does not invalidate the conclusions of the thought-experiment that we could draw from it. Therefore, whilst those reasons remain 'purely technical', then the Parfitian conclusions that spring from them are valid; but if it should be realised that there is a *logical* impossibility in such a supposed scenario, then our conclusions must be revisited. So, as things stand, I would endorse the logical possibility of Hughes and Parfit's visions of self (and therefore death), whilst holding reservations about the achievability of such situations.

¹⁰² As Rodney A. Brooks terms it, see Brooks (2002), p.205.

¹⁰³ Brooks (2002), p.206.

¹⁰⁴ Ibid.

Aside from practical considerations, however, we must also consider a different sense of likelihood. Hughes talks as if such total revisions of 'self' and 'death' were merely round the corner. However, in a similar way to the mind-upload enthusiasts in Brooks's example, there is an over-enthusiasm which blinds him to what is actually involved in such a process of change.

As I intimated in the discussion of the expanded self in relation to religious perspectives, the different notions that such approaches to self and death employ are embedded in daily practices as part of what Wittgenstein and Phillips would term 'forms of life'. In arguing that such changes are possible, therefore, I have not glibly accepted that death is about to be conquered, or that 'cosmic consciousness' is just an attitude of mind. Rather, I have tried to argue that such revisions, whilst potentially possible (given a certain interpretation of things), represent a subterranean shift of deep-seated values and attitudes that are embedded in an individual's deepest habits of mind and body. In light of this, I argued, we can best understand traditional religious/mystical discipline and life as an attempt to change and evolve the individual at these very depths. However, the techno-glibness aside, Hughes's contentions represent an almost cavalier ignorance of what is involved in such potential change.

To illustrate this, imagine what would be involved in the production of a 'hive mind'. If we set aside the question of consciousness for the moment, the whole of society would have to function as a collective, with clearly defined roles for 'workers', 'soldiers', and their attitude towards the 'queen'. This would involve a suppression or weeding-out of individuality, together with prescriptive behaviour for every member of the collective, and some sort of method of bonding individuals together and to their common purpose. Even if this were undertaken on a small scale, the habits of identity and role would have to be embedded into every aspect of life.

Robert Rupert also points out certain problems for the current enthusiasm for attributing collective unity to existing or potential 'group minds' and the possibility that they may possess 'mental states'. As he argues, the ability to form mental representations would seem to be central to our concept of a mind, and yet

our best accounts of mental content seem to exploit architectural features that groups typically lack: many of our best explanations of how mental representations get their content assign a privileged role to perceptual or quasiperceptual processing, thereby requiring a cognitive architecture that group systems typically do not possess. Problems specific to the application of the various theories of content further dim the hopes that group systems, considered as genuinely autonomous cognitive systems, might comprise representational states.¹⁰⁵

This might not be an insuperable problem – for instance, Simon J. Evnine¹⁰⁶ has also argued that, since individuals certainly possess mental states, it may be possible to think of persons *as* groups, thereby supplying a model for group minds which includes individual 'nodes' of representation which are focused and organised around

¹⁰⁵ Rupert (2005), p.184.

¹⁰⁶ Evnine (2005).

a central purpose (a Teilhardian ‘Omega point’, perhaps). In fact, the idea of persons as groups may provide an understanding of the split brain phenomena considered earlier at the end of chapter 6: if persons already *are* conglomerates of distinct conscious entities, then the model for a group mind is already there. However, even if this is the case, getting the global (macrocosm) to mirror the personal (microcosm) is no little task.

A further aspect of difficulty is therefore involved in why individuals would seek to undertake such a task. Arguably, we are already driven by evolutionary and biological needs, and so there would have to be good reasons (in these terms) for us to alter or tailor our behaviour in this way – unless, in some way, such developments represented an actual development of those same drives, and were in fact an expression of them. However, Hughes does not (unlike Nietzsche)¹⁰⁷ seem to entertain this latter possibility, but merely advocates a sort of ‘catalogue’ of possible choices which a consumer-like entity or entities could choose from. In this sense, ironically, his contention that the ‘liberal individualistic assumptions’ are under threat is actually undermined by this notion that we can choose to ‘evolve’ in this way; in other words, what lies at the bottom of his various alternative futures is the notion of liberal-individualistic choice.

Conclusion

In this chapter, I have argued that there are various different possible conceptions of ‘death’ and ‘self’ which, given the observer-relative nature of such concepts, are logically possible – if not immediately (or ever) likely. I have also tried to show that the notion of immortality as embodied in the religious notion of the ‘expanded self’ (as I have called it) is not in itself irrational. In fact, such traditions as embody this attitude may simply be advocating a non-essentialist picture of identity and personhood, whilst following it through to more radical yet legitimate conclusions. This, obviously, leaves open a pathway for reassessing religious attitudes to self and death – to what I have termed ‘multiple parallel redefinitions’. I do not want to necessarily argue that my interpretation of religious doctrine and practice in all cases is the ‘correct’ one – and thus commit what I take to have been the Wittgensteinian fallacy – yet I think it fair to argue that, given certain *current and accepted* assumptions of a scientific picture of the world, such ‘expanded’ notions of self are rational and possible.

In addition to the religious picture, I have also considered various science-fiction-type examples of ways in which the notions of self and death can be diversified. However, in doing so, I have also argued for a distinction between logical possibility (on the one hand) and likelihood and achievability (on the other); the first, I argue, reveals the true nature of self and death, whilst the latter involves a sort of ‘leap of faith’ – in fact, both the religious and technological concepts share this aspect (though one for spiritual immortality, the other for physical survival). Finally, therefore, I criticised the idea (found in such Hughes and Leary), that such possible revisions were both imminent and likely.

¹⁰⁷ Though, of course, Nietzsche distinguishes between ‘will to power’ and ‘will to survive’ – see, for instance, Nietzsche (1990a), section 13.

But where does all this leave us in relation to the definition of death? This is the subject of the next and final chapter.

9. Conclusion: Redefinitions of Death

Introduction

Having concluded that the concept of death is an observer-relative one – and is therefore potentially variable in relation to a wide range of factors – it remains for me to briefly outline what possible consequences this will have for the practical determination of death, and to provide some suggested avenues for future consideration.

However before doing so, I will first summarise the arguments which have led me to this point.

Summary of Arguments

Firstly, I point out that the current problem we face in defining death has arisen in relation to advances in medical technology, but such advances have merely revealed a latent problem¹ (which in turn is linked to the nature of concept formation). I have then argued that most suggested resolutions of the debate to date have involved an attempt to isolate sufficient criteria to disambiguate problematic scenarios in favour of one or the other of the traditional binary oppositions ('alive' or 'dead'). In addition, these attempts have mainly taken a strictly biological approach, seeking to define death in terms which are independent of any moral, social, or other value-driven concerns. However, as we have seen, such an approach is problematic.

