Analytic Functionalism as a Foundation for the Contention that a Non-Biological Machine (Android) can be Viewed as Both a Legal and a Moral Person

by

David Calverley

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved March 2011 by the
Graduate Supervisory Committee:

Brad Armendt, Chair
Andrew Askland
Joan McGregor

ARIZONA STATE UNIVERSITY

May 2011

ABSTRACT

This Thesis contends that if the designer of a non-biological machine (android) can establish that the machine exhibits certain specified behaviors or characteristics, then there is no principled reason to deny that the machine can be considered a legal person. The thesis also states that given a related but not necessarily identical set of characteristics, there is no principled reason to deny that the non-biological machine can make a claim to a level of moral personhood. It is the purpose of my analysis to delineate some of the specified behaviors required for each of these conditions so as to provide guidance and understanding to designers seeking to establish criteria for creation of such machines. Implicit in the stated thesis are assumptions concerning what is meant by a non-biological machine. I use analytic functionalism as a mechanism to establish a framework within which to operate. In order to develop this framework it is necessary to provide an analysis of what currently constitutes the attributes of a legal person, and to likewise examine what are the roots of the claim to moral personhood. This analysis consists of a treatment of the concept of legal personhood starting with the Greek and Roman views and tracing the line of development through the modern era. This examination then explores at a more abstract level what it means to be a person. Next, I examine law's role as a normative system, placing it within the context of the previous discussions. Then, criteria such as autonomy and intentionality are discussed in detail and are related to the over all analysis of the thesis. Following this, moral personhood is examined using the animal rights

movement of the last thirty years as an argument by analogy to the question posed by the thesis. Finally, all of the above concepts are combined in a way that will provide a basis for analyzing and testing future assertions that a non-biological entity has a plausible claim for legal or moral personhood. If such an entity exhibits the type of intentionality and autonomy which humans view as the foundation of practical reason, in combination with other indicia of sentience described by "folk psychology", analytic functionalism suggests that there is no principled reason to deny the android's claim to rights.

DEDICATION

To Bentley, my wife, without her, this adventure would have been impossible.

My love and gratitude always

ACKNOWLEDGMENTS

This Thesis is based in large part on papers and presentations made over the last six years. It has benefitted from the input of the participants at the many conferences I have attended as well as the work of anonymous editors and reviewers who commented on written papers. I thank then for their time, efforts, and insights. With the exception of Chapter 5, the bulk of this work has appeared in previously published works. Permission to use copyrighted work is gratefully acknowledged. A complete list of previously published work is contained in the References section.

TABLE OF CONTENTS

Page

PREFACE...... vii

CHAPTER

1 RIGHTS TALK.. 1

2 FUNCTIONALISM AND ITS ANTECEDENTS............................... 5

3 HUMANS, PERSONS & PROPERTY... 16

4 LAW AS A NORMATIVE SYSTEM 24

5 REASON AS A BASIS FOR LAW AND ETHICS, A CLASSICAL

. VIEW 30

6 LAW AND PRACTICAL REASON 49

7 LAW AND FOLK PSYCHOLOGY 53

7a. Intentionality..54

7b. Autonomy and Responsibility..61

7c. An Ethical Aside ……………………………………..……. 72

8 ANIMAL RIGHTS 79

9 CONCLUSION 87

REFERENCES 90

PREFACE

The principal thesis of this presentation is that at least one reading of analytic functionalism can support the conclusion that a non-biological machine can be found to be a legal person for purposes of the 14th Amendment to the United States Constitution. Furthermore, drawing an analogy between animal rights and rights for androids, it is suggested that an even lower threshold can be established for the purpose of ascribing moral rights. In order to establish this thesis, I first lay out a description of analytic functionalism which supports the view that a non-biological machine can exhibit the same functional roles as humans. The theory of mind posited by analytic functionalism is premised on the common-sense platitudes of folk psychology. Consequently, to explore my thesis I examine the suggestion that these platitudes may be carried over into the way we define a legal person. We need to be mindful, however, of the fact that folk psychology is susceptible to two interpretations, an external view and an internal one. I argue that "folk psychology internal", as I refer to it, and law are based on a similar understanding of the way humans predict and evaluate their own behavior and that of others. I further suggest that law is a pragmatic system of rules developed by society to regulate its members. I then show that this system of rules is based upon the belief that the objects of its attention, *legal persons,* are governed by practical reason. The similarity between "folk psychology internal" and law is then made explicit by reference to two particular concepts which are used by both; *autonomy* and *intentionality*. However, to be secure in our position

we need to also meet the direct challenge to folk psychology posed by eliminativist materialism. This can be done by "folk psychology internal" on the reading I adopt. Other criticisms of analytic functionalism are mentioned but are addressed only in passing because they do not block the argument I make. From the foundation created by this series of positions, I then argue that under a reading of "folk psychology internal", once a non-biological machine meets criteria such as autonomy and intentionality, there is no theoretical reason to deny a claim that it possesses sufficient attributes of mental states such that it can be deemed to be a *legal person* for Constitutional purposes. Next I examine the animal rights movement as it has developed over the last thirty years and draw an analogy between animals and androids. Based on this analogy I suggest that analytic functionalism allows for the ascribing of moral rights to androids which operate at a level at least as complex as that of animals. Finally, if this is so, then a cautionary word is added which could affect future development in this area of endeavor.

On a regular basis the general media carries reports of ever increasingly sophisticated activities being performed by robots or other types of machines. (Naik, 2009). Robots are being trained to serve food and drinks, tend to infirm patients and any number of other tasks. (Daly, 2010). In Japan, Hiroshi Isiguro has designed androids that look like real live people. (Guizzo, 2010). Robots are becoming so ubiquitous that the Japanese government has established a committee

to draw up safety guidelines for the keeping of robots in homes and offices. (Faiola, 2005). What issues might arise if an android reached a stage at which the public perceives that the android should be subject to some form of legal constraint and entitled to some form of legal protection? (Inayatullah, 2001). It seems likely that at some point a political subdivision capable of making and enforcing laws will try to apply them directly to an android rather than just to the designer or builder. (McNally and Inayatullah, (n.d.)). In examining this possibility it is necessary to address some moral and ethical issues that could arise from the creation of such an android. Here, the term 'android' is used interchangeably with the term 'non-biological machine', both of which are intended to capture the sense developed by Ishiguro (2006), that is, to refer to very humanlike robots that can give us insights into human behavior and cognition (MacDorman and Ishiguro, 2006). By analyzing the animal rights movement we can draw lessons that could be applied to androids as they develop from crude machines to something closer to 'human'. The arguments for animal rights developed over the last 30 years have substantially changed the views of many. Perhaps we can determine whether those ideas can shed light on how best to design androids (Mazlish, 1993)

Chapter 1

RIGHTS TALK

> There are many other ways of trying to analyze the various kinds of situations that arise in our lives when an action or policy is good or bad, right or wrong, moral or immoral. But in the Western world at the present time, rights talk is one of the most common ways of formulating moral issues, whether the issues involve only individuals or are between individuals and communities or even when they involve the relationships of whole communities with one another. Werhane, 1985. (p. 4)

Although rights talk is fraught with confusion, and occasionally intentional distortion for partisan reasons, it is still the 'normal currency of ordinary political discourse' (Waldron, 1988). In discussing rights, there is an often-ignored distinction between what are termed 'moral rights' and 'legal rights', and what are moral persons and what are legal persons. The classical formulation of a legal person is an entity that is the beneficiary of 'rights', one that can own property and has the capacity to sue and be sued. This does not mean that inanimate objects, such as mountains, cannot be legal persons for limited purposes (Stone, 1974). However, in the much challenged classical sense, the term legal person is viewed as a fiction used by judges to meet a particular need and not a real determination of status, as long as the target does not have intelligence or will (Gray, 1909).

In legal philosophy there are two major strands of thought that have dominated the debate concerning the relationship between law and morality over the last 200 years, each with a distinctive approach to the origin of rights. In general terms they focus on how open or closed a legal system should be and on whether a legal system should rely on outside factors such as morality or ethics (Schauer, 2004).

The first, legal positivism, derives from Bentham's idea, later expanded upon by his protégé, John Austin, that law derives its power from the sovereign and is the expression of the will of the sovereign enforced by the appropriate control structures. Legal rights, as expressed by the modern positivists Hans Kelsen (1967) and Richard Tur (1987), are whatever the law says they are, independent of morality:

> There is therefore, in this view, no necessary relation between any given set of human characteristics (say, the ability to reason and reflect) and legal personality. Moreover, there is no minimal threshold level of intelligence required to constitute a person. Once a legal right is in evidence, so is a person. Modern legal personality lacks a persistent character over time and place; even its beginnings and cessations are not easy to recognize. Rather, legal personality is better regarded as groupings of rights and duties whose content depends on such factors as age, sex and mental ability (all regarded as natural categories), as well as legal purpose and jurisdiction. Davis and Naffeine, 2001. (p. 54)

In contrast, the natural law view is premised on the belief that there are independent factors, external to man-made law, which give rise to moral obligations. This view, grounded in Aristotle but primarily identified with the Catholic philosophers Augustine and Thomas Aquinas, is functional in outlook. Man has an end ('telos') to which he aims, and his rationality drives the choices he makes in this quest. For Aristotle, man's telos is happiness or flourishing. For Aquinas, it is sharing in God's goodness. With Hugo Grotius, and later John Locke, the natural rights perspective developed. Here, rights are viewed as inherent in man and derive, not from the sovereign, or even from God, but from the person's status as a human being. It is from this tradition that the American Declaration of Independence derived its concepts of inalienable rights such as life, liberty and the pursuit of happiness.

Most modern debates in moral philosophy and jurisprudence have focused on utilitarian views derived from Bentham's positivist tradition and their contrasts with natural rights, or the liberal pluralistic view based on Locke's idea of rights. It has only been in the last 50 years that a neo-Aristotelian natural law tradition has been revived outside the Catholic Church's moral theology and has formed the basis of what has come to be called 'virtue ethics'. The incompatibility of competing views of legal rights will play an important role in how one evaluates an android. The distinction between property and person must also be kept in mind. Roman law and, later, medieval English law, placed great emphasis on one's status as property (slave), or person (freeman). What something

was, the slot into which it fit, was paramount to the question of what rights it possessed. Rights were derivative of status. With the rise of liberal individualism, individuals are viewed as possessing rights

Chapter 2

FUNCTIONALISM AND ITS ANTECEDENTS

In order to establish the thesis of this paper it is first necessary to set forth a view of analytic functionalism which will support our later speculation. Antecedents of ideas which later evolved into functionalism were described by philosophers such as Herbert Feigl, J.C.C. Smart and psychologist UT Place in their development of the identity theory. This theory held that mental states were identical with brain states. In Feigl's words,

> the identity thesis which I wish to clarify and to defend asserts that the states of direct experience which conscious human beings "live through," and those which we confidently ascribe to some of the higher animals, are identical with certain (presumably configurational) aspects of the neural processes in those organisms. Feigl, 1958, in Chalmers 2002. (p. 69)

One consequence which followed from this premise was the suggestion that all mental states could be directly reduced to physical states. The identity theory stressed the position that mental states stood in a direct empirical relationship to brain states. Consequently, this meant that mental states, such as those held by humans, could only exist in human physical states. Interestingly, in language which was later echoed by Searle (1980) and others, Feigl (1958, in Chalmers 2002, p. 71) made the following observation, "As regards the mental life of robots, or as Scriven's (1953) 'androids,' I

cannot believe that they could display all (or even most) of the characteristics of human behavior unless they were made of the proteins that constitute the nervous systems – and in that case they would present no puzzle." As we will see in a moment, this chauvinism was one of the reasons why the identity theory was soon replaced by a more expansive reading which came to be known as functionalism.

Hilary Putman (1973), in "The Nature of Mental States", argued that mental state types, such as being in pain, could not be identical to any particular brain state. From this he concluded that the mental state "being in pain" was in actuality a functional state of the entire organism and not simply a brain state or a disposition to behave in a particular way. This in turn, he argued, would allow for the functional state to exist in different organisms, hence be multiply realizable. This concept avoided the limitation found in the identity theory and allowed for mental states to exist in physical forms different from the physical forms found in humans. As an example of how this theory would work in practice he described a Turing machine, where the look up table of the design served as the mental state type, hence the term "machine functionalism" was applied to his ideas.

In contrast, David Armstrong argued from a type of conceptual analysis that mental states should be defined in terms of the causal role they played. Mental states are the things we mean when we talk about the mind. They are actual internal states which have external causal effect. "The analysis proposed may be called the Causal analysis of the mental concepts. According to this view, the concept of a mental state essentially

involves, and is exhausted by, the concept of a state that is *apt to be the cause of certain effects or apt to be the effect of certain causes*." (emphasis in original). (Armstrong 1981, in Chalmers 2002, p. 82).

Following Armstrong's lead, David Lewis (1972) set forth the tenets of the theory which would form the basis for analytic functionalism.

> In this sequel, I shall uphold the view that psychophysical identifications thus described would be like theoretical identifications, though they would not fit the usual account thereof. For the usual account, I claim, is wrong: theoretical identifications in general are implied by the theories that make them possible – not posited independently. This follows from a general hypothesis about the meanings of theoretical terms: that they are definable functionally, by reference to causal roles. Applied to common-sense psychology – folk science rather than professional sciences, but a theory nonetheless – we get the hypothesis of my previous paper that a mental state M (say, an experience) is definable as the occupant of a certain causal role R – that is, as the state, of whatever sort, that is causally connected in specified ways to sensory stimuli, motor response, and other mental states. Lewis 1972 in Chalmers 2002. (p. 88)

Lewis then proceeded to draw upon the ideas of Frank Ramsey to postulate a theory where the referents of T terms occupy a causal role with respect to O terms, and to

one another. By utilizing the concept of the existential quantifier as defined by Ramsey, Lewis articulated a way to formulate functional definitions of mental state kinds without the risk of circularity inherent in the causal theoretic functionalism of Armstrong. Applying common-sense psychology, as the term generating source, as in the hypothesis cited above, resulted in the following:

> Think of common-sense psychology as a term-introducing scientific theory, though one invented long before there was any such institution as professional science. Collect all the platitudes you can think of regarding the causal relations of mental states, sensory stimuli, and motor responses. …
>
> Add also all the platitudes to the effect that one mental state falls under another – 'toothache is a kind of pain' and the like. … Include only platitudes which are common knowledge among us – everyone knows them, and so on. For the meanings of our words are common knowledge, and I am going to claim that names of mental states derive their meaning from these platitudes.
>
> Form the conjunction of these platitudes; or better, form the cluster of them – a disjunction of all conjunctions of most of them. (That way it will not matter if a few of them are wrong.) This is the postulate of our term-introducing theory. The names of mental states are the T-terms. The O-terms used to introduce them must be sufficient for speaking of stimuli and responses, and for speaking of causal relations among these and states of unspecified nature.

> From the postulate, form the definition of the T-terms; it defines the mental states by reference to their causal relations to stimuli, responses, and each other. When we learn what sort of states occupy those causal role definitive of the mental states, we will learn what states the mental states are… Lewis 1972 in Chalmers 2002. (p. 92)

From this we can see that in Lewis's view folk psychology, or common-sense psychology, is the theory of mind which underlies our everyday talk about beliefs, desires and other types of mental states. The specific terms such as "belief", "desire", and so on, are the terms which make up the T terms. By Ramsiyfying the sentence through use of the existential quantifier, it is possible to derive a complete theory of mind from those terms i.e. we provide an analysis of the theoretical significance of the terms. As with Putnam's machine functionalism, because the mental state type is a functional state type, it is possible that it can be multiply realized. We can also see from this that functional roles are complex relational properties which consist of inputs, outputs and relations between internal mental states. In this sense they are second order properties. It is our everyday talk which plays the role of the provider of the platitudes from which the relationships are implicitly drawn. Furthermore, as in the identity theory of Armstrong, causal properties define the relationship between inputs, relational mental states and outputs. These platitudes cohere and form the theory of mind known as folk psychology. That is, they form a theory about how our minds work.

