CORE SELF-AWARENESS AND PERSONHOOD

CORE SELF-AWARENESS AND PERSONHOOD

By

JEFF PERZ, B.A. Hons.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Arts

McMaster University

MASTER OF ARTS (2003) McMaster University

(Philosophy) Hamilton, Ontario

TITLE: Core Self-Awareness and Personhood

AUTHOR: Jeff Perz, B.A. Hons. (University of Toronto)

SUPERVISOR: Professor E. Gedge (Boetzkes)

NUMBER OF PAGES: x, 143

ABSTRACT

All beings who possess the capacity for core self-awareness are moral persons and ought to be legal persons. More specifically, I argue that core self-aware beings ought not to be used merely as a means. This moral prohibition ought to be legally enforced and such enforcement can only be effectively accomplished with legal personhood status. Moreover, the moral prohibition that core self-aware beings ought not to be used merely as a means constitutes the essence of moral personhood. This prohibition is defended with four mutually supportive justifications: Kantian ethics, utilitarianism, ethical empathism and a principle of equal consideration of interests.

The moral frameworks appealed to either support the thesis directly or do so after philosophically questionable elements have been removed form them. These frameworks are ultimately justified by an appeal to Aristotelian ethics. Although Aristotle concludes that only those who are capable of abstract rational contemplation can embody the good that is the proper subject of moral philosophy, it is briefly claimed within this thesis that Aristotle's undefended premises assume this conclusion. This claim regarding Aristotle's conclusion about rational beings is not defended herein and is left for a future work.

The thesis that all beings who possess the capacity for core self-awareness are persons, or ought not to be used merely as a means, is relatively rare in philosophical discourse. The present work is original because its essential claim is defended with a synergy of seemingly disparate traditional moral theories, a new moral theory and a principle of equal consideration of interests. It is a significant contribution to

philosophical knowledge because the question of who counts in ethics, or who is the proper subject of moral discourse, is fundamental to moral philosophy. An important political implication of this thesis is that non-human animals are persons.

ACKNOWLEDGEMENTS

I express deep gratitude for the incalculable help of my thesis supervisor Dr. Elisabeth Gedge and second reader Dr. Wilfrid J. Waluchow. Dr. Gedge witnessed the metamorphosis of this thesis from one that focused upon medical treatment decisions for mentally incompetent patients – including whether or not patients whose higher brain functions, including all forms of self-awareness, have been permanently destroyed are the valid recipients of scarce medical resources – to the present thesis that narrows its focus to ask who the proper subject of moral discourse is. I very much appreciate Dr. Gedge's complete ability to simultaneously make discerning objections and truly constructive suggestions. The forthright yet warm and sensitive way in which she voices her insightful criticisms was highly conducive to my productivity and reflects her amiable character. Dr. Waluchow's insistence on clarity, attention to detail and rigor of argument was likewise invaluable. I laud him for pushing me to make this thesis manifestly stronger than it would have been without his help. I am also grateful to my third reader, Dr. Spiro Panagiotou, whose objections during my defence have inspired me to further inquire into ancient Greek philosophy. I also thank the chairperson of my examining committee, Dr. Violetta Igneski, for making a very high-spirited defence flow smoothly. Special thanks go to Dr. David Sztybel for his comments on an earlier draft of this thesis. I give warm thanks to Carolyn Chau for providing moral support during my defence. For their stalwart integrity, passion and insights they have shared with me over the past three years, I

express my gratitude and deepest appreciation for Professor Gary L. Francione and Adjunct Professor Lee Hall.

I thank my parents, Aileen Perz and Norbert Perz, for providing me with some of the early resources to pursue higher education that eventually led to the writing of this thesis. I also thank my grandparents, Rudolf and Else Perz and Florence and George Leonard for their financial contributions to my education over the years.

Robust thanks go to Floyd Switzer, my 11th grade philosophy teacher, for originally instilling within me a strong love of wisdom and of moral philosophy in particular. I am fundamentally indebted to Rachel Ann O'Shea for watering the seed of my confidence and for providing me with the first knowledge I received that has allowed me to become more than I am. Without her, I very likely would have never begun to develop the character traits required for post-graduate work and life in general. Credit for the information and inspiration that has resulted in the essence and continued progression of my personal development rests squarely on the shoulders of Axel Molema, my Yoga teacher, and his teacher before him. The courage I relied upon to complete this project was largely the result of positive desire that was nurtured by him. I thank Axel for my continued aspiration toward cultivating positive character traits and eradicating negative ones.