Firstly, it assumes that the essential features of a living human being can be isolated to the exclusion of such 'observer-relative' features (to use Searle's term), and that a concept of death as an 'intrinsic' feature of the world is therefore possible. However, as Gervais has rightly argued, such a strictly-biological approach mistakenly assumes that all of the essential features of a living human being *are* strictly biological, whereas in fact the attempt to ring-fence one set of criteria (and to exclude another) always involves an extra-biological 'decision of significance': what is significant to one observer may not be so to another, and – more importantly – this variance cannot be resolved in a way which excludes observer-relative features.

This conclusion, I argued, is further supported by an analysis of the nature of the strictly-biological approach, which is essentialist in nature, and therefore open to the well-established criticisms that all forms of essentialism face. So, treating the problem of the definition of death as basically a form of the problem of category membership, I showed that the former concern can be seen to be similarly misguided: there are no means whereby what we mean by 'death' can be fixed, and there would seem to be a potential variance in the nature of the concept which resists the essentialist approach. In other words, what we mean by death is not to be understood by reference to a fixed set of criteria (strictly-biological or otherwise), but is rather an 'epistemically-thick' concept the application of which is much more potentially diverse and changeable than has previously been supposed.

¹ A point independently made by Hughes (2001), p.8.

To further enhance this observation, I next looked at various candidates for death criteria which could in turn be considered 'intrinsic' (borrowing Searle's ontology), and thus which could be used to argue that (a) death consisted of the cessation of certain intrinsic (non-observer-relative) features of the world, and (b) the goal of philosophers in the definition of death debate should be to identify these criteria (thereby 'fixing' the notion of death – in both senses). However, as was shown, even if we adopt a moderate realist position such as Searle's, the main feature which we would seek to base an intrinsic definition of death upon (i.e. function) cannot itself be considered an intrinsic feature. This conclusion represents the death knell of the functional approach, and thereby the strictly-biological approach itself, which is itself function centred. What we are left with, then, is the contention that biological features are, at best, relevant to a definition of death (what I term the 'partly-biological view').

With the above, we reach a point whereby the possibility of a unified definition of death for all organic life becomes extremely problematic, for it is the cessation of functions necessary for life which would unite human death with that of animals, plants, etc. However, with the demise of the notion that function can be a non-observer-relative feature of the world (and thus exist without humans), we are left with – at best – a two-fold definition, involving humans and the remainder of living entities. However, admitting this state of affairs, we need not (it is argued) thereby abandon the search for the set of intrinsic features which determine *human* death. This approach therefore ushers in two further criteriological candidates: the capacity for consciousness and personal identity respectively. Analysis of the former, however, revealed that, whilst it may be possible to consider the capacity of phenomenal consciousness (to use David Chalmers's distinction) as an intrinsic feature of human beings (and even certain animals), there would appear to be *no necessary reason* to identify this capacity as essential to a human being: living human beings certainly exist without it, and to base a definition of death upon its absence would be highly controversial – and moreover, arbitrary. Furthermore, this approach fails to conclusively resolve the problem diagnostically. Capacity for psychological consciousness involves an observer-relative attribution of function, and so cannot play a part in a definition of death involving intrinsic features; however, not only do we face extreme difficulties in distinguishing between those features which do or do not involve conscious participation (phenomenally speaking), we face similar difficulty in being able to ascertain the presence (or absence) of phenomenal consciousness itself. Thus, the criterion fails on a number of fronts.

Notions of personal identity and the capacity for personhood however, do not fare any better. In as much as we would seek to identify a physical basis for personhood, the brain would seem to be the obvious locus; and yet, once again, why should the vehicle for consciousness be the only or central feature of significance? This aside, however, there are a host of other problems highlighted not only by Parfitian thought experiments, but also by findings in so-called 'split-brain' experiments. It is these latter issues which suggest that we do not in fact need science fiction scenarios to show us that the notion of a unified self or 'person' is – arguably – a convenient fiction. The failure of this approach is therefore the final stop in the search for potentially intrinsic features – in relation to *all* forms of death – and with it, we reach the conclusion that death *cannot* be an intrinsic feature of a mind-independent, *human-independent* world.

Following this, and the conclusion that death must be – even according to a realist’s position such as Searle’s – a non-intrinsic feature of the world, I suggested what I take to be more apt models for understanding the concept of death and its potential variability. In doing this, I applied Searle’s notion of constitutive and regulative rules, and showed that, in comparison with such social practices as the game of football, the main difference between football and death is not one of kind, but rather the intransigency which we consider the respective constitutive rules to have. In other words, football may change much more freely than death simply because the constitutive rules which define death practice are informed by deeper and more entrenched values, and are more closely tied in with other ‘deep’ concepts (such as self) which – whilst being similarly observer relative – are similarly resistant to change. Furthermore, it was shown that the type of ambiguity we face in defining death has similar parallels in the development of the rules in football – an ambiguity which can be resolved, but perhaps only if we give up the notion that death is a completely different *type* of debate.

However, whilst such change is slow-moving, this is not to say that it is not possible, or that it is not in fact already taking place. In chapter 8, I therefore set out various alternatives to traditional and accepted notions of death and self in an attempt to show that such alternatives (if interpreted in certain ways) are in fact much more viable and legitimate than may be supposed – or, at least, no more *illegitimate* (logically speaking) than the traditional notions themselves. Finally, however, I strike a note of caution: to argue that the notions of self and death are potentially variable and mutable is still not to argue that such variance and mutability will inevitably come about. So, whilst I argue that, in fact, it *has* come about in certain places (notably in religious contexts), this in turn is linked to deeply embedded changes in lifestyle and habits of thought. In this sense, I reject what I take to be the glib liberal individualist position as represented by Hughes and others. So, in conclusion, whilst we are faced with multiple parallel possible redefinitions of death, the extent to which we must accommodate them is currently relatively limited (though this may not always be so).

Given the above summary, it may be seen that my main concern in this thesis has been with resisting prevalent contemporary approaches to the definition of death which, in my eyes, distort and misrepresent the nature of the problem. They do this, I argue, out of a desire to maintain certain features of the traditional notion of death – though, of course, in as much as they differ, the features which they favour differ also. Most seek to maintain the binary opposition between ‘alive’ and ‘dead’, with no possible intermediate state (which, incidentally, I have shown is an *extra*-logical assumption); most also seek to maintain the notion that death must also be ‘irreversible’ (which, again, the combination of technological advance and the role played by function would seem to make highly problematic); some favour a unified definition, whereby human death is shared with all other living creatures, whilst others admit that an at-least-twofold definition is necessary (both of which would, in the light of subsequent discussion, seem to be highly optimistic); finally, to varying degrees, most commentators seem to hold that death is something that should exist separately from human concerns, and may in fact be an intrinsic feature of a mind-independent world (which has also been shown to be problematic).