Folk psychology has another meaning however, and this view has different implications for our analysis. This alternate argument suggests that Lewis's version of folk psychology is an external view i.e., it is an argument concerning a theory about the external world as we know it, not the internal nature of the individuals who created the platitudes in the first place. As put by Stich and Ravenscroft (1994, p. 460) "(Lewis's) folk psychology 'ain't in the head'". What they have offered as an alternative is that there is an internal view of folk psychology which is the way in which we explain human behavior, both our own and that of others. In this sense folk psychology captures the ability that people have to explain behavior in mental state terms. It is the "knowledge representation" which connects the way we perceive behavior and our prediction or explanation of that behavior. This interpretation of folk psychology, because it allows for the predictive ability of the agents who use it in daily discourse, appears to be the view that is most closely related to the way the law operates in our lives. In later work, Stich and Nichols (2003) have grouped this second theory under the rubric "mindreading" and stated the thesis in the following fashion:

> …(O)n the mindreading account, folk psychology is the theory that people actually use in recognizing and attributing mental states, in drawing inferences about mental states, and in generating predictions and explanations on the basis of mental state attributions. It is hard to see why someone who thinks, as functionalists do, that mental state terms get their meaning by being embedded in a theory would want to focus on the platitude-based theory whose principles

> people can easily acknowledge, rather than the richer theory that is actually guiding people when they think and talk about the mind. (p. 241)

There are a number of related arguments which have developed from this distinction. One, the simulation theory, suggests that people use their own mind as a model of the minds of others to run mental simulations of how they think others would behave in like circumstances. In Stich and Nichols (2003), there is an attempt to form a hybrid theory which combines mindreading and the simulation theory. Another attempt, using modularity, appears in Braddon-Mitchell (2004). It is however, suggested that these refinements add nothing to our current analysis and can be set aside at this time. In fact, as stated by Ravenscroft (2004),

> It is implausible that the theory of mind implicit in our everyday talk about mental states is simply identical to the internally represented theory of mind which underpins our capacity to mentalize. Rather, it is likely that folk psychology (internal) is partly inaccessible to consciousness, and that folk psychology (external) is an articulation of that fragment of folk psychology (internal) which is available to conscious reflection. It follows that our everyday talk about the mental is only a rough guide to folk psychology (internal). Ravenscroft, 2004. (p. 8)

This suggests that perhaps there is not so much of a difference between the two views of folk psychology as we might initially think. However, as I point out below, there is still an important distinction which we can exploit to develop our further arguments.

Having set out the basic premises of analytic functionalism and established its reliance on folk psychology, it is necessary, before moving on, to address a major criticism of folk psychology posed by eliminativist materialism. This criticism is based on the argument that folk psychology is simply wrong, massively so. As a result, it cannot support any of the claims made for it in the functionalist account. It must await further definition to decide what is actually true, but the conclusion can be made that the answer is not folk psychology. As framed by Churchland (1988), the issue is as follows:

> As the eliminative materialist sees it, the one-to-one match-ups will not be found, and our common-sense psychological framework will not enjoy an intertheoretic reduction, because our common-sense psychological framework is a false and misleading conception of the causes of human behavior and the nature of cognitive activity. On this view, folk-psychology is not just an incomplete representation of our inner natures; it is an outright misrepresentation of our internal states and activities. Consequently, we cannot expect a truly adequate neuro-scientific account of our inner lives to provide theoretical categories that match up nicely with the categories of our common-sense framework.

> Accordingly, we must expect that the older framework will simply be eliminated, rather than be reduced, by a matured neuroscience. (emphasis in original). (p. 43)

While this passage refers for the most part to the identity theory and the way that theory contended that types of mental states are identical to types of physical states, it is equally important for its critique of folk psychology. To support his contention, Churchland points to the failure of previously held “folk ideas” such as ‘caloric fluid’ which was supposed to flow in our bodies and explained ‘heat’, or ‘phlogiston’ which was thought to be a spirit-like substance which was released from a body when metal rusted or wood burned. Other similar examples are the Ptolemaic view of the universe, and witches. According to Churchland, (1988, p. 44 - 45) “the concepts of folk-psychology – belief, desire, fear, sensation, pain, joy, and so on – await a similar fate … when neuroscience has matured to the point where the poverty of our current conceptions is apparent to everyone…”

However, functionalism and its reliance on folk-psychology is not undermined by eliminative materialist arguments of the type just presented. As explained by Stich and Ravenscroft (1994), both “folk psychology internal” and simulation theory survive the eliminative materialist attack. They are able to do this because they are based on a premise that avoids the argument entirely.

> Internal accounts use the label 'folk psychology' for the knowledge structures that actually underlie skills (cognitive and behavioral capacities) … So on internal accounts(,) folk psychology plays a major role in the explanation of our ability to predict and explain each other's behavior. But on internal construals of folk psychology, the eliminativist's argument may turn out to be incoherent. For it is entirely possible that the knowledge structure underlying our common-sense psychological skills consists entirely of instructions, or it may be a connectionist devise or mental model that does not map comfortably on to a set of sentences or propositions. The eliminativist who adopts an internal reading of "folk psychology" must make the risky bet that none of these options will turn out to be correct. For if one of them is correct, then premise (2) in the eliminativist's argument can't be right, since folk psychology is not the sort of thing that can be either true or false. Stich and Ravenscroft, 1994. (p. 463 - 464)

How then does this help our thesis that a non-biological machine can, under a reading of analytic functionalism, be viewed as legal person? It happens, I suggest, because any such non-biological machine will have to conform to "folk psychology internal". The designer, and any outside observer, will, because she is actually programming the machine, know that it is made up of an explicit knowledge structure; whether it is a connectionist model or a mental model is essentially irrelevant. Its basic structure, by definition, is one we will understand and instantiate in a meaningful way. However, as we will see below, because the knowledge structure will be based upon more than a

simple look up table, this proposed non-biological entity will be more than the Turing type machine described by Putnam. Despite this, it will still meet the requirements of the analytic functionalist model and will be realized in a substrate that is non-biological. As Kim (1992) suggests, perhaps this will mean that we will need a new psychology for this invention if we are to reduce its mental states to physical states, but as I argue, we may not need a new kind of law to accommodate our non-biological machine. This conclusion requires us to show that it does not matter for legal purposes that the entity lacks phenomenal awareness. The criticism of functionalism found in Frank Jackson's knowledge argument, or David Chalmer's "hard problem", (Chalmers, 1995), which suggests that any such functional entity will lack *qualia* or feel, should not, as we will see later, stand as an impediment to our ultimate conclusion.

Chapter 3

HUMANS, PERSONS & PROPERTY

In presenting arguments which would tend to support the idea that a non-biological machine can in some way be developed to a point where it, or a guardian acting on its behalf, could make a plausible claim that it is entitled to legal recognition, many factors are implicated. One in particular which needs initial clarification, is the distinction between the related concepts of human, person and property.

The word “person” is derived from the Latin word “persona” which originally referred to a mask worn by a human who was conveying a particular role in a play. In time it took on the sense of describing a guise one took on to express certain characteristics. Only later did the term become coextensive with the actual human who was taking on the persona, and thus became interchangeable with the term “human”. Even as this transformation in linguistic meaning was taking place, the concepts of person and human remained distinct.

To Greeks such as Aristotle, slaves and women did not possess souls. Consequently, while they were nominally human, they were not capable of fully participating in the civic life of the City and therefore not recognized as persons before the law. Because they were not legal persons, they had none of the rights possessed by full members of Athenian society. Similarly, Roman law, drawing heavily from Greek antecedents, made clear distinctions, drawing lines between property and legal persons,

but allowing for gradations in status and in the case of slaves, permitting movement between categories.

Legal theory has historically drawn a distinction between property and person, but with the implicit understanding that person equates to human. Locke (1689b/1975) did attempt to make a distinction between the two in his *Essay Concerning Human Understanding.* There, he was concerned with drawing a contrast between the animal side of man's nature and what we customarily call man. *Person* in his sense belongs 'only to intelligent Agents capable of a Law, and Happiness and Misery' (Locke, 1689b: chapter XXVII, section 26). Until recently there had not been a need to make more precise distinctions. Since the expression of different opinions by Strawson (1959) and Ayers (1963), the concept of human and person has become much more a topic for philosophical speculation. To define a category of rights holder that is distinguishable from human, but none the less comprehensible, we can resort to a form of 'fiction'. This proposition is subject to much debate, but at least one use of the fiction is in the comparison between a being of the species *Homo sapiens* and the legal concept of the corporation as a person (Note, 1987). With androids, we avoid relying on this fiction by making a clear distinction between what we define as a 'human' and what we define as a 'person' (Calverley, 2005 a, b, c).

> Only when a legal system has abandoned clan or family responsibility, and individuals are seen as primary agents, does the class of persons coincide with the

> class of biological individual human beings. In principle, and often in law, they need not.… The issue of whether the class of persons exactly coincides with the class of biologically defined human being – whether corporations, Venusians, mongolian idiots, and fetuses are persons – is in part a conceptual question. It is a question about whether the relevant base for the classification of persons requires attention to whether things look like "us," whether they are made out of stuff like "ours," or whether it is enough that they function as we take "ourselves" to function. If Venusians and robots come to be thought of as persons, at least part of the argument that will establish them will be that they function as we do: that while they are not the same organisms that we are, they are in the appropriate sense the same type of organism or entity. Rorty, 1976. (p. 322 - 323)

The distinction between human and person is controversial (MacDorman and Cowley, 2006). For example, in the sanctity of life debate currently being played out in the USA, serious arguments are addressed to the question whether a human fetus becomes a person at conception, or at a later point of viability (Warren, 1991; Ramey, 2005b). Similar questions arise at the end of life: Do humans in a persistent vegetative state lose the status of legal person while still remaining human at the physical/genetic level? Likewise, children and individuals with serious mental impairments are treated as persons for some purposes but not for others, although they are human. Personhood can be defined in a way that gives moral and legal weight to attributes that we ultimately define

as relevant without the requirement that the entity either be given the full legal rights of humans or burden them with the duties those rights entail. (Calverley, 2005b).

Others have stated “[i]n books of the Law, as in other books, and in common speech, ‘person’ is often used as meaning a human being, but the technical legal meaning of ‘person’ is a subject of legal rights and duties.”. (Gray, 1909, p. 27). However, by qualifying this definition with the caution that it only makes sense to give this appellation to beings which exhibit “intelligence” and “will”, Gray equated person with human. From this we can infer that the real issue is to determine just what attributes the law is particularly interested in defining so that it can specify the who, or what, to which it applies.

We also see a similar type of development in the concepts of person and property as society developed in the Middle Ages in Western Europe, more particularly in England. There, the concepts of a legal person and property became intertwined. Over time, a person was defined in terms of the status he maintained, and obligations he owed, in relationship to real property. With Locke, these ideas coalesced into the view that a person had a property right in his own labor. By exercising this right he exercised control over things such as land and the goods he produced. It was from this view of a person as property holder that the so called Fiction Theory of corporate personality initially derived. The validity of this proposition as a basis for legal personhood, as we will see later, is subject to much debate, but at least one generally accepted instance where the fiction has been used can be shown by the comparison between a being of the

species homo sapiens and the legal concept of the corporation as a person. (Note, 1987) Because synthetic entities such as corporations were authorized by their state granted charters of organization to own property, they were deemed to be "persons". In the earliest cases the idea that the corporation was an artificial entity was based solely on this derivative claim. For example, in the case Trustees of Dartmouth College v. Woodward, 17 U.S. 518 (1819), English precedent, and a specific grant of rights from King George III, was relied upon by Chief Justice Marshall to support the conclusion that a group of trustees could act as a body corporate and exercise rights such as holding real property. It was only later, following the US Civil War, when courts were forced to respond to arguments based on the anti-slavery amendments to the United States Constitution that the concept of a corporation as the equivalent of a natural person began to be articulated. The answer was that the use of the term "person" in the language of the 14th Amendment to the Bill of Rights,[1] was broad enough to apply to artificial groupings of people not just humans. This idea, based on the view that corporations are nothing more than a grouping of individual persons who have come together for a particular purpose, has come to be known as the Aggregate Theory. These ideas will be revisited later when we examine the concept of intentionality as it applies to our analysis.

---

1 Section 1. All persons born or naturalized in the United States, and subject to the jurisdiction thereof, are citizens of the United States and of the state wherein they reside. No state shall make or enforce any law which shall abridge the privileges or immunities of citizens of the United States; nor shall any state deprive any person of life, liberty, or property, without due process of law; nor deny to any person within its jurisdiction the equal protection of the laws.

Perhaps the most obvious and well known social institution where the tension concerning the distinction between person and property came to the attention of the Courts was with the institution of slavery. As a preliminary note, although slavery as practiced in the seventeenth, eighteenth, and nineteenth centuries, particularly in the Americas, had, at least superficially, a strong racial component, most of the actual writings and cases from contemporary sources indicate that race actually played only a small part in the legal discussions of the institution. The theoretical underpinnings were non-racial in origin and related more to status as property than to skin color. (Tushnet, 1975) This was also true in other countries at the same time such as Russia with its system of serfdom, which was clearly as oppressive as American slavery.

In looking at the reported cases decided at that time we can see that the real struggle the Courts were having was with the justification of defining a human as property, that is, as a non-person for purposes of the law. In a famous English case, Somerset's case, *98 Eng. Rep. 499, 1772*, a slave was brought from the Americas to England by his owner. When he arrived, he argued that he should be set free by his master. The master's response was that he was property and should not be free. The Court stated that there was no issue with the black man's humanity, he clearly was human. As a slave, he had been deprived of his right to freedom and was treated as property in parts of the world. However, because there was no provision in English positive law that permitted a human being to be deprived of his freedom and treated

as property, he could not be deemed a slave under English law. Note however, the careful way in which the ruling was limited to the fact that it was positive law which did not allow the enslavement of this human and thereby led to the conclusion that Somerset could not continue to be held as a slave. The clear implication is that if positive law had been different, the result might also have been different. The Court was drawing a clear distinction between Somerset's status as a human and his status as a legal person. Similar theoretical justification can be seen in early cases decided in the United States, until the passage in most southern states of what are generally called the Slave Acts. However, in perhaps one of the most egregious and incendiary rulings by the United States Supreme Court, the Dred Scott case *(60 US 19 How. 393, 1857),* Chief Justice Taney reached a conclusion opposite to that of Somerset's case, ruling that the Constitution did not extend to protect a black man because at the time of its passage, the meaning of person did not include slaves. This left the regulation of slavery up to each state. From this we can conclude that it is the exercise of positive law, expressed in making, defining and through manipulating, the definition of the legal concept of person, that is the defining characteristic of these cases. It is not the slave's status as human being.

From the above discussion we can draw the conclusion that person and human are distinct concepts recognized by the law. Property, in Locke's sense of being the end product of mental or physical effort, is also recognized by the law. However, because a person is able to own property, a person can be defined in a way that allows us to draw a

distinction between property and a legal person. It is now our task, as Rorty (1976) suggested, to find those attributes of the non-biological artifact which functionally indicate that we can make a principled distinction between it and the property it produces, thereby coming to the conclusion that it can support the claim that it is a legal person. But before doing so we need to look at the institution we call “law”.

Chapter 4

LAW AS A NORMATIVE SYSTEM

> Law is a socially constructed, intensely practical evaluative system of rules and institutions that guides and governs human action, that help us live together. It tells citizens what they may, must, and may not do, and what they are entitled to, and it includes institutions to ensure that law is made and enforced. Morse 2004, (p. 158).