Very special thanks go to Nathalie Hellot, who opened my mind and spirit to the possibility of, and gentle faith toward, recognizing self-awareness in whatever shape it may take.

DEDICATION

For

the 60 billion every year[†]*,*

179 million every day,

124 000 every minute,

2067 every second.

[†] According to the United Nations Food and Agriculture Organization, 47.9 billion non-human animals were killed for food in 2001 alone. This figure does not include some "non-slaughter" deaths that are generally not reported and does not include deaths in unprivileged countries that have no reporting procedure in place. This figure also does not include fish and other marine animals who were killed for food. Moreover, it does not include the non-human animals who were directly or indirectly killed for other purposes such as vivisection, (fur) clothing, (circus) entertainment, companion animal breeding and subsequent "shelter" killing and so on. Therefore, an extremely conservative estimate of the number of non-human animals who are killed as a result of human exploitation every year is 60 billion. According to Agriculture Canada, over 640 million non-human animals are killed for food in Canada every year. This figure does not include fish and other marine animals.

"Philosophy can lead the mind to water but only emotion can make it drink."[‡]

Tom Regan

[‡] Patrice Greanville, "The Search for a New Global Ethic," *Animals' Agenda*, December 1986, p. 40, Quoting Tom Regan.

CONTENTS

CHAPTER 1: Introduction

SECTION I: Thesis

All beings who possess the capacity for core self-awareness ought not to be used merely as a means. The definition of core self-awareness will be specified and defended in chapter two. This thesis is relatively rare in philosophical discourse, although there are three notable exceptions.[1] Due to the relatively unconventional nature of this thesis, it will be defended with a two-pronged approach. Chapters three through five will discuss relevant aspects of three different moral theories and argue that these theories either support the thesis directly or do so after philosophically questionable elements have been removed from them. The thesis will further be defended in chapter six by an appeal to a principle of equal consideration of interests that is found within all of the moral theories discussed. Chapter seven will discuss the implications of this thesis for legal and moral personhood.

Although I use Kantian language to frame this thesis, the arguments in support of it will not be based in the moral philosophy of Immanuel Kant. Kant maintains that his categorical imperative, including its second formulation which I have significantly altered above, stems from the nature of reason itself and hence only applies to rational beings—where reason is narrowly defined in terms of logical consistency and related concepts. In chapter three, I will argue that the basis for Kant's view, namely that the rational good

[1] Tom Regan, in *The Case for Animal Rights*, argues that all beings who are "subjects of a life" (e.g. human and non-human animals of one year of age and older) have the basic right not to be used as a mere means. Gary L. Francione, in *Introduction to Animal Rights*, argues that all sentient beings have the aforementioned right. David Sztybel, in *Empathy and Rationality in Ethics*, argues that all conscious beings have this right; see chapter five in this thesis.

will is good in itself and bridges the logical gap between *a priori* and *synthetic* claims of moral truth, is questionable. Despite having rejected the basis for Kant's moral theory, I will not abandon it entirely. In the spirit of principalism, as first formulated by Tomas Beauchamp and James Childress, I place value on "Kantian" ideas such as the importance of duty, intention and the interests of the individual in making moral decisions. So, in this limited sense, I will use "Kantian" like arguments to support this thesis. Moreover, in the following chapter, I will draw attention to existing support for my thesis found in classical utilitarianism. Like Kant's moral theory, it will be shown that the foundation upon which classical utilitarianism rests is ultimately not subject to philosophical "proof," but nevertheless has immense value. The tensions between these two different moral theories will not become relevant because I will limit myself to discussing cases of using others merely as a means that do not involve ethical dilemmas or true conflicts of interest. Hence, all things being equal, the two theories should be in agreement with each other with respect to the cases I consider. Importantly, it will also be argued that although classical utilitarianism is accurately described as "act" utilitarianism, it is—at the least—very conducive to the "rule" that one ought not treat others merely as a means.

My discussion of moral theory is admittedly non-exhaustive. It is beyond the scope of this thesis to offer an extensive critical analysis of each of the moral theories considered. *My purpose in touching on meta-ethics is strictly limited to showing that my relatively unconventional thesis regarding beings who have core self-awareness is compatible with two classical moral theories*. The *method* in Aristotelian ethics will be discussed in chapter three, section three, sub-section B and its *assumptions that are*

relevant to this thesis will be briefly discussed at the beginning of the same sub-section. A direct argument from feminist ethical theory will not be provided, as such arguments can be found in other works.[2] Moreover, I will also discuss an entirely new moral theory that includes beings who have core self-awareness that was developed by philosopher David Sztybel.