However, having sought to resist these various approaches and their attendant metaphysical and ontological assumptions, it now remains to briefly outline some

possible avenues whereby the above conclusions may have some practical benefit for medical practice. Obviously, given my above conclusions concerning the potential variability of death, it would be inappropriate for me now to identify what I take to be the 'correct approach' to defining death. Furthermore, even in suggesting operative parameters for medical practice, I do not have space to do respective issues involved adequate justice – such an undertaking would represent a considerable one in itself, for which the foregoing arguments may only be seen as a sort of basis or 'stepping off' point. What I wish to do, therefore, is merely to outline what I take to be the central concerns for medical practice. It may be objected that, in doing so, my approach has been purely 'negative', and that I have not, as such, 'resolved' the problem of the definition of death. However, if I have performed this negative duty effectively, then I will consider my arguments to have done their job – after all, as the Hippocratic Oath states, 'First, do no harm.' Therefore, in resisting what I take to have been heavy-handed approaches to resolving the debate, and which furthermore I take to have misrepresented the nature of the debate for their own purposes, I am content to have shown these to have been problematic, and thereby to set future debate on a different course.

An Informed Death

To contrast two literary examples, few would share Hamlet's view that death is to be considered a 'consummation devoutly to be wished'; many, however, would agree with Casanova's observation in his voluminous memoirs:

Death is a monster which drives an attentive spectator from the great theater before the play in which he is infinitely interested is over. This alone is reason enough to hate it.²

It is this reluctance to face our inevitable demise, therefore, combined with fear of pain and suffering (and what might lie beyond), that colours the public perception of death. Consequently, the thought that there might be any potential ambiguity in determining death, let alone in defining it, creates understandable anxiety and concern.

One task for the definition of death is therefore to provide unambiguous criteria which not only act as a clear means of determining death in ambiguous cases, but also reassure the public that the decisions involved are capable of standard application to all potential situations. For instance, if we adopt a criterion of death which allows for variance in application of – for instance – concepts of 'viability' or 'quality of life', or that even allow for moral or economic concerns to influence decision making, then this would undoubtedly cause widespread unease. What is needed, therefore, is the agreement of a uniform definition which would not allow for variations of application.

However, given my conclusions concerning the nature and status of the concept of death, is this in fact possible? As I have so far argued, whilst certain situations may seem to reveal no problematic ambiguities as regards the life-status of the individual, there are others which – as we have seen – reveal the determination of death to be a highly-fraught exercise.

² Casanova (1966), vol.1, p.35.

To bring these issues to life (as it were), I will now return to the scenario that I first outlined in my introduction.³ If you recall, I envisaged the case of a young woman who had fallen into a coma and suffered brain anoxia and subsequent brain damage. She was kept alive on a respirator, fed naso-gastrically, and otherwise maintained in a deep coma. After a while, she partially 'revived', and entered a persistent vegetative state, whereby she could breath unaided, and responded to stimuli. However, all tests revealed that she had suffered irreversible damage to the cerebral hemispheres (the 'upper' brain) and would never again resume cognitive function.

It should first of all be noted that it is not intended here that this example represent a *diagnostic* dilemma. That there may be diagnostic problems involved (such as the extent of brain function lost, the possibility of partial recovery) may be true of such a case, but for my purposes we shall consider these to be irrelevant, and the focus to be primarily one of definition. In other words, given that she will never recover, what status do we consider her to have?

As already noted, it is possible to respond to this situation in a number of different ways, and for the response to determine the future course of action. For instance, she may be considered 'dead', in light of the permanent incapacity for cognitive function (and personhood), and allowed to die; she may be considered 'alive', in view of her continued spontaneous function, and her life-support maintained; we may even talk of the death of the 'person' and the survival of the 'organism'. However, as I hope I have shown, all these moves represent value-based assertions as to the nature of a human being, quality of life, etc. Which, then, should represent public policy and accepted medical practice, and how should we decide between them?

In relation to the above situation, therefore, I will argue that we are faced with a number of possible avenues. As I have suggested, I consider these various possibilities to involve what I term 'multiple parallel redefinitions of death'. They are parallel in the sense of having equal status (logically speaking) and in that they embody legitimate but divergent expressions of value. Therefore, since (I argue), a unified definition may be impossible, we are faced with two options: to impose a unified definition; to legislate for diversity. Of course, the latter is the more democratic and liberal move, whilst the former will no doubt be the one favoured by those who hold that the wise must legislate for the less so. However, in reality, a compromise will be the best approach: individual liberty and choice must always be constrained by its effect on society as a whole, and authoritarian prescription will always be challenged by the need for such individual liberty. Accordingly, I will express my suggestions as a brief list of concerns which future debate may centre around.

- (1) Embedded Values. Owing to the prevalent traditional concept of death, medical practice and public policy reflect a similar set of entrenched values to the ones already questioned. In rejecting these, therefore – or at least, their status – practice in both these areas would need to be revised. The long-term goal of such revision would be to more clearly distinguish between factual pronouncements (the patient permanently lacks capacity for cognitive function) and value-based assertions (the patient is 'really' dead).

³ See chapter 1, 'A Hypothetical Case'.

The latter, obviously, should not be assumed by medical practitioners or those who influence and decide public policy. That the patient is ‘really’ dead may be a conclusion that family and friends may arrive at, but in the above scenario, is it not one that could be arrived at by strictly-biological criteria. Such decisions should not therefore form part of a rationale for certain medical decisions. In other words, we should not declare death because we need hospital beds, or organs for transplant, or even because cognitive function is permanently lacking. Consequently, if decisions to remove life support are made for (ultimately) financial or logistical reasons, then this should be made explicit, and the debate shift to its appropriate arena (i.e. funding), and not be masked by shifting the definition of death.

- (2) Expertise and Authority. Similarly, there should be an attempt to make, as far as possible, a clear distinction between those decisions which are in the hands of friends and family (or personal preference – of which more shortly) and those which are determined by medical facts. Out of necessity and habit, the families and friends of the individual will look to the advice of doctors, medical professionals, and the dictates of public policy for advice on what to do. Therefore, such advice should reflect the observer-relative status of death and its consequent potential variability. As far as possible, expertise should advise (and bodies legislate) for the degree of choice which the nature of the concept of death would seem to imply. Therefore, in identifying those situations where individuals have choice as to what is to be done, this should be explicitly stated in terms which make clear to those making the decision that the ‘facts’ can only decide so much for them, and that there are cases where pronouncement of ‘death’ involves a value-based ‘decision of significance’. Obviously, this will necessarily involve a public programme of education – both to inform the public of situations where decisions may have to be made, and in those where public policy takes such decisions out of their hands. Once again, in the latter case, the decision should be clearly depicted for what it is (e.g. motivated by extraneous factors), and not concerning the definition of death itself. Incidentally, this would differ from Gervais’s call for a programme of education in that (a) its purpose would not necessarily be to arrive at a unified definition of death, or to ‘inform’ the public of one already arrived at by experts, (b) the form of contractarianism would be focused on increasing freedom of choice and public awareness in relation to death.
- (3) Death and the Individual. As implied by the previous point, admitting that there is a leeway for interpretation as to what death is, it would make sense to allow individuals to determine exactly what should be their fate should they find themselves in a regrettably ambiguous near-death state. In such cases, a version of the organ donor card might be employed (a ‘Death Card’), or some sort of ‘living will’ which informed those left with the decision-making of the wishes of the individual who cannot now express his wishes. Once again, this would have to be based on informed choice, and there would have to be strict legal guidelines in place to forestall abuse of the process. However, seeing as we are simply talking about decisions which are currently made or influenced by medical professionals and public policy (and should perhaps not be), we are merely here empowering

individual choice in an area where such a thing is most apt (i.e. the individual's own fate).