This definition, on its face, seems to be elegant and concise but, like an iceberg, it is deceptive. The tip only hints at the complexity of what is under the surface. Rather than simply applying the definition, let us begin by setting a foundation for the discussion which follows. In order to determine whether "law" has any normative value when it is used to evaluate the idea of machine consciousness we first need to gain at least a basic understanding of how this thing we call "law" is formulated at a conceptual level. By understanding what we mean when we speak of law, where it derives its ability to regulate human conduct, we can perhaps begin to formulate criteria by which some aspects of law could also be used to test the idea that something we have created in a machine substrate is a new form of conscious being. If this can be done in a way that is meaningful both to those who will be faced with deciding how to regulate such an entity and to the designers who are actually making the effort to create such an artifact, then it is worth the effort. As with most endeavors, it is often the question one asks at the outset which determines the nature of the debate and directs the form of the ultimate outcome.

If we want to design a machine consciousness which we will claim is the equivalent of a human, we should determine as early as possible in the process whether the result we seek will stand up to scrutiny and will, in the end, satisfy criteria which are consistent with the way humans govern themselves and view each other.

Applying the tools of analytic jurisprudence, philosophers of law examine the ways in which law is distinguishable from other normative systems. Historically, this has been done by conceptual analysis or intuition pumping with little or no empirical analysis. Only recently have empirical facts and evolutionary analysis started to be used to evaluate legal concepts in an attempt to naturalize law. (Jones, 1997). There are many variations and nuances in legal theory, it is generally acknowledged that there have been two major historic themes which have, for the last few hundred years, dominated the debate about what "law" means.

One of the most familiar ideas to western societies is the concept of natural law, which was originally based on the Judeo-Christian belief that God is the source of all law. It was this belief which underpinned most of western civilization until the Enlightenment period. Prominent thinkers such as Augustine and Thomas Aquinas are two examples of this predominant orthodoxy. In essence, natural law proponents argue that law is inextricably linked with morality and therefore an 'unjust law is no law'.

With the Enlightenment came a decreasing emphasis on God as the giver of all law, and an increasing development of the idea that humans possessed innate qualities which gave rise to "law". As members of society, humans were capable of effecting their own decisions and consequently were entitled to govern their own actions based upon their intrinsic worth as individuals. While this concept was originally suggested by Hugo Grotius (1625) and later refined by John Locke (1689a/1967) it arguably reached its most notable actual expression in the system of laws ultimately idealized by the drafters of the United States Declaration of Independence. Drawing on a similar argument and applying it to moral philosophy, Immanuel Kant hypothesized that humans were, by the exercise of their reason, capable of determining rules that were universally acceptable and applicable, and in turn able to use those rules to govern their conduct. (Kant, 1785).

More recently, John Finnis, building on ideas reminiscent of Kant, has outlined what he calls basic goods (which exist without any hierarchical ranking), and then has posited the existence of principles which are used to guide a person's choice when there are alternative goods to choose from. These principles, which he describes as the "basic requirements of practical reasonableness", are the connection between the basic good and ultimate moral choice. Derived from this view, law is the way in which groups of people are coordinated in order to effect a social good or to ease the way to reach other basic goods. Because law has the effect of promoting moral obligations it necessarily has binding effect (Finnis, 1980). Similarly, Lon Fuller argued that law is a normative system for guiding people, and must therefore have an internal moral value in order to

give it its validity. Only in this way can law fulfill its function which is to subject human conduct to the governance of rules (Fuller, 1958; 1969). Another important modern theorist in this natural law tradition is Ronald Dworkin. Dworkin advocates a thesis which states in essence that legal principles are moral propositions grounded on past official acts such as statutes or precedent. As such, normative moral evaluation is required in order to understand law and how it should be applied (Dworkin, 1978).

In contrast to the basic premise of natural law, that law and morality are inextricably intertwined, stands the doctrine of legal positivism. Initially articulated by Jeremy Bentham, and derived from his view that the belief in natural rights was "nonsense on stilts" (Bentham, 1824), criticism of natural law centered around the proposition that law is the command of the sovereign, while morality tells us what law ought to be.   This idea of law as a system of rules "laid down for the guidance of an intelligent being by an intelligent being having power over him", was given full voice by Bentham's protégé, John Austin. In its simplest form this idea is premised on the belief that law is a creature of society and is a normative system based upon the will of those ruled as expressed by the sovereign.   Law derives its normative power from the citizen's ability to know and predict what the sovereign will do if the law is transgressed (Austin, 1832).

Austin's position, that law was based on the coercive power of the sovereign, has been severely criticized by the modern positivist H.L.A. Hart who has argued that law

requires more than mere sanctions; there must be reasons and justifications why those sanctions properly should apply. While neither of these positions rule out the overlap between law and morality, both do argue that what constitutes law in a society is based on social convention. Hart goes further and states that this convention forms a rule of recognition, under which the law is accepted by the interpreters of the law, i.e. judges. (Hart, 1958; 1961). In contrast, Joseph Raz argues that law is normative and derives its authority from the fact that it is a social institution which can claim legitimate authority to set normative standards. Law serves an essential function as a mediator between its subjects and points them to the right reason in any given circumstance, without the need to refer to external normative systems such as morality. (Raz, 1975).

It is conceded that the above exposition is vastly over simplified and does not do justice to the nuances of any of the described theories. None the less, it can serve as a basis for the contention that despite the seeming difference between the two views of law, there is an important point of commonality. Returning to the definition with which we started this paper, we can see that it is inherently legal positivist in its outlook. However, its central idea, that law is a normative system by which humans govern their conduct, seems to be a characteristic shared by both major theories of law and therefore is one upon which we can profitably ground some further speculation. To the extent that law requires humans to act in accordance either with a moral norm established in accordance with a theological, or natural theory, or to the extent it is a normative system based on one's recognition of and compliance with a social created standard of conduct, it is

premised on the belief that humans are capable of, and regularly engage in, independent reflective thought, and are able to make determinations which direct their actions based upon those thoughts. Described in a slightly different way, law is based on the premise that humans are capable of making determinations about their actions based on reason.

Chapter 5

REASON AS A BASIS FOR LAW AND ETHICS, A CLASSICAL VIEW

The place which reason occupies in natural law theory is derived from the views of the classical writers in this tradition, specifically, the central role played by reason in the writings of Aristotle, Cicero and Aquinas. It is from the point of view of these writers that we can begin to understand the significance of the premise that reason is at the center of the concept of natural law.

It is generally accepted that Aristotle was one of the first writers and theoreticians to explicitly look at and compile in a systematic fashion, the ideas which generally are viewed as precepts of a natural law perspective of ethical behavior. Ethical behavior in this sense means the way one lives ones life in order to be a fully functioning human being. Aristotle himself describes the inquiry in the *Nicomachean Ethics*, in the following fashion:

> Every art and every inquiry, and similarly, every action and pursuit, is thought to aim at some good; and for this reason the good has rightly been declared to be that at which all things aim. … If then, there is some end of the things we do, which we desire for its own sake (everything else being desired for the sake of this), and if we do not choose everything for the sake of something else... clearly this must be the good and the chief good. Book I, Chap. 1 and Chap. 2.

From this starting point Aristotle views his task as one directed at determining just what this good is, if it even exists, and more importantly how a man is supposed to go about achieving this good. The preliminary position he establishes is that happiness (eudaimonia), in the sense of living well or faring well, constitutes the general consensus of what is good. But as he goes on to state, the achievement of pleasure or honor or wealth is not what is meant by happiness in the sense that he is seeking, rather, what he seeks is the ultimate good. Consequently, he rejects this list as the definition of the goal to which action is aimed.

The inquiry he pursues then narrows in focus.

> Therefore, if there is an end for all that we do, this will be the good achievable by action. ... (I)f there is only one final end, this will be what we are seeking, ... Now we call that which is in itself worthy of pursuit more final than that which is worthy of pursuit for the sake of something else and that which is never desirable for the sake of something else more final than the things that are desirable both in themselves and for the sake of that other thing, and therefore we call final without qualification that which is always desirable itself and never for the sake of something else. Book I, Chap. 7.

Once again the conclusion he reaches is that happiness is the final end of action.

> Now such a thing happiness above all else, is held to be: for this we choose always for its self and never for the sake of something else.... Happiness, on the other hand, no one chooses for the sake of these (other virtues) nor, in general, for anything other than itself. Book I, Chap. 7

He then goes on to state that the same applies to self-sufficiency of happiness, reaching his ultimate point that "happiness then is something final and self-sufficient, and is the end of action."

However, because Aristotle does not view happiness simply as a state or condition of existence but rather as a virtuous activity of the soul, he believed it was necessary to provide a clear account of just exactly what happiness means. To do this he first states that it is necessary to ascertain the function of man because "the good and the "well" is thought to reside in the function...". It is here, in the analysis of man's function where Aristotle first identifies what he terms the "rational principle". The thing which sets man apart from plants and animals is that man has "… an active life of the element that has a rational principle….". Then, from this, Aristotle derives the argument that if we can agree that "…the function of man (is) to be a certain kind of life, and this (is) to be an activity or actions of the soul implying a rational principle, and the function of a good man (is) to be the good and noble performance of these, and if any action is well performed when it is performed in accordance with the appropriate excellence... human

good turns out to be the activity of soul exhibiting excellence, and if there are more than one excellence, in accordance with the best and most complete." Book I, Chap. 7.

We can see from this that happiness is a part of life motivated and directed by the rational principle. It is activity of the soul in accordance with perfect virtue. As a result, Aristotle, in Book I Chap. 13, turns to a discussion of virtue, and in particular, human virtue. But, he is careful to note that it is not virtue of the body he seeks but virtue of the soul. Here again Aristotle sets up a dialectic, as with so many other aspects he discusses. First, there is an irrational portion or element of the soul and a second or rational aspect. But even the irrational is bifurcated into two forms. One he believes is widely shared by man and animals, and is the kind of nutritive growth oriented aspect found in "nurslings and embryo". The other aspect can be influenced and directed by the rational principle. In this sense there is a tug-of-war, so to speak, between the irrational and the rational part of the soul. An example which shows this point is the fact that humans can give and take advice and are susceptible to exhortations or admonitions.

It is because of this split which he sees between the rational and irrational aspects of the soul that Aristotle further postulates that virtue is similarly bifurcated into what he calls the intellectual and the moral virtues. Broadly speaking the intellectual virtues arise from nature but require experience and time to come to fruition. On the other hand, the moral virtues do not arise from nature and can come about only by habit. It is the

intellectual virtues where reason plays its most important role, and where, in particular, we see reason related directly to a theory which can properly be called natural law.

Aristotle identifies five intellectual virtues: science, craft expertise, intuitive understanding, practical reasoning/wisdom, and theoretical reasoning/wisdom. For our purposes here, I intend to address only the last two because it is here I believe we will find the relationships between reason and natural law spelled out most clearly. First, I address practical wisdom. Aristotle, as we saw above, argues that part of the irrational soul is in tension between reason and irrationality. This tug-of-war leads man to try to find the middle ground, that is, the Golden Mean. The way this process is accomplished is by the irrational part of the soul being affected by right reason. But this in turn requires us to determine "... what is the right rule and what is the standard that fixes it." Book VI, Chap. 1.

As he did with the irrational part of the soul, Aristotle also draws a distinction between two aspects of the rational part of the soul. He states "... there are two parts which grasp a rational principle -one by which we contemplate the kind of things whose originative causes are invariable, and one by which we contemplate variable things." Book VI Chap. 2. This is considered further, and results in a distinction on the one hand between a state of the intellect which is not practical or productive but is concerned with the good and bad, truth and falsity directly, and, on the other hand, a state of the intellect which is practical where the good state is truth in agreement with right desire. This latter

quality, identified as practical wisdom, is "... a reasoned and true state of capacity to act with regard to human goods." Theoretical wisdom on the other hand, is based on scientific contemplation of the invariable aspects of nature, and is concerned with matters of first principle. In combination with intuitive reason, man is able to grasp the first principles with which science is concerned. It is this combination which Aristotle calls philosophic wisdom. "... (P)hilosophic wisdom is scientific knowledge combined with intuitive reason, of the things that are highest by nature." Book VI Chap.7. "... (P)ractical wisdom on the other hand is concerned with things human and things about which it is possible to deliberate."

While arguing that theoretical wisdom is higher than practical wisdom, Aristotle none the less clearly states that practical wisdom is an important requirement for human goodness. "It is clear then, from what has been said, that it is not possible to be good in the strict sense without practical wisdom or practically wise without moral virtue." Book VI Chap. 13. However, Aristotle believes that because practical wisdom is aimed at man, and finds its expression through politics, it is not the highest form of happiness. This highest form of happiness, he explains in Book X, derives from theoretical wisdom, the contemplation of what is true or false, good or bad, in nature.

After rejecting the idea that happiness is mere amusement, Aristotle focuses on happiness as it has been discussed throughout the preceding books of the "Ethics".

> If happiness is activity in accordance with virtue, it is reasonable that it should be done in accordance with the highest virtue; and this will be that of the best thing in us. Whether it be reason or something else that is this element which is thought to be our natural ruler and guide and to take thought of things noble and divine, whether it be itself also divine or only the most divine element in us, the activity of this in accordance with its proper virtue will be perfect happiness. That this activity is contemplative we have already said." Book X Chap. 7.

From this he draws the further insight that "... this activity is the best (since not only is reason the best thing in us, but the objects of reason are the best of knowable objects); and, secondly, it is the most continuous, since we can contemplate truth more continuously than we can do anything."

He then turns to a somewhat more practical consideration and attempts to show that happiness needs less of material goods than do the political virtues, and likewise needs less activity than do the virtues such as honor or courage. If happiness can be said to depend on anything it is leisure. But query, if happiness depends on leisure, is not leisure the greater good? Aristotle sidesteps this point by incorporating leisure into happiness as a component of it, giving it merely a supporting role to reason.

> ... but the activity of reason, which is contemplative, seems both to be superior in serious worth and to aim at no end beyond itself, and to have that its

> pleasure proper to itself (and this augments the activity), and the self-sufficiency, leisureliness, unweariedness (so far as this is possible for man), and all the other attributes ascribed to the supremely happy man are evidently those connected with this activity, it follows that this will be the complete happiness of man if he be allowed a complete turn of life (for none of the attributes of happiness is incomplete). Book X Chap.7

Aristotle, immediately after this passage sounds a cautionary note, which establishes without doubt his commitment to the idea that happiness and more particularly the activity of reason which constitutes its motive force can only be viewed as coming from the natural law. "But such a life would be too high for man; for it is not in so far as he is man that he will live so (in happiness), but in so far as something divine is present in him; and by so much as this is superior to our composite nature is its activity superior to that which is the exercise of the other kind of virtue. If reason is divine, then, in comparison with man, the life according to it is divine in comparison with human life." Book X Chap. 7.

So, Aristotle urges, "... so far as we can, make ourselves immortal, and strain every nerve to live in accordance with the best thing in us; for even if it be small in bulk, much more does it in power and worth surpass everything... And what we said before will apply now: that which is proper to each thing is by nature best and most pleasant for each

thing; for man, therefore, the life according to reason is best and pleasantest, since reason more than anything else is man. This life therefore is also the happiest. Book X Chap 7.

Finally at least for our purposes here, we can tie this idea up with one additional quote from Book X:

> Now he who exercises his reason and cultivates it seems to be both in the best state of mind and most dear to the gods. For if the gods have any case for human affairs, as they are thought to have, it would be reasonable both that they should delight in that which it was best and most akin to them (i.e. reason) and that they should reward those who love and honor this most, as caring for the things that are dear to them and acting both rightly and nobly. And that all these attributes belong most of all to the philosophers is manifest. He, therefore, is the dearest to the gods. And he who is that will presumably be also the happiest; so that in this way too the philosopher will, more than any other, be happy.