As alluded to above, after supporting my thesis with moral theory, I will suggest a principle of equal consideration of interests and argue that it applies to all those who possess core self-awareness. Next, I will attempt to connect this principle with the idea that those who are subject to it should not be used merely as a means. I will respond to relevant objections throughout. Finally, as a matter of practical implementation and not as a matter of philosophical argument, I will draw attention to work that shows that everyone who *morally* ought not to be treated merely as a means must be considered to be a *legal* person. Therefore, it will be concluded that all beings who possess core self-awareness ought to be legal persons. The distinction and relation between legal and moral matters will be discussed in the following section and in chapter seven. In my closing remarks, I will entertain the view that not only should those with core self-awareness be regarded as legal persons, but as moral persons as well.

[2] Carol J. Adams, *The Sexual Politics of Meat: A Feminist-Vegetarian Critical Theory*, 10th Anniversary ed. (New York: Continuum Publishing Corporation, 2000); Carol J. Adams and Josephine

SECTION II: The Distinction and Relation Between Law and Morality

A) Law and Morality as Complementary

Throughout this thesis, legal and moral status, equality, rights, standing and personhood will be discussed. In this discussion, I employ a positivist theory of law. Citing H.L.A Hart's classic work of analytic jurisprudence, *The Concept of Law*, Gary L. Francione notes that legal positivism maintains that if a given legal rule exists within an efficacious legal system and was adopted through the accepted process of that system such as being passed by the legislature or ruled by a court, then the law exists and is valid. Conversely, natural law theory has the additional requirement that the law must conform to some moral standard. Francione, Hart and positivists generally, however, maintain that a law can both be a valid rule of an efficacious legal system and either be moral or immoral, just or unjust.[3] In short, the theory of legal positivism adopted here maintains that there can be a contingent, but not necessary, connection between moral claims and legal rules.

The arguments in support of this thesis will be *moral* ones. To say that something is "*seriously or fundamentally*"[4] immoral is to say that it morally ought not to be done. To say that something seriously or fundamentally immoral ought not to be done is to say that it morally ought to be prohibited. The claim that something morally ought not to be done would lose much of its normative force if there were no moral prohibition or support for a

---

Donovan, eds., *Animals & Women: Feminist Theoretical Explorations* (Duke University Press, 1995); Carol J. Adams, *The Pornography of Meat* (New York: Continuum Publishing Corporation, 2003).

[3] Gary L. Francione, *Animals, Property and the Law* (Philadelphia: Temple University Press, 1995), p. 95.

[4] See sub-section B, below.

mechanism to enforce that moral prohibition. For example, the claim that rape and murder are immoral would lose much of its moral commitment if it were not advocated to be enforced by the law. Of course, the criteria for using the law to effect moral prohibitions extend beyond the recognition that certain legally prohibited actions are credibly argued to be seriously or fundamentally immoral. These criteria, such as issues of efficacy and proportionality, are background considerations to the essential *moral* claim that seriously or fundamentally immoral actions morally ought to be legally prohibited.[5] Hence, the positivist view that the law, *at its best*[6], should serve as the mechanism of enforcing moral claims such as rape and murder are wrong can be consistently held despite background considerations such as issues of efficacy and proportionality.

---

[5] For example, in chapter seven, it will be argued that the only way to enforce the moral claim made in this thesis is to accord all core self-aware beings legal personhood. At this point in history, doing so would not be legally efficacious in the case of core self-aware non-human animals because "For the most part, the law reflects social attitudes and does not form them. This is particularly true when the behavior in question is deeply embedded in the cultural fabric, as our exploitation of animals undoubtedly is. As long as most [human] people think that it's fine to eat animals, use them in experiments, or use them for entertainment purposes, the law is not likely to be a particularly useful tool to help animals. If, for example, Congress or a state legislature abolished factory farming, that would drive the cost of meat up and there would be a social revolt!" (Friends of Animals, "An Interview with Gary L. Francione on the State of the U.S. Animal Rights Movement," *Act·ionline*, Summer 2002, p. 29. and http://www.friendsofanimals.org/action/summer2002/summer2002garyfrancione.htm, Quoting Gary L. Francione) Thus, the law cannot efficaciously enforce the moral claim made in this thesis at this time in history. What is required in order for this to take place is a substantial shift in the widespread social attitudes that are embodied by the institutionalized practices of non-human animal exploitation. As will be argued in chapter seven, when a critical mass of humans reject the use of other animals as mere means to their ends, legal personhood status will be efficacious to enforce the moral claim argued for in this thesis. Issues of efficacy (and proportionality) may be relevant to different aspects of using the law to effect moral prohibitions but they have no bearing on the truth or falsity of the claim that seriously or fundamentally immoral actions morally ought to be enforced. Again, it will be argued in chapter seven that the only way to enforce the moral claim made in this thesis is with legal personhood status. As a preliminary but nevertheless entirely separate matter, the only way to efficaciously establish the legal personhood of certain core self-aware beings is to first bring about a non-violent social revolution in which a critical mass of humans reject their instrumental use.