Obviously, the above points are not exhaustive of the concerns and issues, but they do at least suggest a fruitful area of future research and debate. If I could pick out a main theme, however, it would centre on a clearer distinction between those decisions which are motivated by a consideration of the medical facts, and those which seem to go illegitimately beyond them. Both, I have argued, are value based, but the former admit of a wider commonality of human response, whereas the latter would seem to attempt to force matters.

For instance, we might say that the majority of individuals would agree that a person who had irreversible brain damage to both higher and lower brain, could only breathe with the assistance of a respirator, and whose only signs of life consisted of certain individual 'low-level' processes (such as tissue growth) should be allowed to die. However, to call the woman in our scenario 'dead' and to allow that status to form the basis of switching off life support, seems to me to be an illegitimate move, and an abuse of the concept. By all means, if it is agreed by all parties (especially, perhaps, based on some previously stated wish of the individual herself), then life-support should be withdrawn, but to call such a status 'dead' is to define death not in line with technological inability (that 'we can do nothing more'), but rather in relation to a dubious set of metaphysical and value-based assumptions of which we are barely conscious. Commonly, I feel, death *is* partly defined by technological inability: therefore, as technology advances, so will possible redefinitions of death change and multiply. It is therefore this greater awareness of our implicit assumptions that we should strive for; not so that we can define death 'objectively' – that, arguably, will never be possible – but so that we can approach our final end in an informed way, aware both of what we value in life, and how that in turn must determine how we leave it.

Bibliography

Agich, George J. and Royce P. Jones. 1986. 'Personal Identity and Brain Death: A Critical Response.' *Philosophy and Public Affairs* 15: 267-274.

Alexander, Leo. 1949. 'Medical Science under Dictatorship.' *New England Journal of Medicine* 241: 39-47.

Altham, J. E. J. and Ross Harrison, eds. 1995. *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams*. Cambridge: Cambridge University Press.

Amos, Jonathan. 2005. 'Science Intends to Tag All Life.' BBC News: Science and Nature (February 10, 2005). <http://news.bbc.co.uk/1/hi/sci/tech/4251309.stm> (accessed February 6, 2008).

Anderson, Walter Truett. 1997. *The Future of the Self: Exploring the Post-Identity Society*. New York: Tarcher.

Andrews, Keith, Lesley Murphy, Ros Munday, and Clare Littlewood. 1996. 'Misdiagnosis of the Vegetative State: Retrospective Study in a Rehabilitation Unit.' *British Medical Journal* 313: 13-16.

Andó, Shōei. 1970. *Zen and American Transcendentalism*. Tokyo: Hokuseido Press.

Archer, Margaret S. 2000. *Being Human: The Problem of Agency*. Cambridge: Cambridge University Press.

Aristotle. 1924. *Metaphysics*. Translated by W. D. Ross in 2 vol. Oxford: Clarendon.

Armstrong, D. M. 1973. 'Epistemological Considerations for a Materialist Theory of Mind.' *Philosophy of Science* 40: 178-93.

Armstrong, Karen. 2000. *Buddha*. London: Weidenfeld & Nicolson.

Arnold, John D., Thomas Zimmerman, and Daniel C. Martin. 1968. 'Public Attitudes and the Diagnosis of Death.' *Journal of the American Medical Association* 206: 1949-54.

Aston, Ken. 2008. 'History of the Game Football'. www.kenaston.com.
<http://www.kenaston.org/History/History02.htm> (accessed June 8th 2008).

Ayer, A. J. 1946. *Language, Truth and Logic*. London: Penguin.

Balog, Katalin. 1999. 'Conceivability, Possibility, and the Mind-body Problem.' *The Philosophical Review* 108: 497-528.

Bambrough, Renford F. 1960. 'Universals and Family Resemblances.' *Proceedings of the Aristotelian Society* 40: 207-22.

- Bauby, Jean-Dominique. 1997. *The Diving-Bell and the Butterfly*. London: Harper Perennial.
- Bay, Michael, dir. 2005. *The Island*. Warner Bros.
- Beauchamp, T.L. and S. Perlin, eds. 1978. *Ethical Issues in Death and Dying*. Englewood Cliffs, NJ: Prentice-Hall.
- Beardsmore, R.W. 1992. 'The Theory of Family Resemblances.' *Philosophical Investigations* 15: 131-46.
- Becker, Laurence C. 1975. 'Human Being: The Boundaries of the Concept.' *Philosophy and Public Affairs* 4: 335-59.
- Bedau, Mark. 1992. 'Where's the Good in Teleology?' *Philosophy and Phenomenological Research* 52: 781-806.
- Beecher, Henry K. 1968. 'A Definition of Irreversible Coma. Report of the Ad Hoc Committee of the Harvard Medical School to Examine the Definition of Brain Death.' *Journal of the American Medical Association* 205: 337-40.
- Berger, Peter L., and Thomas Luckman. 1966. *The Social Construction of Reality*. Harmondsworth: Penguin Books.
- Berlin, Brent. 1978. 'Ethnobiological Classification.' In *Cognition and Categorization*, ed. Eleanor Rosch and Barbara B. Lloyd, 9-26. Hillsdale, NJ: Erlbaum.
- Black, Peter M. 1978a. 'Brain Death (First of Two Parts).' *The New England Journal of Medicine* 299: 338-44.
- Black, Peter M. 1978b. 'Brain Death (Second of Two Parts).' *The New England Journal of Medicine* 299: 393-401.
- Black, Peter M. 1978c. 'Definitions of Brain Death.' In *Ethical Issues in Death and Dying*, ed. T.L. Beauchamp and S. Perlin, 5-10. Englewood Cliffs, NJ: Prentice-Hall.
- Blackmore, Susan. 2003. *Consciousness: An Introduction*. London: Hodder & Stoughton.
- Block, Ned. 1981. 'Psychologism and Behaviourism.' *The Philosophical Review* 90: 5-43.
- Bondeson, Jan. 2001. *Buried Alive: The Terrifying History of Our Most Primal Fear*. New York: W.W. Norton & Co.
- Brante, Thomas, and Margareta Hallberg. 1991. 'Brain or Heart? The Controversy over the Concept of Death.' *Social Studies of Science* 21: 389-413.
- Brett, Matthew, Ingrid S. Johnsrude, and Adrian M. Owen. 2002. 'The Problem of Functional Localization in the Human Brain.' *Nature Reviews: Neuroscience* 3: 243-9.