The connection which Aristotle discusses above and almost seemingly treats as an afterthought; that reason is most akin to godlike qualities, was made explicit by Cicero in *The Laws.* In particular in Book 1 Chap. 17 he starts his inquiry as follows: "we must clarify the nature of justice, and that has to be deduced from the nature of man." He then goes on to note that law is the highest reason, a force of nature, and it is from law that

justice is derived. For this reason he indicates that he "... shall look to nature for the origins of justice."

With this preamble, Cicero then proceeds to make explicit his view of the connection between God, man and reason.

> The creature of foresight, wisdom, variety, keenness, memory, endowed with reason and judgment, which we call the man, was created by the supreme God to enjoy a remarkable status. Of all the types and species of living creatures he is the only one that participates in reason and reflection, whereas none of the others do. What is there, I will not say in man, but in the whole of heaven and earth, more divine than reason, (a faculty which, when it has developed and become complete, is rightly called wisdom)?
>
> Since, then, there is nothing better than reason, and reason is present in both man and God, there is a primordial partnership in reason between man and God. But those who share reason also share right reason; and since that is law, we men must also be thought of as partners with the gods in law. Furthermore, those who share law share justice. Now those who share all these things must be regarded as belonging to the same state; and much the more so if they obey the same powers and authorities. Book 1, Chap 22 and 23.

Cicero then argues that reason came directly from God. At the proper time in the life of the universe when the human race was sown on earth, the race was "... endowed with the divine gift of mind; that whereas men derived the other elements in their makeup from their mortal nature -- elements which are fragile and transitory -- their mind was implanted in them by God." Book 1 Chap 24. Similarly moral excellence resides in both man and God because it is the completion and perfection of nature.

This in turn directs us back to the original inquiry "... nothing is more vital than the clear realization that we are born for justice, and that what is just is based, not on opinion, but on nature." Book 1 Chap 28.

From this he concludes "... we have been made by nature to share justice amongst ourselves and to impart it to one another.... For those who have been endowed by nature with reason have also been endowed with right reason, and hence with law, which is right reason in commanding and forbidding; but if with law, then with justice too. But reason has been bestowed on everybody; therefore the same applies to justice." Book 1 Chap 33.

In looking at the Stoic position we see distinct similarities to the arguments propounded by Aristotle. But, as noted at the outset of this section, the connection between reason and God is made explicit. Reason is implanted by God and is the core attribute which differentiates humans from animals. Cicero states the same in Book V of

his work “On Moral Ends”. “We seek a life in which the virtues of both mind and body are fully realized. This is where the supreme good is to be found,… In humans, however, the mind, and especially reason, is absolutely paramount. Reason is the source of virtue, and virtue is defined as the perfection of reason.” Book V, Paragraphs 37 and 38. Like Aristotle, Cicero believes that the supreme good is to found in this life if one learns to live in accordance with virtue.

The third major work where we become aware that reason has a direct impact on one's conception of natural law, is the complex development found in Thomas Aquinas, and particularly in the Summa Theologica, Book I Part II Questions 90 to 97 *Treatise on Laws.* Here we see a point of view which, while relying substantially upon concepts initially set forth by Aristotle, goes well beyond Aristotle, and Cicero, in its conclusions. Aquinas’ concept of God and the inter-relation of reason in God and man can also be seen to have much in common with both Aristotle and with the Stoic view of Cicero, at least at a superficial level, but ends with a much different result in so far as the ability of man to achieve happiness in this life. As an initial observation it is quite understandable why in Aquinas, there is a heightened sense of connection between man and a theological God. He is writing a theological treatise which is intended to argue for the exalted position of the Roman Catholic Church in society, particularly political society. His view of God, and the role God plays in the world, is intended to bolster the proposition that God is supreme and man’s role is to become one with God in blessedness, not in this

world, as Aristotle believed, but in the life after death where goodness will be rewarded and vice will be punished.

Aquinas begins with an initial inquiry into the “essence" of law, and immediately answers

> ... that law is a rule and measure of acts that induces persons to act or refrain from acting.... And the rule and measure of human acts is reason, which is the primary source of human acts.... For it belongs to reason to order us to our end, which is the primary source regarding our prospective action…. Reason has from the will the power to induce activity... since reason commands means because one wills ends. But an act of reason needs to rule the will regarding the means commanded in order that the willing have the nature of law. And we in this way understand that the will of the ruler has the force of law. ST I –II, Q 90, A1.

However, while seemingly starting out at a point closely akin to that of Aristotle, Aquinas appears to reject the idea that happiness, in the sense of living well, is the end to which humans aim. True, his view is still a functionalist view as was Aristotle's, but the end toward which humans aim is not just human happiness but blessedness. Aristotle and Aquinas recognized that human happiness in this life requires right reason to direct external action and govern internal emotions, as well as a rightly ordered will regarding

the requisite images of human actions and emotions, but Aquinas goes further and maintains that living in this way is the only way to attain man's end which is eternal blessedness, and that can only be finally gained in the after life.

In outline form, Aquinas' view in the "Treatise on Laws", posits four types of law; eternal law, natural law, human law and divine law. Reason plays a central role in each, but most particularly in his view of natural law. Again, this is because Aquinas, as did his predecessors, viewed man as a rational being. But Aquinas takes great pain to establish that man is not the only rational being. God is also rational, as well as being eternal.

The eternal law, which is God's plan for all creatures, applies to both rational and irrational creatures. But they share in this law in different ways "… one way as things partake of the eternal law in a conscious way; a second by acting and being acted upon as things partake of the eternal law by reason of causes acting on them. And irrational creatures are subject to the eternal law in a second way, as I have said. But because rational natures, along with what is common to all creatures, have something proper to them as rational, they are consequently subject to the eternal law in both ways." ST I-II, Q 93, A 6.

However, only the virtuous rational being is fully part of the eternal law. "… (T)he virtuous are completely subject to the eternal law, as they always act in accord with

it. And the wicked are indeed incompletely subject to the eternal law regarding their own actions, as they incompletely recognize and incompletely incline to goodness." ST I-II, Q 93, A 6. But the price the wicked pay for this failure to recognize the eternal law is that they suffer.

Turning next to natural law, Aquinas states that

> … the precepts of the natural law are related to practical reason as the first principles of scientific demonstration are related to theoretical reason. For both the precepts of the natural law and the first principles of scientific demonstrations are self-evident principles….
>
> And so the first indemonstrable principle is that one cannot at the same time affirm and deny the same thing….And as being is the first thing that without qualification falls within our understanding, so good is the first thing that falls within the understanding of practical reason. And practical reason is ordered to action, since every efficient cause acts for the sake of an end, which has the nature of good. And so the first principle in practical reason is one based on the nature of good, namely, that good is what all things seek. Therefore, the first precept of the natural law is that we should do and seek good, and shun evil. And all the other precepts of the natural law are based on that precept, namely, that all the things that practical reason by nature understands to be human goods or evils belong to precepts of the natural law as things to be done or shunned.

> And since good has the nature of end, and evil the nature of the contrary, reason by nature understand to be good all the things for which human beings have a natural inclination, and so to be things to be actively sought, and understands contrary things as evil and to be shunned. Therefore, the ordination of our natural inclinations ordains the precepts of the natural law. ST I-II, Q 94, A 2.

Aquinas then presents some examples of how these precepts are ordained. The third is the most salient for our discussion. " … Human beings have inclinations for good by their rational nature, which is proper to them…. And so things that relate to such inclinations belong to the natural law…."

This is elaborated upon further, "… since the rational soul is the specific form of human beings, everyone has an inclination from one's nature to act in accord with reason. And this is to act virtuously. And so in this regard, all virtuous acts belong to the natural law, since one's own reason by nature dictates that one act virtuously." ST I-II, Q 94 A 3.

Aquinas also argues that reason is the power which controls all other human powers. "As the power of reason in human beings rules and commands other powers, so reason needs to direct all the natural inclinations belonging to the other powers. And so it is universally correct for all persons to direct all their inclinations by reason." ST I –II, Q 94, A 4, RO 3.

Having established that natural law is the way in which God's creatures participate in the eternal law and that man's special position is derived from and dependent upon man's rational nature, Aquinas then goes on to explain that while human beings by nature have the ability and inclination to be virtuous, virtue does not come to us automatically. There needs to be training. Because, as he says, virtue consists mostly of restraining one's self from excess pleasures, it is necessary, particularly for youth, to have training in virtue. For those who are naturally inclined to be virtuous, paternal training or admonitions are probably enough. However, there are people who are wicked and prone to vices. These people need to be restrained by force and fear. This force, he suggests, is supplied by human law. ST I –II, Q 95, A 1.

Similarly, "…human reason cannot partake of the complete dictates of God's reason but partakes of them in its own way and incompletely. And so regarding theoretical reason, we by our natural participation in God's wisdom know general principles but do not specifically know every truth, as God's wisdom does. Just so regarding practical reason, human beings by nature partake of the eternal law as to general principles but not as to particular specifications of particular matters, although such specifications belong to the eternal law. And so human reason needs to proceed further to determine the particular prescriptions of human law." ST I – II, Q 90, A 3, RO 1.

In this sense human participation in the eternal law is limited and needs augmentation. Likewise, because human nature is an imperfect form of reason when compared to God's reason, humans need more direction, particularly in order to reach their true end.

> … (L)aw directs our acts in relation to our ultimate end. And human beings, if they were indeed ordained only for an end that did not surpass the proportion of their natural ability, would not, regarding reason, need to have any direction superior to the natural law and human laws derived from the natural law. But because human beings are ordained for an end of eternal blessedness, which surpasses their proportional natural human capacity, … God needed to lay down a law superior to the natural law and human laws to direct human beings to their end. ST I-II, Q 90, A 4.

And this additional law, according to Aquinas, is the divine law, which is revealed to humans by God. "The natural law partakes of the eternal law in proportion to the capacity of human nature. But human beings need to be directed in a higher way to their ultimate supernatural end. And so God gives an additional law that partakes of the eternal law in a higher way" ST I –II, Q 90, A 4, RO 1. and this, according to Aristotle is divine law.

From this we can see that while Aquinas draws upon a similar basic understanding of the rational nature of humans, and ascribes this rationality to the core of the nature of being human, the results is quite different from the result reached by Aristotle and the Stoics. In each of the earlier cases, man, in general, and for Aristotle, the philosopher in particular, is capable, if trained properly, of acting in ways which will lead to the attainment of man's function (ergon) in the form of happiness here on earth. For Aquinas, reason here on earth serves a different function because the true nature of man is not just earthly happiness, but happiness which can only be achieved by the attainment of eternal blessedness. Only through reason, exercised on earth, augmented by God's assistance in the form of divine law, can man reach his ultimate goal.

In conclusion, we have seen that reason is central to the understanding of humans held by all three of the classical writers. Humans are viewed as the quintessential rational beings. It is this attribute of rationality which sets them apart from other creatures. It is suggested that this classical view of rationality is at the heart of the folk psychology view, which will be discussed below.

Chapter 6

LAW AND PRACTICAL REASON

> Human action is distinguished from all other phenomena because only action is explained by reasons resulting from desires and beliefs, rather than simply by mechanistic causes. Only human beings are fully intentional creatures. To ask why a person acted a certain way is to ask for reasons for action, not the reductionist biophysical, psychological, or sociological explanations. To comprehend fully why an agent has particular desires, beliefs, and reasons requires biophysical, psychological, and sociological explanations, but ultimately, human action is not simply the mechanistic outcome of mechanistic variables. Only persons can deliberate about what action to perform and can determine their conduct by practical reason. Morse, 2004, (p. 160)

Similarly, Gazzaniga and Steven, express the idea as follows:

> At the crux of the problem is the legal systems view of human behavior. It assumes (X) is a "practical reasoner", a person who acts because he has freely chosen to act. This simple but powerful assumption drives the entire legal system.
>
> …
>
> The view of human behavior offered by neuroscience is simply at odds with this idea. …neuroscience is in the business of determining the mechanistic actions of the nervous system. The brain is an evolved system, a decision-making

> device that interacts with its environment in a way that allows it to learn rules to govern how it responds. It is a rule based device that, fortunately, works automatically.
>
> …
>
> Neuroscience will never find the brain correlate of responsibility, because that is something we ascribe to humans, not to brains. It is a moral value we demand of our fellow, rule-following human beings. … The issue of responsibility … is a matter of social choice. …We are all part of a deterministic system that some day, in theory, we will completely understand. Yet the idea of responsibility is a social construct and exists in the rules of society. It does not exist in the neuronal structure of the brain. (p. 67).

The point is clear that the law looks at the motivation of the actor and the ability of the actor to control actions based upon those motivations. Gazzaniga has expressed similar arguments as follows:

> Neuroscience seeks an empirically valid model of human nature and human behavior – one that has predictive power and allows us to understand better the relation between our brains and mental lives. The law seeks to bring about conformity of individuals' behavior to certain codes in order to maintain order in society. Waldbauer and Gazzaniga, 2001, (p. 364)

While this perspective is not universally accepted by philosophers of law, (Goodenough, 2001), for the purpose of this paper it has been used as the basis from which to argue that proponents of machine consciousness have significant hurdles to overcome to prove an assertion that a machine consciousness can be seen to be a legally responsible entity. Interestingly enough, it is possible that while this view of law may affect machine consciousness design, it is equally likely that if such design is ultimately successful, we may have to revisit some of the basic premises of law.

Why take this exposition as the starting place of our analysis? The answer is straightforward. Gazzaniga's and Morse's presentations were commissioned for a conference organized by the American Association for the Advancement of Science on Neuroscience and Law. In September 2004, Morse presented his ideas to the President's Council on Bioethics. Gazzaniga was a member of the President's Council. Consequently, it seems to me that the ideas expressed are positioned in such a way that they can uniquely influence the future course of deliberations at the highest level of government, at least in the United States. By stating that neuroscience can never affect law because law is based on a humanly constructed concept of responsibility, a bias is created against any reductive materialist argument to the contrary. How the question is framed often defines the answer received. My purpose here is to change the frame of reference slightly. Rather than look at how things exist in the world today, thereby forcing the debate into the terms set forth by Morse, ask whether we can, in the present state of knowledge, posit a scenario where the skepticism expressed can be tested

differently. That is, not by looking backward in evolutionary time and trying to decide why humans have laws and chimpanzees don't (Morse, 2004a), but rather to look forward and to determine if there is a set of conditions which, if they came to pass, would plausibly require us to reevaluate our position in the world and the relevance of law to that position.

Chapter 7

LAW AND FOLK PSYCHOLOGY

My limited contention here is that folk psychology gives us a starting point from which to begin a defense of the thesis of this paper. While it may not be easy to determine whether the various aspects we will discuss are necessary and sufficient for the minimum requirement of legal personhood, it is possible to get a sense of what would be acceptable to most of us, at least in a general way, if we were asked what it means to be a legal person. Through this type of inquiry we can begin to identify some of the common-sense platitudes which will form the basis of our functional definition of a legal person. Certainly under some theories of law, such as positivism (Kelsen, 1967), it is logically possible to argue that law could simply define a legal person to be anything the law makers choose it to be, much like Humpty Dumpty in <u>Through the Looking Glass</u>, "nothing more and nothing less". If, rather than being viewed as a closed system which makes up its own rules and simply applies them to its objects, law was viewed as a limited domain which, drew upon factors outside the law to define its concepts (Schauer, 2004), we could articulate the concept of a person by using folk psychology platitudes which are related more to function. This point will become clearer in the subsequent discussion.