[6] Of course, according to legal positivism, it is not necessarily the case that the law serves this

In light of the foregoing analysis, when I speak of "moral and legal status," "moral and legal equality," "moral and legal rights," "moral and legal standing" and "moral and legal personhood" or use those terms without qualifying them as moral or legal, I am not conflating moral and legal concepts. Rather, I am asserting *both* that someone has moral standing (or moral equality, rights, personhood, etc.) *and* ought to have legal standing *for the purpose of* enforcing that moral standing.

B) Limited Sense of Not Using Someone Merely as a Means

Throughout this thesis, I avoid the use of rights language unless a particular philosopher whom I am considering uses it. This is because moral and legal rights are ultimately justified by moral theories and it is simpler to discuss those theories directly rather than going through the medium of rights language. Rights language, however, is relevant at present because it helps to distinguish cases of using someone merely as a means that are "seriously or fundamentally" immoral from those that are not. In particular, basic rights are relevant in the former cases but not the latter.

Francione maintains that every human has the basic right not to be used as a mere means. This right is basic insofar that it is a necessary condition for the enjoyment of any other right. For instance, Francione argues that the right of free speech or liberty would be meaningless[7] without the basic right not to be used as a mere means.[8] He notes that this

---

function. Rather, there is a moral claim here that the law, at its best, *ought* to serve this function.

[7] By "meaningless," Francione presumably means that a non-basic right cannot be exercised with any degree of efficaciousness without a basic right, or that the former would not be useful or valuable without the latter. It might be objected that certain inmates of concentration camps, for example, may have exercised the non-basic right of free speech to some degree and thereby gained improvements in their living

view of basic rights is found within Immanuel Kant's concept of innate equality that "grounds our right to *have* [other] rights."[9] Moreover, Francione cites a similar view found in Henry Shue's *Basic Rights*. Shue holds that, although basic rights are not more valuable or intrinsically satisfying to enjoy than non-basic rights, the latter cannot be enjoyed if the former is sacrificed. Moreover, non-basic rights can be sacrificed in order to secure basic rights but the reverse is not possible. That is, Shue argues that the sacrifice of a basic right in order to secure a non-basic right would be self-defeating because the latter could not be enjoyed in the absence of the former. Francione notes that the most important basic right that Shue identifies is the basic right to physical security, which encompasses the rights not to be murdered, tortured, raped or assaulted.[10] "[I]f I have no

---

conditions. As such, the non-basic right to free speech may not be inefficacious or valueless in the absence of basic rights. If, however, the concentration camp inmates are being used merely as a means, then any "improvements" in their living conditions are instrumental to the efficiency of using them in this way. For example, the Nazis played classical music in their concentration camps because doing so was thought to lessen overall tension and insurgency and keep the lines to the gas chambers moving more smoothly. This present claim regarding non-basic rights not being valuable without basic rights is defended in chapter seven, section one. Notwithstanding this claim, Francione argues in another work that "proto-rights" are mechanisms that can protect the interests of individuals who are still being used as a mere means beyond the extent required to maximize their being exploited efficiently. Proto-rights are distinguishable from non-basic rights insofar that the former *necessarily* serve to chip away at the legal property status of the individual who is being used as a mere means. In other words, proto-rights interfere with the ability of an exploiter to use the exploited individual in a way that would maximize benefits to the exploiter. In this way, proto-rights are the "building blocks" of rights. Based upon essential aspects of rights theory, Francione explains and argues for five criteria that are necessary conditions for proto rights: an incremental change must constitute a prohibition, there must be a prohibition of an identifiable activity that is constitutive of the exploitative institution, the prohibition of a constitutive activity must recognize and respect a non-institutional interest, interests cannot be "tradable" and the prohibition should not substitute an alternative, and supposedly more "humane" form of exploitation. For example, an absolute prohibition on using concentration camp inmates as the non-consenting subjects of hypothermia research would qualify as a proto-right in a context in which they are still being used as a mere means. For a reasoned defence of this claim, see: Gary L. Francione, *Rain Without Thunder: The Ideology of The Animal Rights Movement* (Philadelphia: Temple University Press, 1996), pp. 190-219.