- Brody, Howard. 1981. *Ethical Decisions in Medicine*, 2nd edn. Boston: Little, Brown.
- Brody, Howard. 1983. 'Brain Death and Personal Existence: A Reply to Green and Wikler.' *The Journal of Medicine and Philosophy* 8: 187-96.
- Brooks, Rodney A. 2002. *Robot: The Future of Flesh and Machines*. London: Penguin.
- Browne, Alister. 1983. 'Whole-Brain Death Reconsidered.' *Journal of Medical Ethics* 9: 28-31.
- Burgess, Anthony. 1986. *Little Wilson and Big God*. London: Vintage.
- Burton, Tim, dir. 1999. *Sleepy Hollow*. Paramount.
- Byrne, P., S. O'Reilly, and P. M. Quay. 1979. 'Brain Death: An Opposing Viewpoint.' *Journal of the American Medical Association* 242: 1985-90.
- Capron, Alexander Morgan. 1978. 'A Statutory Definition of the Standards for Determining Human Death: An Appraisal and a Proposal.' In *Ethical Issues in Death and Dying*, ed. T.L. Beauchamp and S. Perlin, 60-73. Englewood Cliffs, NJ: Prentice-Hall.
- Carter, Rita. 1998. *Mapping the Mind*. London: Weidenfeld & Nicolson.
- Carter, Rita. 2002. *Consciousness*. London: Weidenfeld & Nicolson.
- Casanova, Giacomo. 1966. *History of My Life: Volumes 1 & 2*. Trans. Willard R. Task. Baltimore: The Johns Hopkins University Press.
- Chalmers, Alan. 1999. *What is this Thing Called Science?* 3rd edn. Buckingham: Open University Press.
- Chalmers, David. 1996. *The Conscious Mind*. Oxford: Oxford University Press.
- Chatterjee, Satischandra and Dhirendramohan Datta. 1960. *An Introduction to Indian Philosophy*. 6th edn. Calcutta: University of Calcutta.
- Churchland, Paul. 1981. 'Eliminative Materialism and the Propositional Attitudes.' *Journal of Philosophy* 78: 67-90.
- Cockburn, David. 2001. *An Introduction to the Philosophy of Mind*. Basingstoke: Palgrave Macmillan.
- Conze, Edward, ed. 1959. *Buddhist Scriptures*. London: Penguin Books.
- Cox, Gary. 1999. 'Heidegger and Sartre on Death', *Cogito* 13:171-5.

- Cranford, R. E. 1978. 'Minnesota Medical Association criteria. Brain Death. Concept and Criteria.' *Minnesota Medicine* 61: 561-3.
- Crick, Francis. 1994. *The Astonishing Hypothesis*. Sydney: Simon & Schuster.
- Daly, Chris. 2000. 'Natural Kinds.' In *The Concise Routledge Encyclopedia of Philosophy*, 612-3. London: Routledge.
- Davidson, Donald. 1984. 'On the Very Idea of a Conceptual Scheme.' In *Inquiries into Truth and Interpretation*, 183-198. Oxford: Oxford University Press.
- Dawkins, Richard. 1986. London: Penguin.
- Dennett, Daniel. 1971. 'Intentional Systems.' *The Journal of Philosophy* 68: 87-106.
- Dennett, Daniel. 1991. *Consciousness Explained*. London: Penguin.
- Department of Health. 1998. 'A Code of Practice for the Diagnosis of Brain Stem Death.'
- Descartes, René. 1986. *A Discourse on Method, Meditations and Principles*. Translated by John Veitch. London: J. M. Dent & Sons.
- Detlefsen, Michael. 2000. 'Gödel's Theorems.' In *The Concise Routledge Encyclopedia of Philosophy*, 319. London: Routledge.
- Diamond, Cora. 1991. *The Realistic Spirit*. Cambridge, MA: MIT Press.
- Dick, Phillip K. 2000. 'We Can Remember It for You Wholesale.' In *We Can Remember It for You Wholesale: Vol. 5 of Collected Stories, 157-74*. London: Orion Books.
- Doepke, Frederick. 1992. 'Identity and Natural Kinds.' *The Philosophical Quarterly* 42: 89-94.
- Donne, John. 1999. *Devotions upon Emergent Occasions and Death's Duel*. London: Vintage.
- Eccles, J. C. 1967. 'Evolution and the Conscious Self.' In J. D. Roslansky and J. C. Eccles, eds. *The Human Mind*, 11. Amsterdam: North Holland Pub. Co.
- Eccles, J. C. 1970. *Facing Reality: Philosophical Adventures by a Brain Scientist*. New York: Springer-Verlag.
- Evans, D. W. and D. J. Hill. 1989. 'The Brainstems of Organ Donors are not Dead.' *Catholic Medical Quarterly* 40: 113-21.
- Evans, Martyn. 1990. 'A Plea for the Heart.' *Bioethics* 4: 227-31.

- Evans, Martyn. 1994. 'Against the Definition of Brainstem Death.' In *Death Rites: Law and Ethics at the End of Life*, 1-10. London: Routledge.
- Evnine, Simon J. 2005. 'Containing Multitudes: Reflection, Expertise and Persons as Groups.' *Episteme* 2: 57-64.
- Flew, A. G. N. 1964. 'Theology and Falsification.' In John Hick, ed. *The Existence of God*, 225-8. London: Collier Macmillan.
- Flew, A. G. N. 1955. 'Death.' In A. MacIntyre and A. G. N. Flew, eds. *New Essays in Philosophical Theology*, 267-72. London: SCM Press.
- Gardener, Sebastian. 1999. *Kant and the Critique of Pure Reason*. London: Routledge.
- Gardner, Helen, ed. 1972. *The Metaphysical Poets*. London: Penguin Books.
- Gaylin, Waylard. 1974. 'Harvesting the Dead.' *Harper's Magazine*. September: 23-30.
- Gazzaniga, M. S., J. E. Bogen, and R. W. Sperry, 1962. 'Some Functional Effects of Sectioning the Cerebral Commissures in Man.' *Proceedings of the National Academy of Sciences* 48: 1765-9.
- Geach, Peter. 1972. 'The Fallacy of *Cogito Ergo Sum*'. In Oswald Hanfling, ed. *Fundamental Problems in Philosophy*, 59-61. Oxford: Blackwell.
- Gellner, Ernest. 1959. *Words and Things*. Harmondsworth: Penguin Books.
- Gervais, Karen. 1986. *Redefining Death*. London: Yale University Press.
- Giddens, Anthony. 1986. *The Constitution of Society: Outline of the Theory of Structuration*. California: University of California Press.
- Glaser, Paul Michael, dir. 1987. *The Running Man*. USA: Tristar Pictures.
- Goldman, William. 1995. *William Goldman: Four Screenplays with Essays*. New York: Applause Books.
- Goodman, Nelson. 1983. *Fact, Fiction, and Forecast*, 4th edn. Cambridge, MA: Harvard University Press.
- Goodman, Russell B. 2002. *Wittgenstein and William James*. Cambridge: Cambridge University Press.
- Green, Michael, and Daniel Wikler. 1980. 'Brain Death and Personal Identity.' *Philosophy and Public Affairs* 9: 105-133.
- Hacking, Ian. 1999. *The Social Construction of What?* Cambridge, MA: Harvard University Press.