The thesis under examination implies that so long as the non-biological machine has a set of physical/behavioral states deemed relevant to law, such as

autonomy and intentionality, then, if we could infer from those states a set of mental states, the non-biological machine could be a legal person with independent existence separate and apart from its origins as property. Given the wide range of entities and the variety of types of conduct that the law has brought within its scope, we need to identify those aspects of what Leonard Angel called "functional simimorphy". (Angel, 1989) Certainly there is just this type of relationship when we look at corporations, and I suggest that nothing we have seen so far requires us to categorically rule out non-biological entities from the equation.

If we accept that looking at folk psychology is a legitimate exercise in order to ascertain whether an artificial entity can by definition be viewed as a legal person, then we can further speculate on the effects of this attribution in light of our premise concerning functionalism. Let us begin by looking at what acceptance of the idea that law follows the folk psychology model might suggest about intentionality.

7a. Intentionality

There are at least two meanings of the word "intentionality", and those meanings should not be confused because they are not the same. From a philosophical point of view, starting with Franz Brentano (Brentano, 1924/1973) and continuing through later commentators, the philosophical idea of intentionality has referred to the ability of our thoughts to be about something or to represent something. For example,

the sentence, “The White House is in Washington D.C.”, is a sentence about the location of the White House, and also a statement about a feature of Washington D.C., namely that it is the site of the White House. John Searle has stated that “Intentionality is that feature of the mind by which mental states are directed at, or are about or of, or refer to, or aim at, states of affairs in the world.” (Searle, 1999) Further, there is no necessary connection between intentionality in this sense and consciousness. One can have an intentional belief even while one is asleep. Even Churchland (1988, p. 63), the main proponent of eliminative materialism, makes this point of distinction. “In the technical vocabulary of philosophers, such states (propositional attitudes) are said to display intentionality, in that they ‘intend’ something or ‘point’ to something beyond themselves: … (A caution: this use of the term “intentionality” has nothing to do with the term “intentional” as meaning “done deliberately”.)”

It is this latter usage mentioned by Churchland which defines intentionality in the vernacular of the law, and, as we will see in a moment, it is also what is meant by intentionality in folk psychology.

> The law clearly treats people as intentional agents and not simply as part of the biophysical flotsam and jetsam of the causal universe. … (L)aw and morality are systems of rules that at the least are meant to guide or influence behavior … They operate within the domain of practical reason. …

> All things being equal, intentional action or forbearance is the only aspect of the human condition that is fully "up to us," that is fully within our control, and that can be fully guided by and produced by our reason. Morse, 2004a, (p. 367 - 368)

From this we see that intentionality in the legal sense is concerned more with the concept that people act for reasons which they themselves control. In other words, they act deliberately.

According to the folk psychology conception, the direct cause of an "intentional action" is an "intention". An intention is action directed. It is based, in a hierarchical relationship, upon two inter-related but different concepts, a desire for an outcome, and a belief about the consequences of the act before it takes place. Both components are necessary to form an intention, so in this sense the intention is derived from the presence of desire and belief, but they are not sufficient for us to take the next step and ascribe intentionality on the strength of the existence of an intention alone. Similarly, desire alone, or belief alone is not sufficient to give rise to the intention. In order for an action to be performed intentionally, the intention has to be present as we have noted, but more is required. In order for there to be intentionality, there must be an intention coupled with action and accompanied by the skill to perform the act and awareness that the act is being performed. The awareness component specifies the agent's state of mind at the time of acting

(knowing what he or she is doing), and the skill component refers to the agent's ability and skill to perform the action he or she intends. (Malle and Knobe, 1997) (Malle and Nelson, 2003).

It is important to note that desire differs from intention in that intention represents a mental state which leads to an action. It is based on reasoning and involves commitment to act whereas desire is not based on reasoning or commitment. We see from this that it is the intention that is relevant for the legal system because it is the point at which acting begins. Action is the end result of the process of being a practical reasoner.

We need to be careful in this regard, because, as Malle and Nelson (2003) strongly suggests, the folk definition actually differs from the legal definition in large part because there is no single clearly understood legal definition which is consistent. The law in many places is hopelessly confused when it uses the concept of intention. In fact, some of the legal definitions are often at odds with the folk sense of intentionality. For our purposes this simply means that we cannot apply legal definitions of intentionality taken from statutes or cases without further understanding the components which make them up and the instances to which they refer. To do so runs the risk of creating further confusion in our effort to lay out the parameters of what it would take for a non-biological machine to become a legal person. However, because we are interested in the term in a theoretical sense we can

still use the common-sense definition to argue in favor of our claim that intentionality can serve as a component of legal personality. Again, a word of caution is in order. Recent work by Knobe (2003) and Nadelhoffer (2005) seems to point to the fact that a person's view of the morality of an action also has significant impact on the determination of whether it was intentional or not. This has interesting implications for any position which is based on a purely positivist view of law and may indicate that a natural rights view is more pervasive at a folk level. If this is the case, designers of non-biological machines may have to take this moral component into account as well. (Wallach and Allen, 2009).

If a non-biological machine can form an intention derived from a desire and a belief, and can act intentionally based upon that intention, there seems again to be no theoretical reason why it could not be viewed as a legal person by the average lay person applying common sense. But our intuition seems to tell us that simply saying this is not very convincing. We need to look further for examples to bolster our contention that a non-biological machine can exhibit intentionality such that it meets our criteria for legal personhood.

As previously noted, there has been considerable debate concerning the nature of the corporation as a legal entity. First it was viewed as a mere fiction, then justified on the basis that in reality it was an aggregation of individuals acting in concert. There is another suggestion which I will now examine to see if it can help us

to determine if intentionality has any bearing on how a non-biological machine might be treated before the law. If we set aside the distinction between a legal person and a moral person for the time being and focus more on the underlying conditions required for there to be any sort of "intentionality", we see that the topic has been the subject of some debate in the literature. Peter A. French is perhaps the person most noted for advocating the idea that a corporation is something more than a mere legal fiction, or an aggregation of human employees or shareholders. His view argues that the corporation has natural rights and should be treated as a moral person, in part because it can act intentionally. In this context, French uses the term "intentionally" in virtually the same sense as common-sense psychology. Thus, it offers some meaningful basis for comparison to the situation we are considering.

French's premise is that "… to be a moral person is to be both an intentional actor and an entity with the capacity or ability to intentionally modify its behavioral patterns, habits, or modus operandi after it has learned that untoward or valued events (defined in legal, moral or even prudential terms) were caused by its past unintentional behavior." French, 1984. (p. 165).

Needless to say, French is not without his critics. Donaldson (1982) argues from an Aggregate Theory stance that the corporation cannot have a single unified intention to act. He then goes on to argue that simply having intention is not enough to make the claim that the actor has moral agency, a position at odds with most of

the animal rights movement. Werhane (1985) carries this point further and, using the example of a computer system, argues that the appearance of intentionality does not necessarily mean that it acts out of real desires or beliefs. In other words intentionality does not imply that it is also free and autonomous. This is a point I will return to in the next section, for now I suggest that while I recognize Werhane's point, I disagree that such a system is impossible to construct. One example of a theory which could lead to just such a functional artificial agent is set forth in Pollock (2006). Further, drawing on Daniel Dennett's ideas concerning intentional systems, one can certainly argue that Werhane's position requires one to accept the premise that only phenomenological intentionality counts for moral, and perhaps legal purposes, but that does not appear to be supported by folk intuition. Likewise, critics of functionalism could make a similar argument against our thesis. For example, Frank Jackson's knowledge argument and David Chalmer's suggestion of the "hard problem', both rest on the premise that functionalism fails to account for the 'feel', or qualia which attend a mental state. But that does not appear to be supported by intuition when it comes to law. Functional intentionality is probably enough in a folk psychology sense to convince people that a non-biological system is acting intentionally. Solum suggests as much in the following language:

> How would the legal system deal with the objection that the AI does not really have "intentionality" despite its seemingly intentional behaviors? The case against real intentionality could begin with the observation that

> behaving as if you know something is not the same as really knowing it. … My suspicion is that judges and juries would be rather impatient with the metaphysical argument that AIs cannot really have intentionality. … Solum, 1992, (p. 1268).

If the complexity of the non-biological machine's behavior did not exceed that of a thermostat, then it is not likely that anyone would be convinced that it really possessed mental states—that it really believed things or desired things. But if our daily interaction was with such machines exhibiting symptoms of complex intentionality (of a human quality), the presumption might be overcome irrespective of any "phenomenal" feel in the machine itself. (Rivaud, 1992).

We can see therefore, that we cannot rule out the possibility that a non-biological machine can be developed which would meet the criteria of intentionality set forth above. I also suggest that Werhane's criticism concerning lack of autonomy could also be met by a properly designed machine. That is the topic to which I turn next.

7b. Autonomy and Responsibility

If asked whether humans are different from animals most people would say yes. Humans are generally believed to act according to their own intrinsic autonomy. While animals are recognized as having a similar level of autonomy they still are viewed as

controlled by instinct. People have little conceptual problem understanding that a cat, for example, "has a mind of its own", but is still not the same as a human. The same cannot be said initially about an android, even though as we will see later, the ascription of anthropomorphic characteristics to such creations can confuse the issue. Most people would contend that the android is a "mere machine", and, because this is the case, it needs to be programmed in some fashion by a human creator. This human programmer in turn has to determine the rules and regulations which will govern the conduct of the android both as it interacts with humans and also with other entities. (Wallach and Allen, 2009). Hugo deGaris (1989 and 1990), suggests that this may in fact not be possible if the android (in de Garis' terminology an "artilect") can truly evolve its way around any such constraints. One of the basic problems in ascribing rights and duties is this question of autonomy. Can the android actually be a moral subject if it is bounded by constraints imposed upon it by its "creator" or programmer? If the android is bounded by inherent, predetermined constraints, such as those contained in Asimov's well known Laws of Robotics, it is not truly free to make choices. (Asimov, 1950).

Autonomy has a number of potential meanings in the context of machine intelligence. Hexmoor, Castelfranchi, and Falcone (2003) draw a number of distinctions between the different types of interactions relevant to systems design and artificial intelligence. First there is human to agent interaction where the agent is expected to acquire and conform to the preferences set by the human operator. In their words, "(a) device is autonomous when the device faithfully carries the human's preferences and

performs actions accordingly.” Another sense is where the reference point is another agent rather than a human. In this sense the agents are considered relative to each other and essentially negotiate to accomplish tasks. In this view “(t)he agent is supposed to use its knowledge, its intelligence, and its ability, and to exert a degree of discretion.” In a third sense there is the idea mentioned before that the agent can be viewed as manipulating “…its own internal capabilities, its own liberties and what it allows itself to experience about the outside world as a whole.” Margaret Boden, in a similar vein, writes about the capacity of the agent to be original, unguided by outside sources. (Boden, 1996) It is in this third sense where I suggest that the term autonomy comes closest to what the law views as crucial to its sense of responsibility. In each of the first two senses of autonomy discussed above, there appears to be a referent to which the agent always defers in making its decision. In the first sense it is the human operator. Similarly in the second sense, it is the weighting or value placed upon the various decisions, weighting which is determined, not by the agent, but by the operator who is setting the conditions for the agent’s interactions with other agents. In each of these situations there appears to be a “controlling” entity which is setting the parameters of action. From a philosophical and legal sense this would strongly imply that the agent is not the competent causal agent of a consequence that has legal significance.

When pressed to describe what these distinctions imply in the context of legal rules, many people would respond that autonomy means that we are responsible for our actions because we have free will, that is, our actions are not predetermined. Note

however, that Morse (2004) argues that this is a mistake in that free will is not necessarily a criteria for responsibility in a legal sense. Outside of the legal contect, the debate has been couched in slightly different terms. In the view of the "incompatibilist", in order for people to be held responsible for their acts they must have freedom to choose amongst various alternatives. Without alternatives there can be no free will. (van Inwagen, 1983; Kane, 1996). The incompatibilist position has been strongly attacked by Harry Frankfurt, who called their argument the "principle of alternate possibilities". (Frankfurt, 1969/1988) Frankfurt has argued that it is possible to reconcile free will with determinism in his view of "personhood". His conclusion is that people, as opposed to animals or other lower order beings, possess first and second order desires as well as first and second order volitions. If a person has a second order desire it means that she cares about her first order desires. To the extent that this second order desire is motivated by a second order volition, that is, wanting the second order desire to be effective in controlling the first order desire, the person is viewed as being autonomous so long as she is satisfied with the desire. The conclusion is that in such a case the person is autonomous. (Frankfurt, 1971/1988)

It should be noted that in this context Frankfurt is using the term person as the equivalent of human. Others would argue that person is a broader term and more inclusive, drawing a clear distinction between person and human. (Strawson, 1959; Ayers, 1963) As is clear from the previous sections, my preference is to use the term

human to apply to homo sapiens and the term person to beings irrespective of species boundaries.

It is also helpful in this regard to compare Frankfurt's position with Kant's belief that autonomy is viewed as obedience to the rational dictates of the moral law. (Herman, 2002) Kant's idea that autonomy is rational differs from that of David Hume who argued that emotions are the driving force behind moral judgments. Hume seems to be an antecedent of Frankfurt's concept of "satisfaction" if the latter's essay on love is understood to equate "satisfaction" with "emotion". (Frankfurt, 1994/1999). Transposing these contrasting positions into the language used earlier to describe law, I suggest that it is possible to equate Frankfurt's sense of autonomy with the concept of responsibility. As discussed above with regard to intentionality, humans are believed to be freely capable of desiring to choose and actually choosing a course of action. Humans are believed to be capable of changing desires through the sheer force of mental effort applied in a self reflexive way. Humans are therefore, as practical reasoners, capable of being subject to law so long as they act in an autonomous way, they are responsible. Again, as in our earlier argument, I suggest that autonomy and responsibility are mental state terms which define the knowledge structure which relate the relevant input and output such that the functional role is sufficient for us to conclude that if a non-biological machine exhibited those behaviors, it would be deemed to be a legal person.

Let's look at this more closely. Assume for example that I program an agent so that it enters virtual space, say the Internet, to perform a task I set for it such as locating a particular set of documents. I give it various search criteria, i.e., set the search parameters, then leave it to its own devices in determining how best to accomplish the task and fulfill its duty. Assume further that the agent in fact proceeds to do as directed, but in the process commits what you and I and the world would view as an egregious harm. Perhaps in order to get the document it has to fraudulently represent to another agent that it is authorized to access a particular computer. Perhaps it determines that the best way to obtain the document requested is to copy it from a site where there is a charge for access. In order to avoid this fee, it manipulates another computer to access a third person's bank account to make the payment. Because the initial directions did not explicitly rule out these courses of action, the artifact is not constrained from following them. In each of these cases the law would have little difficulty in ignoring the "autonomy" of the agent and ascribing legal responsibility to the person who programmed the computer. As explained in Heckman and Wobbrock (1999), the law, using various well established rules such as strict products liability, or the law of dangerous instrumentalities, would have little difficulty in determining that the real actor in this scenario is the programmer, the person who sets the chain of action into motion.

As a further aside, space considerations preclude analysis of the meaning of the term "agent" as it is used in philosophy and in law. Suffice it to say that law has a technical understanding of the term agent which implies that the agent is directed and

controlled by a principal which may, if one is not careful, predispose one to conclude that the first two senses of autonomy are the only legally relevant ones. (ALI Restatement, Agency, 1958) Another topic beyond the scope of this paper, which provides a basis for speculation, is whether a philosophical agent can act morally without exhibiting free will, mental states or responsibility. (Floridi and Sanders, 2004)

In the third sense of autonomy mentioned above, the answer is not so straightforward. Change the above scenario slightly and assume that our initial point of departure is merely a stated desire we have to read a particular document. Our "friend", a conscious machine, hears our expression, and motivated by friendliness and social convention, decides to get the document for us as a birthday gift. Acting upon this determination the non-biological machine then proceeds to commit similar proscribed acts as mentioned above. Note that this sequence of activities would fulfill our intentionality criteria discussed previously. Here, I suggest, something more has happened, something more human like. If, in this scenario, the agent is autonomous in the sense described by Frankfurt, what I call the strong sense of autonomy, then it is conceivable that the law could directly affect the question of how we effectively evaluate such an artifact. If we adopt the strong definition of autonomy, and argue that if it is achieved in a machine, as it would be in the above example, then at least from a functional viewpoint, we could assert the machine is the equivalent of a human in terms of its being held responsible. As noted earlier, one would expect to be met with the objection that such a conclusion simply begs the question about whether the artifact is

phenomenally conscious. (Werhane, 1985; Adams, 2004) But once again, in the limited area we are examining we can put this argument to one side. For law, and for the idea of a legal person we are examining, it simply may not matter. Functional results are probably enough. Certainly there is a long history in the law of both animals and inanimate objects being put on "trial" for their various malfeasances. For a fascinating account of this phenomenon see generally Beirne (1994).