[8] Gary L. Francione, *Introduction to Animal Rights: Your Child or the Dog?* (Philadelphia: Temple University Press, 2000), p. 93.

[9] *Ibid.*

[10] *Ibid.*, pp. 93-94.

right to physical security and you have a right to kill me at any time, then my possession of the right to drive or to vote becomes meaningless."[11] Francione concludes that the minimal condition for membership in the moral community is having the basic right not to be treated as a thing: "if you are not going to be a thing that has no protected interests—then you cannot get less protection than this right affords."[12]

It might be objected that Francione's contention that non-basic rights in the absence of basic rights would be "meaningless"[13] and Shue's contention that these rights could not be "enjoyed" are vague. Presumably, these terms refer to the fact that non-basic rights cannot be effectively used or have any value without the presence of basic rights. The claim that the basic right to physical security is highly valuable or has great moral significance and the non-basic right to drive (*without an accompanying basic right*) is less so presupposes a moral standard that is used to distinguish between these rights. This distinction between basic and non-basic rights is ultimately supported by moral theories such as Kantian ethics, utilitarianism, Aristotelian ethics, feminist ethics and so on. All of these theories would no doubt agree on the simple point that the right to physical security is serious or fundamental, while the right to drive considered in itself is less so. Thus, following Francione and Shue, the former is a basic right and the latter is a non-basic right. If this thesis were expressed in rights language, it would state that all beings who possess the capacity for core self-awareness have the basic right not to be used merely as

[11] *Ibid.*, p. 94.
[12] *Ibid.*, p. 95.
[13] *Supra*, note 7.

a means.[14] The actual language of this thesis, however, is that all beings who possess the capacity for core self-awareness ought not to be used merely as a means, and this is to be understood in a basic or fundamental sense as described above.

[14] This is Francione's thesis in *Introduction to Animal Rights* (*Supra* note 8) regarding sentient beings, who Francione argues necessarily also possess core self-awareness. Francione bases his thesis on the assumption, taken from conventional wisdom, that "we may prefer human interests over animal interests, but that we may do so only when it is necessary and that we therefore ought not to inflict unnecessary suffering on animals." (Francione, *Introduction to Animal Rights*, *Op. cit.*, p. xxiii.) The present thesis is different from Francione's in that it neither makes this assumption at the outset nor bases subsequent arguments upon it.

CHAPTER 2: Core Self-Awareness

SECTION I: Definition

Self-awareness is the capacity to experience oneself as existing. This capacity can take many forms, including the familiar form of having psychological continuity through time, self-oriented beliefs and a concept of self that (insofar as it is a concept) can only be understood and expressed through language. In the next two sections, however, it will be shown that the capacity to experience oneself as existing requires neither psychological continuity, the possession of beliefs about oneself, a concept of self nor the capacity for language. Thus, the definition of core self-awareness that will be defended below is the bare capacity to experience oneself as existing. The constituent definitions of "experience," "oneself" and "existing" will be implicated in the following discussion of José Luis Bermúdez.

SECTION II: The Paradox of Self-Consciousness

A) Outline

The title of this section is taken from a book by Bermúdez. Bermúdez uses the terms "self-consciousness" and "self-awareness" interchangeably[15], although he mostly employs the former. Bermúdez maintains that what I call core self-awareness includes both what he calls primitive and more developed forms of self-consciousness, both of which can take different forms and exist at the non-conceptual, non-linguistic level. He

[15] José Luis Bermúdez, *The Paradox of Self-Consciousness* (Cambridge, Massachusetts: The MIT Press, 1998), p. 147, note 16. In this note, Bermúdez uses the term self-awareness when referring to the

argues that it is logically and empirically impossible for self-consciousness to necessarily require the possession of linguistic self-reference (and any concepts, beliefs, psychological continuity or self-knowledge that such self-reference might require). More specifically, Bermúdez argues that a paradox is created that makes self-consciousness itself impossible if it is assumed that conceptual and linguistic forms of self-consciousness are the only forms of self-consciousness. Bermúdez resolves the paradox by arguing and citing evidence for the existence of content-bearing, self-representational states that are possessed by beings who lack both conceptual and linguistic abilities. The states possessed by these beings can consist of primitive (or more developed) self-consciousness and are also possessed by beings who do have conceptual and linguistic abilities. Bermúdez then argues how non-conceptual self-consciousness makes the existence of conceptual self-consciousness possible.