- Heil, John. 1998. *Philosophy of Mind: A Contemporary Introduction*. London: Routledge.
- Hick, John. 1990. *Philosophy of Religion*. London: Prentice-Hall.
- Hoffman, John C. 1979. 'Clarifying the Debate on Death.' *Soundings* 62: 430-47.
- Honderich, T. 1981. 'Philosophical Law-like Connections and their Problems.' *Inquiry* 24: 277-303.
- Hughes, James J. 2001. 'The Future of Death: Cryonics and the Telos of Liberal Individualism.' *Journal of Evolution and Technology* 6.
<http://www.jetpress.org/volume6/death.htm> (accessed on 8th June 2008).
- Humberstone, Lloyd. 2000. 'Many-Valued Logics, Philosophical Issues In.' In *The Concise Routledge Encyclopedia of Philosophy*, 612-3. London: Routledge.
- Hume, David. 1966. *Dialogues Concerning Natural Religion*. Norman Kemp Smith, ed. New York: Bobbs-Merroll.
- Hume, David. 1967. *A Treatise of Human Nature*. Oxford: Oxford University Press.
- Hume, David. 1975. *Enquiries Concerning Human Understanding And Concerning the Principles of Morals*. L. A. Selby-Bigge and P. H. Nidditch, eds. Oxford: Oxford University Press.
- Inwagen, Peter van. 1997. 'Materialism and the Psychological-Continuity Account of Personal Identity.' *Noûs* 31: 305-319.
- Jackson, F. 1982. 'Epiphenomenal Qualia.' *Philosophical Quarterly* 32: 127-36.
- Janik, Allan. 1985. *Essays on Wittgenstein and Weininger*. Amsterdam: Rodolpi.
- Jaynes, Julian. 1976. *The Origin of Consciousness in the Breakdown of the Bicameral Mind*. London: Allen Lane.
- Jewison, Norman, dir. 1975. *Rollerball*. USA: United Artists.
- Jonas, H. 'Against the Stream: Comments on the Definition and Re-definition of Death.' In T.L. Beauchamp and S. Perlin, eds. In *Ethical Issues in Death and Dying*, 51-60. Englewood Cliffs, NJ: Prentice Hall.
- Kaufman, Sharon R. 2000. 'In the Shadow of "Death with Dignity": Medicine and Cultural Quandaries of the Vegetative State.' *American Anthropologist* 102: 69-83.
- Kluge, Eike-Henner W. 1984. 'Cerebral Death.' *Theoretical Medicine and Bioethics* 5: 209-31.
- Krauss, Lawrence M. 2001. *Atom*. London: Little, Brown and Company.

- Kripke, Saul A. 1982. *Wittgenstein on Rules and Private Language: An Elementary Exposition*. Oxford: Blackwell.
- Ladyman, James. 2002. *Understanding Philosophy of Science*. Oxford: Routledge.
- Lamb, David. 1978. 'Diagnosing Death.' *Philosophy and Public Affairs* 7: 144-53.
- Lamb, David. 1985. *Death, Brain Death and Ethics*. London: Croom Helm.
- Lampert, Laurence. 2001. *Nietzsche's Task*. New Haven: Yale University Press.
- Leary, Timothy, with R. U. Sirius. 1997. *Design for Dying*. San Francisco: Harper Collins.
- Leibniz, Gottfried Wilhelm. 1973. *Philosophical Writings*. Edited by G. H. R. Parkinson, translated by Mary Morris and G. H. R. Parkinson. London: J. M. Dent & Sons.
- Leidig, Michael. 2003. 'Six Years in a Coma then Christiane Wakes...at a Bryan Adams Gig.' *Daily Mirror*, 13 March 2003, 23.
- Leonard, Brett, dir. 1992. *The Lawnmower Man*. USA: Allied Vision Ltd.
- Lever, Anna-Marie. 2008. 'DNA "Barcode" Revealed in Plants.' *BBC News: Science and Nature*, 6 February 2008. <http://news.bbc.co.uk/1/hi/sci/tech/7221946.stm> (Accessed on 6 February 2008).
- Lisberger, Steven, dir. 1982. *Tron*. USA: Walt Disney.
- Lock, Margaret. 1998. 'Deadly Disputes: Hybrid Selves and the Calculation of Death in Japan and North America.' *Osiris* 13: 410-29.
- Lock, Margaret. 1995. 'Transcending Mortality: Organ Transplants and the Practice of Contradictions.' *Medical Anthropology Quarterly* 9: 390-3.
- Lock, Margaret and Christina Honde. 1990. 'Reaching Consensus about Death: Transplants and Cultural Identity in Japan.' In George Weisz, ed. *Social Science Perspectives on Medical Ethics*, 99-119. Dordrecht: Kluwer Academic Publishers.
- Locke, John. 1997. *An Essay Concerning Human Understanding*. Roger Woolhouse, ed. London: Penguin Classics.
- Lormand, Eric. 1996. 'Nonphenomenal Consciousness.' *Noûs* 30: 242-61.
- Luper-Foy, Stephen. 1987. 'Annihilation.' *The Philosophical Quarterly* 37: 233-52.
- Lynn, J. 1981. 'Guidelines for the determination of death. Report of the medical consultants on the diagnosis of death to the President's Commission for the study of ethical problems in medicine and biomedical and behavioural research.' *Journal of the American Medical Association* 246: 2184-6.

- Machado, Calixto, Julius Kerein, Yazmina Ferrer, Liana Portela, Maria de la C. García, and José M. Manero. 2007. 'The concept of brain death did not evolve to benefit organ transplants.' *Journal of Medical Ethics* 33: 197-200.
- MacKay, D. M. 1966a. 'Cerebral Control and the Conscious Control of Action.' In J. C. Eccles, ed. *Brain and Conscious Experience*, 422-44. New York: Springer-Verleg.
- MacKay, D. M. 1966b. 'Discussion' in J. C. Eccles, ed. *Brain and Conscious Experience*, 312-3. New York: Springer-Verleg.
- MacKinnon, D. M. 1955. 'Death.' In A. MacIntyre and A.G.N. Flew, eds. *New Essays in Philosophical Theology*, 261-6. London: SCM Press.
- McManus, Denis. 1995. 'The Mysterious Appeal of "Wittgenstein's Conservatism".' *Wittgenstein Studien* 2/2. <http://www.phil.uni-passau.de/dlwg/ws04/16-2-95.txt> (accessed 8th June 2008).
- Mascaro, Juan, translator. 1962. *Bhagavad Gita*. London: Penguin Books.
- Maslin, K. T. 2001. *An Introduction to the Philosophy of Mind*. Cambridge: Polity Press.
- Mathiesen, Kay. 2005. 'The Epistemic Features of Group Belief.' *Episteme* 2: 161-75.
- Matthews, Gareth B. 1979. 'Life and Death as the Arrival and Departure of the Psyche.' *American Philosophical Quarterly* 16: 151-7.
- McCutcheon, Felicity. 2001. *Religion Within the Limits of Language Alone: Wittgenstein on Philosophy and Religion*. Aldershot: Ashgate Publishing.
- Metcalf, Fred, ed. 1986. *The Penguin Dictionary of Modern Humorous Quotations*. London: Penguin.
- Millikan, R. G. 1984. *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge, MA: MIT Press.
- Millikan, R. G. 1989. 'In Defense of Proper Functions.' *The Philosophy of Science* 56: 288-302.
- Mollaret, P. and M. Goulon. 1959. 'Le coma dépassé mémoire préliminaire.' *Revue Neurologique (Paris)* 101: 3-15.
- Monk, Ray. 1991. *Ludwig Wittgenstein: The Duty of Genius*. London: Vintage Books.
- Nagel, Thomas. 1970. 'Armstrong on the Mind.' *The Philosophical Review* 79: 394-403.
- Nagel, Thomas. 1974. 'What is it like to be a bat?' *The Philosophical Review* 83: 435-50.