In our example, if the machine can conceive of a second order volition, the desire to be a good friend and to comply with social convention, and can as a result affect a first order action, the obtaining of the document, constrained only by the idea that one is satisfied by that result, does that not imply a functionally equivalent characteristic of human action? Going the next step, we can then argue that law acts at the level of this second order volition. It sets parameters which, as society has determined, outline the limits of an accepted range of responses within the circumscribed field which it addresses, say contract law or tort law or criminal law. This would imply that law acts in an exclusionary fashion in that it inhibits particular first order desires and takes them out of the range of acceptable alternatives for action. (Raz, 1975; Green, 1988). Note that this does not mean to imply that these are the only possible responses or even the best responses the actor could make. To the extent that the subject to which the law is directed, (the citizen within the control of the sovereign in Austin's terms), has access to law as normative information, she can order her desires or actions in accordance with law

or not. This would mean, to borrow the terminology of Antonio Damasio (1994), that the law sets the somatic markers[2] by which future actions will be governed.

By acting in a manner where its intentionality is informed by such constraints, and doing so in an autonomous fashion as just described, the artifact appears to be acting in a way which is functionally equivalent to the way we expect humans to act. I suggest that this does not require that the artifact have a universal, comprehensive understanding of the law any more than the average human does. Heuristics, or perhaps concepts of bounded rationality, could provide the basis for making decisions which are “good enough”. (Clark, 2003) Similar arguments have been advanced on the role of emotion in the development of a non-biological machine. (Sloman and Croucher, 1981; Arib and Fellous, 2004; Wallach, 2004) Perhaps, in light of work being done in how humans make decisions (Kahneman, Slovic, and Tversky, 1982; Lakoff, 1987; Pollock, 2006), more pointed analysis is required to fully articulate the claim concerning law’s normative role within the context of autonomous behavior. One further caution, even though I suggest that accepting law as a guide to a second order volition does not diminish the actor’s autonomy, this proposition can be challenged by some theories such as anarchism. (Wolff, 1970/1998)

---

2 Damasio’s “somatic marker hypothesis” states in essence that there are sensations that increase the accuracy of the decision making process. For my purpose here a “somatic marker” …”forces attention on the negative outcome to which a given action may lead, and functions as an automated alarm signal which says: Beware of danger ahead if you choose the option which leads to this outcome.” Damasio, 1994. (p. 173).

Ultimately the question comes down to whether this type of second order volition can be instantiated in a non-biological machine or whether it is exclusively the realm of biological species. Feigl (1953), Searle (1980) and others would say that it can only be accomplished in biological systems. Others however, have started to look at theoretical possibilities where just this type of activity can occur in non-biological systems. (Covrigaru and Lindsay, 1991; Clark, 2003; Holland, 2003; Wallach and Allen, 2009). In actively moving toward a complex non-biological machine it is possible that this question will be answered.

If the designers and builders of non-biological machines are correct, there will come a point where at least functionally speaking, a synthetically created being could achieve some of the attributes of intentionality and autonomy. At some point along a developmental continuum, the android will become a moral subject because of intrinsic qualities it possesses. Taking this point to the next level of abstraction we are presented with the possibility that such an android could plausibly be viewed as having the potentiality to become something more than mere "property", and in fact could be viewed as being a "subject-of-a-life". If, as argued by modern philosophers, particularly of the reductive materialistic view, humans are nothing more than electrical impulses in a biological substrate which are capable of being fully replicated so long as we use a sufficiently complex method of reproduction, is there any difference between the end result of a purely biological procreative process of life such as that which we humans now engage in, in our bedrooms, the process which takes place in a petri dish with

artificial insemination, or the development through emergence of an android based on mathematical algorithms?

If the views of analytic functionalism are accepted, then the functional role equivalent to that of human being can be envisioned to exist for a non-biological substrate. However, from a Catholic religious perspective, without an 'ensoulment' event, any such android would remain a non-human, incapable of ever assuming a position of moral equivalence with humans. 'If the human body takes its origin from pre-existent living matter, the spiritual soul is immediately created by God. Consequently, theories of evolution (or we might add, artificial, non-human entities) which, in accordance with the philosophies inspiring them, consider the spirit as emerging from the forces of living matter or as mere epiphenomenon of this matter, are incompatible with the truth about man'.[3]

From a functionalist point of view, and perhaps even that of a Shinto devotee, however, the premise may be different. If the android is functionally equivalent to a human in the way it acts and responds to stimuli, and exhibits behavioral characteristics associated with human beings, there is no reason to make a distinction between the biological and the non-biological substrate (Ryle, 1949; Kim, 1992).

An example of a system that is arguably functionally conscious is the software program IDA, described by Franklin (2003). According to Franklin, IDA's design parameters include factors such as perception, working memory, associative memory,

3 Address of Pope John Paul II to the Pontifical Academy of Sciences, 22 October 1996. Available online at: http://www.newadvent.org/library/docs_jp02tc.htm. See also, Encyclical of Pope Pius XII, Humani Generis (1950).

emotion and consciousness in the sense of awareness of factors to act on, action selection, constraint satisfaction, deliberation, voluntary action, negotiation, metacognition and learning. In a later article, Franklin (2005) suggested that the characteristics exhibited by IDA may even approach phenomenal consciousness. However, none of the characteristics he describes as existing at a functionally conscious level grant moral consideration to IDA. This point requires further development, but it implies that the delineation of functional characteristics sufficient for moral consideration is well beyond what has so far been achieved.

7c. An Ethical Aside

Moral consideration for an android could present itself as an issue the moment we begin to ascribe anthropomorphic characteristics to the android. Rodney Brooks (2002) has stated that he will know that he is getting close to his goal of building an intelligent robot when his graduate students feel guilty about turning it off. If we are able to create an android that is close to human in its capabilities and appearance, particularly one to which we are willing to cede significant areas of responsibility (e.g. cleaning, caretaking, childrearing), it is very easy to imagine the extension of certain of the moral considerations given to animals.

Although IDA's current characteristics seem insufficient to warrant moral consideration, as discussed above, they at least indicate fertile ground for empirical research into the myriad of questions this position raises. One of the first steps has

already been taken, and that is the development of an android that will allow researchers to study how humans interact with robots (Minato et al., 2004). If human participants respond to robots that speak (Arita et al., 2005) and exhibit gestures (Sugiyama et al., 2006) and facial characteristics that evoke emotional responses (Breazeal, 2002), perhaps we can begin to determine what functional characteristics people will find to be significant to the ascription of moral consideration. Is it empathy (Lee, 2006), language awareness, bipedalism, physical or psychic pain, or some other grouping of human isomorphic characteristics that are the determinant factors? Those identified aspects could then become the starting point for the design of an android. Progress needs to be made to understand exactly what the qualities people will accept as providing a basis for ascription of moral worth are at the various stages of android development (Kahn et al., 2006). This issue is treated in some detail by Ramey (2005a).

In theory, if one is creating an android by an emergent process (de Garis, 1990; Lindblom and Ziemke, 2006; Ziemke and Lindblom, 2006), at some point in that process, the android will reach a level of functioning that elicits the moral concerns we have for animals. Does this cross the line into morally unacceptable experimentation? Is it subject to legal constraints? We must be prepared to recognize that this effort may have unintended consequences. One of the best examples of how such concerns could find expression, and an example that sets forth some of the deep moral risks this would engender, appears in literature in the story of Dr Frankenstein. There, the creature he imbued with life expresses the same type of angst that one would expect to result from the creation of an android with anthropomorphic characteristics but no means to live its

life to the full. The ultimate demand of the ‘monster’was that he be provided with ‘a creature of another sex, but as hideous as [himself]’ (Shelley, 1817, p. 154). Since the middle of the twentieth century, the monster might have found legal redress under the Universal Declaration of Human Rights (Art. 16, UN Gen. Ass. Res. 217A III, 1948), or the European Convention on Human Rights (1953); both specifically protect the right to private and family life. Other examples are the subject of recent films such as *AI: Artificial Intelligence, Bicentennial Man* and the *Terminator* series. If an android with external human characteristics, perhaps even the ‘nannybot’ that raised us, exhibited this same range of activities, it is possible that most people would say that the creation has a claim to be treated morally. Again, much work needs to be done in this area because there is at least some evidence that people have negative reactions to robots that are too human in appearance, a phenomenon known as the uncanny valley (Mori, 1970; MacDorman and Ishiguro, 2006).

Another aspect of this possible unintended consequence derives from the fact that it is viewed as unethical to perform medical experiments on human participants without their clear and unequivocal informed consent. The judgment at the war crimes tribunal at Nuremberg following World War II established ten rules to be applied to experimentation on human participants. The first of these explicitly stated that the ‘voluntary consent of the human subject is absolutely essential. This means that the person involved should have legal capacity to give consent; should be so situated as to be able to exercise free power of choice, without the intervention of any element of force, fraud, deceit, duress, over-reaching, or other ulterior form of constraint or coercion; and should have sufficient

knowledge and comprehension of the elements of the subject matter involved as to enable him to make an understanding and enlightened decision'. (Trials 1949). Informed consent by a competent, autonomous individual is viewed as the cornerstone on which ethical experimentation must be grounded. (Smith, 1993; Garrison and Schneider, 2003). Until the latter part of the twentieth century, there were far fewer constraints on the ability to perform medical experiments on persons with diminished capacities, such as those with serious mental problems, or operating under some other form of legal disability, such as being a person confined to prison for a felony. Today, even in these formerly accepted areas, constraints exist (Frost and Bonnie, 2001). We recognize that in the case of animals, more specifically the higher-order mammals such as the great ape or the dolphin, current laws place strict constraints on experimentation. (Singer and Cavaleri, 1994). Arguments can be made that a wide range of attributes, sentience being only one of them, are sufficient to cause animals to be viewed as proper subjects of moral concern. What then are the implications of such a conclusion if the object of concern is an android that is in the process of being created from scratch? If the entities do not rise to the level of conscious machines, such as a special-purpose robot or industrial welding machine, which can 'think' enough to tell us when our car fender is properly affixed, then there is little to give us moral concern about how it is treated (Brooks, 2002). However, the more attributes the android has, or could potentially develop, the greater the risk that it could implicate moral concerns.

Even more fundamentally, there are concerns that arise at the earliest stages of development of a machine consciousness. The endeavor itself is replete with moral and

ethical pitfalls. If the same logic as urged for animal rights, or for the rights of fetuses, is applied to a machine consciousness, some of these issues could have the potential to curtail radically the development of a conscious entity. If part of the process of developing a machine consciousness is an emergent learning process (Lindblom and Ziemke, 2006), or even a process of creating various modules that add attributes of consciousness such as sentience, nociception, or language, in a cumulative fashion, some could argue that this is immoral. As posed by LaChat (1986), the question becomes 'Is the AI experiment then immoral from its inception, assuming, that is, that the end (telos) of the experiment is the production of a person? . . . An AI experiment that aims at producing a self-reflexively conscious and communicative "person" is prima facie immoral'. (p. 75–76). Must designers of a machine consciousness be aware that as they come closer to their goal, they may have to consider such concerns in their experimentation? Arguably yes, if human equivalence is the ultimate goal. Failure to treat a machine consciousness in a moral way could be viewed as a form of speciesism. (Ryder, 1975). The utilitarian philosopher J. J. C. Smart (1973) has observed 'if it became possible to control our evolution in such a way as to develop a superior species, then the difference between species morality and a morality of all sentient beings would become much more of a live issue'. (p. 67). It is worthwhile noting in this context, that following a strict interpretation of classical natural rights theory would lead to a conclusion that the creation and exploitation of a machine consciousness is allowed, so long as the machine consciousness is viewed as property and safeguards can be installed to assure that it does not become a "runaway." On the other hand, and perhaps more in

accord with modern thinking, if we assume that the machine consciousness has sufficient capability, either inherently, or because it is part of a group where the average mature individual has those characteristics, then it is conceivable to view the rights more from the perspective of liberal individualism and look to ascribe the machine consciousness, as an individual, with "natural rights". The logical extension of this position is to ascribe rights simply by virtue of the fact of the existence of a machine consciousness. In turn this attribution brings us back to LaChat's assertion of prima facie immorality (Calverley, 2003). I suggest that much more work, both at a theoretical level and a practical level, needs to be done in this area in order to safeguard against the possibility that any endeavor to develop a machine consciousness will be terminated on moral and ethical grounds before it can be empirically tested (Calverley, 2004).

In concluding this exposition, let's return to the question of legal perspective. If one has a bias toward legal positivism, then a likely outcome, if faced with the actuality of a machine consciousness, is to simply say that society will decide, hopefully after full and open debate taking into account the moral issues just mentioned, whether to ascribe a particular status to such an artifact. This response, much like the positions taken in the animal rights debate, is for the most part utilitarian based and pragmatic in its orientation. On the other hand if, as just noted, one has a bias toward a more natural law view, it is possible to argue that a machine consciousness has intrinsic rights. In either case we are left with the question of whether there is some legally relevant characteristic which, as Morse and Gazzaniga assert, is presently viewed as purely human but which we

can identify and articulate in such a way as to say that if that same characteristic is exhibited or possessed by another entity then that entity is entitled to be treated in an equivalent manner to humans.

What is necessary now is to pull these disparate strands together and to try to determine if we can present a theory which will allow us to articulate a coherent argument. In the following section I will set forth the foundation for an analogy between animal rights and android rights. By looking at this analogy we might then gather insights to further our argument.

Chapter 8

ANIMAL RIGHTS

In the Hellenic and Judeo-Christian tradition, in contrast to other religious and philosophical traditions such as Shinto, Jainism, Taoism and Buddhism, the predominant premise of philosophical discourse was that animals were inferior to humans (Rao, 2002). A similar contrast can also be drawn with the pantheistic view of native Alaskan people, the Koyukon (Nelson, 1983) and many other traditions. In the West, however, human beings have been described as the masters of the beasts since the Book of Genesis, which states that man was given dominion 'over the fish of the sea, the birds of the air, and the cattle, and over all the wild animals and all the creatures that crawl on the ground' (Genesis, Book 1, verse 26). While not all Western philosophical outlooks have accepted this position, it has found expression in the long-standing Judeo-Christian tradition, which holds that man is special and as a result has been given dominion over animals by God:

> Augustine, followed by Aquinas, accepted the Stoic view that animals can be killed, because, lacking reason, they do not belong in our community. There is an even more far-reaching conclusion in Thomas Aquinas... Intellectual understanding ... is the only operation of the soul that is performed without a physical organ, and infers that the souls of brute animals are not immortal like ours. Sorabji 1993. (p. 201).

As pointed out earlier the Catholic Church continues to hold that the distinguishing feature between man and animal is ensoulment. Driven by this belief, the subservient position of animals as property has remained the primary Western belief for millennia. Because animals were nothing more than property, their moral status elicited little debate or controversy. For example, the champion of individual human liberties and natural rights, John Locke, stated that even though we all, as humans, belonged to God and were God's creatures, we had ownership and control over our own bodies and the labor we performed. When this labor combined with an object of nature, we gained a right to that object in the form of a property right, which could be defended against the attempts of others to take away that right from us. This 'natural right' of ownership extended to animals because they were part of nature and, therefore, could properly be viewed as resources to be used by humans and converted to property (Locke, 1689b/1975; Waldron, 1988).

Rene Descartes stated that animals were no different from machines. If a man were to hit a dog with a hammer, it would be of the same moral significance as hitting a clock with a hammer. Issues might arise if the clock broke and injured a bystander, or if the clock were owned by that bystander, but in so far as the clock, or the dog, was concerned, no moral issues arose. He also believed that language distinguished human beings from animals:

> …there are no men so dull and stupid, not even idiots, as to be incapable of joining together different words, and thereby constructing a declaration by which to

> make their thoughts understood; and that on the other hand, there is no other animal, however perfect or happily circumstanced, which can do the like.
>
> . . . And this proves not only that the brutes have less reason than man, but that they have none at all: for we see that very little is required to enable a person to speak. . . it is incredible that the most perfect ape or parrot of its species, should not in this be equal to the most stupid infant of its kind or at least to one that was crack-brained, unless the soul of brutes were of a nature wholly different from ours. Descartes 1637.

A similar view has been expressed by both Frey (1980) and Macphail (1998); but this raises the question of whether it is the inability to use language comprehensible to humans that is the critical factor in this argument. Recent studies have shown that animals do use limited forms of language, but in a much different way from humans. (Friend, 2004).

Immanuel Kant held that the moral status of animals was derived from their relationship to humans. For Kant, the causing of injury to an animal had moral significance if it lessened one's likelihood of treating humans properly:

> ...since all animals exist only as means, and not for their own sakes, in that they have no self-consciousness, whereas man is the end, such that I can no longer ask: Why does he exist?, as can be done with animals, it follows that we have no

> immediate duties to animals; our duties toward them are indirect duties to humanity. Since animals are an analogue of humanity, we observe duties to mankind when we observe them as analogues to this, and thus cultivate our duties to humanity. . . so if the acts of animals arise out of the same principium from which human actions spring, and the animal actions are analogues of this, we have duties to animals, in that we thereby promote the cause of humanity. Kant 1794. (p. 212)

In contrast to the indirect duty view of Kant, Jeremy Bentham viewed animals themselves as moral agents worthy of consideration. Bentham held this position despite his belief that there were no natural rights; but being the object of moral consideration did not imply that the object had rights. According to Bentham (1789), it was not whether animals could reason or talk, but whether they could suffer that made them the subjects of moral consideration. Some commentators have suggested that Bentham may have had a limited view of the role animals might play in moral life and may have been concerned only with the gratuitous infliction of suffering on animals (Francione, 2000). Nevertheless, Bentham was one of the first to espouse moral consideration for animals solely because of their intrinsic characteristics, and not as a derivative of human considerations. He also went further than others and stated that animals should be treated under what he termed the ‘principle of equal consideration’ as equally weighted with humans in the same situation.

Since the 1975 publication of Peter Singer's book, *Animal Liberation*, there has been a growing debate over whether animals have 'rights'. There have been substantial debates concerning which characteristics of animals should necessitate which 'rights'. Singer himself did not use the term 'rights'. Instead he developed a modified act utilitarian view based on the concept that animals have 'interests' that are sufficient to make them worthy of moral consideration. In contrast, another significant proponent of protections for animals, Tom Regan, explicitly claimed that animals do have rights, and these derive solely from the fact that an animal has inherent value; because the animal is a subject-of-a-life, it has inherent value and is therefore to be treated as worthy of moral consideration (Regan, 1983).

Bentham's idea of 'equal consideration'has been expressed in modern vernacular as follows:

> At a practical level, equal consideration for animals would rule out, most importantly, the routine overriding of animals' interests in the name of human benefit. While equal consideration is compatible with different ethical theories, it is incompatible—if extended to animals—with all views that see animals as essentially resources for our use. Equal consideration may be compatible with some use of animals (and perhaps even human—think of conscription) for human purposes. But unequal consideration implies that animals and humans have such fundamentally different moral standings that the two exist in a hierarchy in which

those at the top may regard those beneath as resources for bettering their own lives. DeGrazia 1996. (p. 47).

Animal rights lawyer and advocate, Gary Francione, argued for the basic right of an animal not to be treated as a thing. ‘The basic right not to be treated as a thing is the minimal condition for membership in the moral community. . . the basic right provides essential protections’ (Francione, 2000, p. 95). This is equated with the concept of ‘equal inherent value’ (compare Bentham), which is interpreted to mean that one human is equal at a moral level to any other human, though not necessarily for all purposes. ‘Although there is a great deal of disagreement about precisely what rights human beings have, it is clear that we now regard every human being as holding the right not to be treated exclusively as a means to the end of another. This is a basic right, and it is different from all other rights; it is a pre-legal right in that it is a necessary prerequisite to the enjoyment of any other right’ (Francione 2000, p. 93). This formulation is Kantian in its essence, but has many different connotations in this context.

What are the specific animal characteristics that motivate these writers to assert their positions, and can we, for our purposes here, learn anything from those factors? For Regan, subject-of-a-life means that individuals ‘have beliefs and desires; perception, memory, and a sense of the future, including their own future; an emotional life together with feelings of pleasure and pain; preference and welfare-interests; the ability to initiate action in pursuit of their desires and goals; a psychophysical identity over time; and an

individual welfare in the sense that their experiential life fares well or ill for them, logically independent of their utility for others and logically independent of their being the object of anyone else's interests'. (Regan, 1983, p.243). Similar listings of characteristics arguably possessed by animals include 'nociception [sensing pain stimuli], consciousness, pain, distress, fear, anxiety, suffering, pleasure enjoyments, happiness, desires (and conation generally), concepts, beliefs, thinking, expectations, memory, intentional action, self-awareness, language, moral agency and autonomy', and have all been considered mental states to be used as measures of whether animals should be ascribed rights. (DeGrazia, 1996, p.96).

In Francione's view, sentience, in the sense of the capacity to have perceptual experiences, particularly pain and pleasure, is sufficient to invoke the full panoply of moral protections for an animal species; but this only means the animals are moral subjects, it does not mean they have the status of moral agents, for it is absurd to think that a well-fed cat could be criminally charged with the equivalent of manslaughter for killing a mouse, not for food, but simply for 'pleasure'. Mary Midgley (1983) has noted that there is nothing inconsistent with treating animals differently from one another or from humans as long as their intrinsic nature is safeguarded.

As we can see, there is a well-developed vocabulary for assigning moral consideration to animals. This does not extend to expanding their status to that of legal persons. For example, while numerous anti-cruelty statutes exist, these do not extend to

allowing the animal to assert rights on its own behalf. Rather, they seek to constrain human action and are based on Bentham's humane treatment principle, which recognizes the moral significance of animals without addressing their interests as legal persons. This, according to Francione (2000, p.54), is because animals are still viewed as property.

The debate concerning animal rights extends from the concept that animals have no inherent rights because they are merely property, to the view that they are entitled to have their interests considered and weighed when making moral choices, to the position that they are entities worthy of full moral consideration because they are subjects-of-a-life and more than things. In sum, animal rights can be viewed as a reaction to the idea that animals are merely property to be exploited by humans as they see fit. Animals are now viewed as having rights or interests sufficient to cause us to ascribe to them moral weight, and they cannot be treated as commodities for man's use and benefit. The significance and scope of the particular characteristics required for this ascription are still not clearly formulated. Once established, they lead to treating animals as moral subjects, but do not necessarily lead to them being viewed as legal person.

Chapter 9

CONCLUSION

Our inquiry began with the question of whether analytic functionalism could serve as a theoretical foundation for our assertion that a non-biological machine could be found to be a legal person. Drawing on the fact that analytic functionalism allows for the possibility of multiple realization, and is therefore not limited as was the identity theory, I suggested that there is no reason to deny the possibility that a non-biological machine could exhibit mental states. Next, by relying upon "folk psychology internal" to provide a theoretical basis for the knowledge structures which form those mental states, I suggested that we could use common-sense terms to define those states. In order to apply this argument to the concept of a legal person it became necessary to determine what, if any, relevant distinctions existed between the related concepts of human and person. In making this inquiry we found that "person" can, and does, refer to more than humans. Particularly in the legal sphere, *person* is a concept quite able to embrace entities which are non-biological in nature. Next we saw that folk psychology provides guidance for us by showing that intentionality, in the sense of being an action oriented intention derived from desire and belief, could carry the weight, at least in a functional sense, of our argument. Finally, we saw that the idea of autonomy if viewed as a phenomenon involving higher order properties, much like a functional role, could act as a further common-sense platitude to support our attribution of mental states to a non-biological machine.

By resorting to an analogy, the question posed as the thesis for this discussion, can, not withstanding the divergence between animals and androids, be applied to the way in which animals are viewed as moral objects worthy of consideration, thereby providing is a meaningful way to look at androids. If the android exhibits characteristics which we are willing to recognize as imparting moral consideration to animals, then treating the android as a being with value, at least in Kant's terms, enhances our own self worth. To do something different would demean us as humans. If the development of the android is one which scales up from the dumb toy level to something which exhibits degrees of functional consciousness, then more nuanced consideration of the ethical implications will be warranted. I believe significant areas of careful research are required to understand exactly what are the qualities people will accept as providing a basis for ascription of moral worth at the various stages of android development.

A long tradition of moral, legal and ethical thought gives us some basis upon which to begin to ground speculation concerning the possible attribution of legal and moral rights to androids. Likewise, the development of the concept of animal rights gives us at least a theoretical construct around which we can begin to argue that androids should have similar rights. Once we begin that process however, we may find that these same threads lead to conclusions about the way to create an android that could give rise to forces bent on curtailing the process. Once people begin to consider the moral implications of the development process of an android, implications for other aspects of life loom large. Not only do we have to consider the impact upon humanity itself, but to

be internally consistent we have to consider the impact on the android; our creation which may plausibly argue, at some point in time, that it is a new form of being capable of being viewed as a moral subject. (Wallach and Allen, 2009). As we move forward, much work needs to be done to articulate the moral, ethical and legal parameters which apply to the concept of android science. The task therefore, is for the designers of non-biological systems to convincingly instantiate these ideas into an actual system. The ultimate goal is to design a set of criteria, which, if embodied in a non-biological machine, would support the contention that such an entity is a legal person. Nothing we have seen in the theories we have examined would automatically rule out the idea of a non-biological machine constituting a legal person. We can, therefore, conclude that the possibility of a non-biological machine having legal independence is theoretically possible if it rests upon the foundation of analytic functionalism, which is in turn based upon "folk psychology internal"

REFERENCES

Adams, W. (2004). Machine consciousness: Plausible idea or semantic distortion? *Journal of Consciousness Studies, 11*(9), 46 - 56.

American Law Institute (1958) *Restatement of the Law, Second: Agency 2d.* St Paul, Minnesota: American Law Institute.

Angel, L. (1989). *How to build a conscious machine.* Boulder: Westview Press.

Aquinas. T. Treatise on laws. Trl. Regan, R., Indianapolis, IN: Hackett Publishing.

Arib, M., and Fellous, J. (2004). Emotions: from brain to robot. *Trends in Cognitive Science 8* (12), 554 - 561. doi:10.1016/j.tics.2004.10.004

Aristotle. *The Nicomachean Ethics.* Trl. Ross, D., Oxford: Oxford University Press.

Arita, A., Hiraki, K., Kanda, T., and Ishiguro, H., (2005). Can we talk to robots? Ten-month-old infants expected interactive humanoid robots to be talked to by persons. *Cognition, 95*, B49– B57. doi:10.1016/j.cognition.2004.08.001

Armstrong, D. (1981). The causal theory of mind. In Chalmers, D., (Ed), *Philosophy of Mind*, (pp. 80 – 87). New York: Oxford University Press (2002).

Asimov, I., (1950). *I Robot.* London: D. Dobson.

Austin, J. (1832/1955). *The province of jurisprudence determined.* London: Weidenfeld and Nicholson. (Original work published 1832)

Ayers, A. J. (1963). *The Concept of a Person*. New York: St. Martin's Press.

Beirne, P. (1994). The law is an ass: Reading E. P. Evans' *The Medieval Prosecution and Capital Punishment of Animals*. *Society and Animals: Journal of Human-Animal Studies, 2*(1), 27 – 46.

Bentham, J. (1789). *An introduction to the principles of morals and legislation*. ed. by J.H. Burns and H.L.A. Hart. London: Methune [1982] (c1970).

Bentham, J. (1824). Anarchical fallacies. http://www.law.georgetown.edu/faculty/events/userfiles/file/LawPhilosophy/Spring2008/Bentham_Anarchical_Fallacies.pdf. L-6.

Boden, M. (1996). Autonomy and artificiality In M. Boden (Ed.). *The philosophy of artificial life,* (pp. 95 – 110). Oxford: Oxford Univ. Press.

Braddon-Mitchell, D. (2004). Folk theories of the third kind. *Ratio (new series)* 17 (3), 277 – 293.

Breazeal, C. (2002). *Designing sociable robots*. Cambridge, MA: MIT Press.

Brentano, F. (1973). *Psychology from an empirical standpoint*. (Rancurello, A., Terrell, D., and McAlister, L., Trans.). London: Routledge and Kegan Paul. (Original work published 1924)

Brooks, R. (2002). *Flesh and machines: How robots will change us*. New York: Pantheon Books.

Calverley, D. (2003). Imagining rights for an artificial intelligence. unpublished paper presented at "Transvision 2003", Yale University, June 2003.

Calverley, D. (2004) Ethical Implications of Artificial Intelligence Design. In *Cognitive, Emotive and Ethical Aspects of Decision Making in Humans and Artificial Intelligence Vol. III,* (ed. Smit, I., Wallach, W. and Lasker, G.) Windsor Canada: IIAS. 33 - 36

Calverley, D. (2005a). Towards a method for determining the legal status of a conscious machine. In *Proceedings of the AISB 2005 Symposium on Next Generation Approaches to Machine Consciousness: Imagination, Development, Intersubjectivity, and Embodiment*. 75 – 84.

Calverley, D. (2005b). Can a non-biological machine become a legal person. In *Cognitive, Emotive and Ethical Aspects of Decision Making in Humans and Artificial Intelligence. Vol. IV*, Windsor, Canada: IIAS. 13 – 19.

Calverley, D. (2005c). Additional thoughts concerning the legal status of a non-biological machine. In Symposium on Machine Ethics, AAAI Fall Symposium 2005. Menlo Park, CA: AAAI Press. 30 – 37

Calverley, D. (2006) Android science and animal rights, does an analogy exist. *Connection Science 18*(4), 403 – 417. DOI:10.1080/09540090600879711

Calverley, D. (2008) Imagining a non-biological machine as a legal person. *AI and Society 22*(4), 523 – 538. DOI 10.1007/s00146-007-0092-7

Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies (2)*, 200 - 219.

Churchland, P. (1988). *Matter and consciousness*. Cambridge, MA: The MIT Press.

Cicero. *The Laws.* Trl. Rudd, N. Oxford: Oxford University Press.

Clark, A. (2003). Artificial intelligence and the many faces of reason. In Stich, S., and Warfield, T. (Eds.), The Blackwell Guide to Philosophy of Mind. Malden MA: Blackwell Publishing.

Covrigaru, A., and Lindsay, R. (Fall 1991). Deterministic autonomous systems. *AI Magazine 12* (3), 110 – 117.

Daly, I. (2010, February 24). Just like mombot used to make. *The New York Times*, D1.

Damasio, A. (1994). *Descartes error*. New York: Harper Collins Publisher.