- Nagel, Thomas. 1979a. 'Death.' Reprinted in *Mortal Questions*, 1-10. Cambridge: Cambridge University Press.
- Nagel, Thomas. 1979b. 'Moral Luck.' Reprinted in *Mortal Questions*, 24-38. Cambridge: Cambridge University Press.
- Nagel, Thomas. 1979c. 'Brain Bisection and the Unity of Consciousness.' Reprinted in *Mortal Questions*, 147-64. Cambridge: Cambridge University Press.
- New Jersey Supreme Court. 1975-6. *In the Matter of Karen Quinlan*, 2nd edn., 2 vols. Arlington, VA: University Publications of America.
- Nietzsche, Friedrich. 1990a. *Beyond Good and Evil*. Translated by R. J. Hollingdale. London: Penguin.
- Nietzsche, Friedrich. 1990b. *Twilight of the Idols/The Anti-Christ*. Translated by R. J. Hollingdale. London: Penguin.
- Nietzsche, Friedrich. 1968. *The Will to Power*. Translated by Walter Kaufmann and R. J. Hollingdale. New York: Random House.
- Nietzsche, Friedrich. 1965. *The Philosophy of Nietzsche*. Edited and translated by Geoffrey Clive. New York: Mentor Books.
- Nolan, Christopher. 1981. *Dam-Burst of Dreams*. London: Pan Books.
- Nolan, Christopher. 1987. *Under the Eye of the Clock*. London: Pan Books.
- Norris, Christopher. 2003. 'The Limits of Whose Language? Wittgenstein on the Philosophy of Science, Mathematics, and Logic.' *Pretexts: Literary and Cultural Studies* 12: 5 – 42.
- Nyíri, J. C. 1982. 'Wittgenstein's Later Work in Relation to Conservatism.' In B. F. McGuinness, ed. *Wittgenstein and His Times*, 44-68. Oxford: Basil Blackwell.
- Olson, Eric T. 1997. 'Was I Ever a Fetus?' *Philosophy and Phenomenological Research* 57: 95-110.
- Page, Sam. 2004. 'Searle's Realism Deconstructed.' *The Philosophical Forum* 35: 249-74.
- Pallis, Christopher. 1983a. 'Whole-Brain Death Reconsidered – Physiological Facts and Philosophy.' *Journal of Medical Ethics* 9: 32-7.
- Pallis, Christopher. 1983b. *ABC of Brainstem Death*. London: British Medical Association.
- Pallis, Christopher. 1985. 'Defining Death.' *British Medical Journal* 291: 666-7.

- Papineau, D. 1993. *Philosophical Naturalism*. Oxford: Blackwell.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Phillips, D. Z. 1970. *Death and Immortality*. London: Macmillan & Co Ltd.
- Phillips, D. Z. 1971. 'Religious Beliefs and Language Games.' In Basil Mitchell, ed. *The Philosophy of Religion*, 121-42. Oxford: Oxford University Press.
- Pitcher, G., ed. 1966. *Wittgenstein: The 'Philosophical Investigations'*. New York: Doubleday Anchor.
- Plato. 1987. *The Republic*. London: Penguin Books.
- Plato. 1969. *The Last Days of Socrates*. London: Penguin Books.
- Poe, Edgar Allan. 1986. *The Fall of the House of Usher and Other Stories*. London: Marshall Cavendish.
- Priest, Stephen. 1991. *Theories of the Mind*. London: Penguin.
- Puccetti, Roland. 1973a. 'Brain Bisection and Personal Identity.' *The British Journal for the Philosophy of Science* 24: 339-55.
- Puccetti, Roland. 1973b. 'Multiple Identity.' *The Personalist*, 54: 203-15.
- Puccetti, Roland. 1976. 'The Conquest of Death.' *The Monist* 59: 249-63.
- Putnam, Hilary. 2002. *The Collapse of the Fact/Value Dichotomy and Other Essays*. Cambridge, MA: Harvard University Press.
- Quine, W. V. O. 1961. 'Two Dogmas of Empiricism.' In *From a Logical Point of View*, 20-46. Cambridge, MA: Harvard University Press.
- Quine, W. V. O. 1969a. 'Ontological Relativity.' *Ontological Relativity and Other Essays*, 26-68. London: Columbia University Press.
- Quine, W. V. O. 1969b. 'Natural Kinds.' In *Ontological Relativity and Other Essays*, 114-38. London: Columbia University Press.
- Ramsey, Paul. 1970. *The Patient as Person*. New Haven: Yale University Press.
- Reiner, Carl, dir. 1984. *All of Me*. USA: Universal Pictures.
- Roddenberry, Gene, creator. 1987-94. *Star Trek: The Next Generation* USA: Paramount Television.
- Rorty, Richard. 1997. 'Realism/Antirealism, and Pragmatism.' In *Realism, Antirealism and Epistemology*, 160. Edited by Christopher B. Kulp. New York: Rowman & Littlefield Publishers, Inc.