Davis, M., and Naffeine, N. (2001). *Are persons property?* Burlington, VT: Ashgate Publishing.

de Garis, H. (Spring 1989). What if AI succeeds: the rise of the 21st century artilect. *AI Magazine 10* (2), 16–22,

de Garis, H. (1990). Moral dilemmas concerning the ultra intelligent machine. *Revue Int. Phil., 44*, (172), 131–138.

DeGrazia, D., (1996). *Taking animals seriously.* New York: Cambridge Univ. Press.

Descartes, R. Discourse on method, chapter 5. Available online at: www.literature.-org/authors/descartes-rene/reason-discourse/chapter-05.html

Donaldson, T. (1982). *Corporations and morality.* Englewood Cliffs: Prentice Hall, Inc.

Dworkin, R. (1978). *Taking rights seriously (revised).* London: Duckworth.

Faiola, A. (2005, March 11) Humanoids with attitude. Washington Post Foreign Service.

Feigl, H. (1958). The “mental” and the “physical”. In Chalmers, D., (Ed), *Philosophy of Mind,* (pp. 68 – 72). New York: Oxford University Press (2002).

Finnis, J. (1980). *Natural law and natural rights.* Oxford: Clarendon Press.

Floridi, L., and Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines, 3*(14), 349-379. Retrieved from http://www.wolfson.ox.ac.uk/~floridi/:

Francione, G. (2000). *Introduction to animal rights*. Philadelphia: Temple University Press.

Frankfurt, H. (1969/1988). Alternate possibilities and moral responsibility. In *The importance of what we care about.* (pp. 1 – 10). Cambridge: Cambridge Univ. Press. (Original work published 1969)

Frankfurt, H. (1971/1988). Freedom of the will and the concept of a person. In *The importance of what we care about.* (pp. 11 – 25). Cambridge: Cambridge Univ. Press. (Original work published 1971)

Frankfurt, H. (1994/1999). Autonomy, necessity and love. In *Necessity, volition and love.* (pp. 129 – 141). Cambridge: Cambridge Univ. Press. (Original work published 1994)

Franklin, S. (2003). IDA: a conscious artifact. *Journal of Consciousness Studies. 10*, (4 – 5), 47–66.

Franklin, S. (2005). Evolutionary pressures and a stable world for animals and robots: a commentary on Merker. *Consciousness and. Cognition. 14*, (1), 115–118.

French, P. (1984). *Collective and corporate responsibility.* New York: Columbia University Press.

Frey, R. G. (1980). *Interests and rights: The case against animals.* Oxford: The Clarendon Press.

Friend, T. (2004). *Animal talk.* New York: Free Press.

Frost, L., and Bonnie, R. (2001). *The evolution of mental health law.* Washington, DC: American Psychological Association.

Fuller, L. (1958). Positivism and fidelity to law- a response to Professor Hart. *71 Harvard L. Rev. 630* .

Fuller, L. (1969). *The Morality of Law (2nd. Ed.).* New Have: Yale University Press.

Garrison, L., and Schneider, C. (2003). *The law of bioethics: Individual autonomy and social regulation.* St. Paul: Thomson-West.

Gazzaniga, M., and Steven, M. (2004). Free Will in the twenty-first century: A discussion of neuroscience and the law. In *Neuroscience and the Law.* (pp. 51 – 70). New York: Dana Press.

Goodenough, O. (2001). Mapping cortical areas associated with legal reasoning and moral intuition. *41 Jurimetrics J. 429.*

Gray, J. C. (1909). *The nature and sources of the law.* Retrieved from http://books.google.com/books?id=G3MqAAAAYAAJ&printsec=frontcover&dq=nature+and+sources+of+the+law&source=bl&ots=vjbvKw9x8g&sig=uFb6d9vPIwIv0lhJkPKnaDuFP7A&hl=en&ei=Jt9ZTbuXG8OC8gbm_LGnBw&sa=X&oi=book_result&ct=result&resnum=4&ved=0CDYQ6AEwAw#v=onepage&q&f=false. p. 27.

Green, L. (1988). *The authority of the state*. Oxford: Clarendon Press.

Grotius, H. (1625). *De Jure Belli ac Pacis Libri Tres.* (F. Kelson, Trans.). Oxford: Clarendon Press.

Guizzo, E. Geminoid F: Hiroshi Isiguro unveils new smiling female android. available online at http://spectrum.ieee.org/automaton/robotics/humanoids/040310-geminoid-f-hiroshi-ishiguro-unveils-smiling-female-android.

Hart, H. L. A. (1958). Positivism and the separation of law and morals. *71 Harvard L. Rev. 593.*

Hart, H. L. A. (1961). *The concept of law*. Oxford: Clarendon Press

Heckman, C., and Wobbrock, J. (1999). Liability for autonomous agent design. *Autonomous Agents and Multi-agent Systems.* 2 (1), 87 – 103.

Herman, B. (2002). Bootstrapping. In Buss, S., and Overton, L., (Eds.), *Contours of agency.* (pp. 253 – 274). Cambridge, Mass: The MIT Press.

Hexmoor, H., Castelfranchi, C., and Falcone, R. (2003). A Prospectus on agent autonomy. In Hexmoor, H., (Ed.) *Agent Autonomy*, (pp. 1 – 10). Boston: Kluwer Academic Publishers.

Holland, O., and Goodman, R. (2003). Robots with internal models. *Journal of Consciousness Studies 10* (4 - 5), 77 - 109.

Inayatullah, S. (2001). The rights of your robots: Exclusion and inclusion in history and future. Available online at http://www.metafuture.org/Articles/rights-of-your-robots.htm

Ishiguro, H. (2006). Android science: conscious and subconscious recognition. *Connection Science, 18*(4), 319 – 332.
DOI: 10.1080/09540090600873953

Jones, O. (1997). Evolutionary analysis in law: An introduction and application to child abuse. *75 N. Carolina L. Rev. 1117.*

Kahn, P. H., Jr., Ishiguro, H., Friedman, B., and Kanda, T. What is a human?—Toward psychological benchmarks in the field of human–robot interaction. In Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 2006.

Kahneman, D., Slovic, P., and Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge Univ. Press.

Kane, R. (1996). *The significance of free will.* New York: Oxford Univ. Press.

Kant, E. (1785/1981). *Grounding of the metaphysics of morals.* J. Ellington Tr., Indianapolis: Hackett.

Kant, E. (1794/1996) *Lectures on ethics.* Heath, P., (Ed.), Schneewind, J. B., (Tr.), New York: Cambridge University Press, 1794/1996.

Kelsen, H. (1967). *Pure theory of law.* Knight, M., (Tr.), Berkeley: Univ. Calif. Press.

Kim, J. (1992). Multiple realization and the metaphysics of reduction. In Chalmers, D., (Ed), *Philosophy of Mind,* (pp. 135 – 149). New York: Oxford University Press (2002)

Knobe, J. (2003). Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology, 16* (2), 309 – 324.

Lakoff, G. (1987). *Women, Fire and Dangerous Things: What Categories Reveal About the Mind.* Chicago: Univ. Chicago Press.

LaChat, M. (Summer 1986). Artificial intelligence and ethics: an exercise in the moral imagination. *AI Magazine, 7*(2), 70–79.

Lee, B. (2006). Empathy, androids and 'authentic experience'. *Connection Science, 18* (4). 403–417. DOI:10.1080/09540090600868847

Lewis, D. (1972). Psychophysical and theoretical identifications. In Chalmers, D., (Ed), *Philosophy of Mind.* (pp. 88 – 93). New York: Oxford University Press (2002).

Lindblom, J., and Ziemke, T. (2006). The social body in motion: cognitive development in infants and androids. *Connection Science 18* (4), December 2006, 403–417. DOI: 10.1080/09540090600868888

Locke, J. (1689a/1967). Two treatises of government. In *Two Treatises of Government: A Critical Edition with an Introduction and Apparatus Criticus by P. Laslett.* London: Cambridge University Press, 1967.

Locke, J. (1689b/1975). An essay concerning human understanding. P. Niddich, Ed., Oxford: Clarendon Press.

MacDorman, K. F., and Cowley, S. J. (2006). Long-term relationships as a benchmark for robot personhood. In *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN).*

MacDorman, K. F., and Ishiguro, H. (2006). The uncanny advantage of using androids in social and cognitive science research. *Interactive. Studies, 7*(3), 297–337.

McNally, P., and Inayatullah, S. (n.d.). The rights of robots: Technology, culture and law in the 21st Century. available at http://www.metafuture.org/Articles/TheRightsofRobots.htm

Macphail, E. (1998). *The Evolution of Consciousness.* Oxford: Oxford University Press.

Malle, B., and Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology, 33*, 101 – 121,

Malle, B., and Nelson, S. (2003). Judging mens rea: The tension between folk concepts and legal concepts of intentionality. *Behavioral Sciences and the Law, 21*, 563 – 580.

Mazlish, B., (1993). *The fourth discontinuity: The co-evolution of humans and machines.* New Haven: Yale University Press.

Midgley, M. (1983). *Animals and why they matter.* Harmondsworth: Penguin Books.

Minato, T., Shimada, M., Ishiguro, H., and Itakura, S. (2004). Development of an android for studying human–robot inter-action. In Orchard, R., Yang, C., and Ali, M. (Eds), *Innovations in Applied Artificial Intelligence: The 17th International Conference on Industrial and Engineering Applications of Artificial Intelligence.* Lecture Notes in Artificial Intelligence, (pp. 424 – 434). Berlin: Springer.

Mori, M. (1970). Bukimi no tani [the uncanny valley]. *Energy, 7*, 33–35.

Morse, S. (2004). New neuroscience, old problems. In *Neuroscience and the law.* (pp. 157 – 198). New York: Dana Press.

Morse, S. (2004a). Reason, results, and criminal responsibility. *Univ. of Illinois L. Rev. (2*), 363 - 444.

Nadelhoffer, T. (2005). Skill, luck, control, and intentional action. *Philosophical Psychology, 18*(3), 341 -352.

Naik, G. (2009, July 14). In search for intelligence, a silicon brain twitches. *The Wall Street Journal*, A14.

Nelson, R. (1983). *Make prayers to the raven.* Chicago: University of Chicago Press,

Note. (1987). The personification of the business corporation in american law. *54 U. Chicago L. Rev. 1441.*

Pollock, J. (2006). *Thinking about acting: Logical foundations for rational decision making*. New York: Oxford University Press.

Putnam, H. (1973), The nature of mental states. In Chalmers, D., (Ed), *Philosophy of Mind.* (pp. 73 – 79). New York: Oxford University Press (2002).

Ramey, C. H. (2005a). For the sake of others: the personal ethics of human–android interaction. In *Proceedings of the CogSci 2005 Workshop: Toward Social Mechanisms of Android Science.* 137–148.

Ramey, C. H. (2005b). The uncanny valley of similarities concerning abortion, baldness, heaps of sand, and humanlike robots. In *Proceedings of the Views of the Uncanny Valley Workshop, IEEE-RAS International Conference on Humanoid Robots*.

Rao, K. R. (2002). *Consciousness studies, cross-cultural perspectives.* Jefferson, NC: McFarland.

Ravencroft, I. (2004). Folk psychology as a theory. http://plato.stanford.edu/entries/folkpsych-theory/

Raz, J. (1975). *Practical reason and norms.* London: Hutchinson.

Regan, T. (1983). *The case for animal rights.* Berkeley, CA: University of California Press.

Rivaud, M. (1992). Comment: Toward a general theory of constitutional personhood: A theory of constitutional personhood for transgenic humanoid species. *39 UCLA L. Rev. 1425.*

Rorty, A. (1976). A literary postcript: Characters, persons, selves and individuals. In Rorty, A. (Ed.), *The identity of persons.* (pp. 300 – 323). Berkeley: University of California Press.

Ryder, R. (1975). *Victims of science: The use of animals in research.* London: Davis-Poynter,

Ryle, G. (1949). *The concept of mind.* London: Hutchinson.

Schauer, F. (2004). The limited domain of law. *90 Virginia Law Review 1909.*

Searle, J. (1980). Minds, brains and programs. *Behavioral and Brain Sciences, 3*, 417.

Searle, J. (1992). *The rediscovery of the mind.* Cambridge: MIT Press.

Searle, J. (1999). *Mind, language and society.* New York: Basic Books.

Shelley, M. (1817/1993). *Frankenstein.* New York: Barnes and Noble.

Singer, P. and Cavaleri, P. (1994). *The great ape project: Equality beyond humanity.* New York: St. Martin's Press.

Singer, P. (1975). *Animal liberation.* New York: Random House.

Sloman, A., and Croucher, M. (1981). Why robots will have emotions. *Proceedings IJCAI*, 197 – 202.

Smart, J. C. C. (1973). An outline of a system of utilitarian ethics. In Smart, J. C. C. and Williams, B., (Eds). *Utilitarianism, for and against.* Cambridge: Cambridge University Press.

Smith, G. P., II, (1993). *Bioethics and the law.* Lanham, MD: University Press of America.

Solum, L. (1992). Legal personhood for artificial intelligences. *70 North Carolina L. Rev. 1231.*

Sorabji, R. (1993). *Animal minds and human morals.* Ithaca, NY, Cornell University Press.

Stich, S. and Nichols, S. (2003). Folk psychology. In Stich, S. and Warfield, T., (Eds.), *The Blackwell guide to philosophy of mind.* (pp. 235 – 255). Malden MA: Blackwell Publishing.

Stich, S. and Ravenscroft, I. (1994). What is folk psychology. *Cognition 50*, 447-468.

Stone, C. (1974). *Should trees have standing? Toward legal rights for natural objects.* Los Altos, CA: Kaufman.

Strawson, P. (1959). *Individuals.* London: Methune.

Sugiyama, O., Kanda, T., Imai M., Ishiguro, H., and. Hagita, N. (2006). Humanlike conversation with gestures and verbal cues based on a three-layer attention-drawing model. *Connection Science, 18,* (4), 403–417. DOI: 10.1080/09540090600890254

Trials of War Criminals before the Nuremberg Military Tribunals under Control Council Law No. 10, Vol. 2, pp. 181-182. Washington, D.C.: U.S. Government Printing Office, 1949.

Tur, R. (1987). The 'person' in law. In Peacock, A., and Gilbert, G., (Eds), *Persons and personality: A contemporary inquiry*, *(pp. 116 – 129).* Oxford: Basil Blackwell.

Tushnet, M. (1975). The American law of slavery, 1810 – 1860: A study in the persistence of legal autonomy. *Law and Society Review 10*, (1), 119 – 184.

van Inwagen, P. (1983). *An essay on free will.* Oxford: Oxford Univ. Press.

Waldbauer, J. and Gazzaniga, M. (2001). The divergence of neuroscience and law. *41 Jurimetrics J. 357.*

Waldron, J. (1988). *The right to private property.* NewYork: Oxford University Press.

Wallach, W. (2004). Artificial morality: Bounded rationality, bounded morality and emotions. In *Cognitive, Emotive and Ethical Aspects of Decision Making in Humans and Artificial Intelligence Vol. III*, (pp. 1 – 6). Windsor, Canada: IIAS

Wallach, W., and Allen, C. (2009). *Moral machines.* Oxford: Oxford University Press.

Warren, M. A. (1991/2000). Abortion. In Singer, P., (Ed). *A companion to ethics*, (pp. 303 – 314). Oxford: Blackwell Publishers (Original work published in 1991).

Werhane, P. (1985). *Persons, rights and corporations.* Englewood Cliffs: Prentice Hall Inc.

Wolff, R. P. (1970/1998). *In defense of anarchism.* Berkeley and Los Angeles: Univ. California Press. (Original work published 1970)

Ziemke, T., and Lindblom, J. (2006). Some methodological issues in android science. *Interactive. Studies 7*(3), pp. 339–342.