- Rosch, Eleanor. 1978. 'Principles of Categorization.' In Eleanor Rosch and Barbara B. Lloyd, eds. *Cognition and Categorization*, 27-48. Hillsdale, N. J.: Erlbaum, 1978.
- Rosenbaum, Stephen. 1989. 'Epicurus and Annihilation.' *The Philosophical Quarterly* 39: 81-90.
- Rowlands, Mark. 2003. *The Philosopher at the End of the Universe*. London: Random House.
- Ruben, David-Hillel. 1997. 'John Searle's *The Construction of Social Reality*.' *Philosophy and Phenomenological Research* 57: 443-7.
- Rupert, Robert. 2004. 'Minding One's Cognitive Systems: When Does a Group of Minds Constitute a Single Cognitive Unit?' *Episteme* 1: 177-88.
- Russell, Bertrand, and A. N. Whitehead. 1903. *Principia Ethica*. Cambridge: Cambridge University Press.
- Russell, Bertrand. 1903. 'On Denoting', *Mind* 14: 479-493. Reprinted in Russell, Bertrand. 1973. *Essays in Analysis*, 103-19. London: Allen and Unwin.
- Russell, Tom. 2000. *Brain Death: Philosophical Concepts and Problems*. Hants: Ashgate.
- Ryle, Gilbert. 1949. *The Concept of Mind*. London: Penguin.
- Saunders, Doug. 2002. 'Is dying necessary? Meet the optimists of the Life Extension Movement, who think 'natural death' is a contradiction in terms.' *The Globe and Mail*, 30 November 2002.
- Sardar, Ziauddin. 2000. *Thomas Kuhn and the Science Wars*. London: Icon Books.
- Scarre, Geoffrey. 2007. *Death*. Stocksfield: Acumen.
- Schacht, Richard. 1983. *Nietzsche*. London: Routledge & Kegan Paul.
- Schrodinger, Erwin. 1944. *What is Life?* Cambridge: Cambridge University Press.
- Schultz, Walter. 2000. 'On the Problem of Death.' *Continental Philosophy Review* 33: 467-86.
- Schwager, Robert L. 1978. 'Life, Death, and the Irreversibly Comatose.' In *Ethical Issues in Death and Dying*, 38-50. Englewood Cliffs, N. J.: Prentice-Hall.
- Searle, John R. 1964. 'How to Derive "Ought" From "Is".' *Philosophical Review* 73: 43-58.
- Searle, John R. 1992. *The Rediscovery of Mind*. Cambridge, MA: MIT Press.

- Searle, John R. 1995. *The Construction of Social Reality*. London: Penguin.
- Searle, John R. 1997. *The Mystery of Consciousness*. London: Granta.
- Searle, John R. 2004. 'Realism Reconstructed: A Reply.' *The Philosophical Forum* 35: 275-80.
- Sedgwick, Peter R. 1995. *Nietzsche: A Critical Reader*. Oxford: Blackwell.
- Sedgwick, Peter R., ed. 2001. *Descartes to Derrida*. Oxford: Blackwell.
- Sellars, Wilfrid. 1997. *Empiricism and the Philosophy of Mind*. Cambridge, MA: Harvard University Press.
- Shakespeare, William. 1963. *Shakespeare's Sonnets*. Edited by Martin Seymour-Smith. London: Heinemann.
- Shakespeare, William. 1972. *King Lear*. Edited by G. K. Hunter. London: Penguin Books.
- Shakespeare, William. 1982. *Hamlet*. Edited by Harold Jenkins. London: Routledge.
- Shrader, Douglas. 1986. 'On Dying More Than One Death.' *The Hastings Center Report* 16: 12-17.
- Singer, Peter. 1993. *Practical Ethics*. Cambridge: Cambridge University Press.
- Sokal, Alan. 1996. 'Transgressing the Boundaries: Towards a Transformative Hermeneutics of Quantum Gravity.' *Social Text* 46: 217-52.
- Solomon, Robert C. and Kathleen M. Higgins. 2000. *What Nietzsche Really Said*. New York: Schocken Books.
- Spottiswoode, Roger, dir. 2000. *The Sixth Day*. USA: Phoenix Pictures.
- Sprigge, T. L. S. 1994. 'Consciousness.' *Synthese* 98: 73-93.
- Stanilard, Hilary. 1972. *Universals*. New York: Anchor Books.
- Stich, S. P. 1983. *From Folk Psychology to Cognitive Science: The Case Against Belief*. Cambridge, MA: MIT Press.
- Stoker, Bram. 1993. *Dracula*. Hertfordshire: Wordsworth Classics.
- Thomson, Judith Jarvis. 1992. 'In Defence of Abortion.' In *Applied Ethics*. Edited by Peter Singer. Oxford: Oxford University Press.
- Tooley, Michael. 1992. 'Abortion and Infanticide.' In *Applied Ethics*. Edited by Peter Singer. Oxford: Oxford University Press.

- Tye, M. 1986. 'The Subjective Qualities of Experience.' *Mind* 95: 1-17.
- Van Doren, Mark, ed. 1973. *The Portable Walt Whitman*. London: Penguin Books.
- Veatch, R. M. 1975. 'The Whole-Brain-Oriented Concept of Death: An Outmoded Philosophical Formulation.' *Journal of Thanatology* 3: 13-30.
- Veatch, R. M. 1986. 'Defining Death Anew: Technical and Ethical Problems.' In Robert F. Weir, ed. *Ethical Issues in Death and Dying*, 59-79. New York: Columbia University Press.
- Verhoeven, Paul, dir. 1990. *Total Recall*. Carolco International N. V.
- Walton, Douglas N. 1987. *Ethics of Withdrawal of Life-Support Systems*. London: Praeger.
- Wang, Guoqi. 2002. 'Habeas Corpus.' *Harper's Magazine*, February: 22-5.
- Weil, Simone. 1951. *Gravity and Grace*. London: Routledge & Kegan Paul.
- Weil, Simone. 1959. 'Concerning the "Our Father".' In *Waiting on God*, 166-77. London: Fontana Books.
- Whedon, Joss, creator. 1997-2003. *Buffy the Vampire Slayer: The Series*. 20th Century Fox Television.
- Wheeler, Samuel C., III. 1988. 'Wittgenstein as Conservative Deconstructor.' *New Literary History* 19: 239-58.
- Whorf, B. L. 1956a. 'The Punctual and Segmentative Aspects of Verbs in Hopi.' In John B. Carroll, ed. *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, 51-6. Cambridge, MA: Massachusetts Institute of Technology Press.
- Whorf, B. L. 1956b. 'Science and Linguistics.' In John B. Carroll, ed. *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, 207-20. Cambridge, MA: Massachusetts Institute of Technology Press.
- Wiggins, David. 1980. *Sameness and Substance*. Oxford: Blackwell.
- Wildiers, N. M. 1968. *An Introduction to Teilhard de Chardin*. Translated by Hubert Hoskins. London: Collins.
- Williams, Bernard. 1981. *Moral Luck*. Cambridge: Cambridge University Press.
- Williams, Bernard. 1995. 'Replies.' In Altham, J. E. J. and Ross and Harrison, eds. *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams*. Cambridge: Cambridge University Press.

Wittgenstein, Ludwig. 1967. *Philosophical Investigations*. G. E. M. Anscombe, R. Rhees, and G. H. von Wright, eds; translated by G. E. M. Anscombe, 3rd edn. Oxford: Basil Blackwell.

Wittgenstein, Ludwig. 1974. *On Certainty*. G. E. M. Anscombe and G. H. von Wright, eds; translated by Denis Paul and G. E. M. Anscombe. Oxford: Blackwell.

Wittgenstein, Ludwig. 1980. *Culture and Value*. G.H. Von Wright, ed.; translated by Peter Winch, 2nd edn. Chicago: University of Chicago Press.

Wright, Larry. 1973. 'Functions.' *The Philosophical Review* 82: 139-168.

Yablo, Stephen. 2000. 'Essentialism.' In *The Concise Routledge Encyclopedia of Philosophy*, 254. London: Routledge.

Youngren, William H. 1980. 'Conceptualism and Neoclassical Generality.' *ELH* 47:705-40.