Rather than undertaking an exhaustive exposition of Bermúdez's entire argument, I refer the reader to his excellent book. In order to lend further credence to this thesis, however, I will outline the arguments in favour of Bermúdez's paradox of self-consciousness in chapter one of his book in order to show that self-consciousness is impossible without the existence of core self-awareness as I have defined it. Bermúdez's solution to the paradox and his extensive positive arguments in favour of non-conceptual self-awareness are fascinating and can be better appreciated by reading his book in its entirety.

---

subsequent section in his book that uses the term self-consciousness.

B) The Traditional Account of Self-Consciousness

Bermúdez lists a number of different self-conscious states (e.g. self-knowledge, the capacity to make plans for the future, pursuing questions such as "who am I?" and so on) and entertains the idea that there is a single ability that they all presuppose; the ability to think about oneself. He notes that self-knowledge requires beliefs about oneself and that having these self-oriented beliefs obviously requires the ability to entertain thoughts about oneself. *Many* of the thoughts about oneself can also apply to other people and objects. For example, the knowledge that one is human deploys certain conceptual abilities that can also be deployed in thinking that others are human. Bermúdez notes, however, that the ability-to-think-about-oneself (considered as a distinct ability) cannot be used to think about other humans and objects: philosophy of mind discourse commonly refers to this ability as that required to entertain 'I'-thoughts.[16] Bermúdez adopts this convention and develops the following definition of 'I'-thoughts. The reader should take note that the following development of the definition of 'I' thoughts leads to a distinction of two different kinds of 'I'-thoughts. This distinction in turn leads to what Bermúdez terms the deflationary theory of self-consciousness. Although the arguments that lead up to the deflationary theory of self-consciousness appear to be entirely sound and valid, as does the theory itself, Bermúdez ultimately rejects it because it leads to a logical paradox. Bermúdez presents the following account of self-consciousness and shows how it accords extremely well with two predominate schools of philosophy of

[16] *Ibid.*, pp. 1-2.

mind. Nevertheless, he ultimately rejects this deflationary account of self-consciousness in favour of a truly valid account.

Bermúdez notes that an 'I'-thought is one that involves self-reference; it can only be thought by thinking of oneself. Not all thoughts that involve self-reference, however, are 'I'-thoughts. Bermúdez provides a cogent argument[17] that concludes that one cannot think 'I'-thoughts if one does not know that one is thinking about oneself. This suggests that 'I'-thoughts involve a distinctive form of self-reference whose natural linguistic expression is the first person pronoun 'I,' as one's use of the first person pronoun requires the knowledge that one is thinking about oneself.[18] Bermúdez proceeds to further explain this claim.

Bermúdez notes that when it is said that a thought has a natural linguistic expression, it is also being conveyed that it is appropriate to characterise the content of that thought in a certain way.[19] That is, something is being said about *what* is being thought and Bermúdez calls this "propositional content."[20] This suggests that thoughts whose propositional contents constitutively involve the first-person pronoun consist of 'I'-thoughts. Bermúdez, however, notes that this definition of 'I'-thoughts is not adequate

[17] "Suppose I think that the next person to get a parking ticket in central Cambridge deserves everything he gets. Unbeknownst to me, the very next recipient of a parking ticket will be me. This makes my thought self-referring, but it does not make it an 'I'-thought. Why not? The answer is simply that I do not know that I will be the next person to receive a parking ticket in central Cambridge. If *A* is that unfortunate person, then there is a true identity statement of the form I = *A*, but I do not know that this identity holds. Because I do not know that this identity holds, I cannot be ascribed the thought that I will deserve everything I get. And so I am not thinking a genuine 'I'-thought, because one cannot think a genuine 'I'-thought if one is ignorant that one is thinking about oneself." *Ibid.* pp. 2-3.

[18] *Ibid.*

[19] *Ibid.*, p. 3.

[20] *Ibid.* "A propositional content is given by the sentence that follows the 'that' clause in reporting a thought, a belief, or any propositional attitude." *Ibid.*

because thought contents can be specified both directly and indirectly.[21] For example, thought content is directly specified by the statement "(1) J. L. B. believes the proposition that he would naturally express by saying, '*I* will be the next person to receive a parking ticket in central Cambridge.'"[22] This content can either be indirectly specified in itself [i.e. "(2) J. L. B. believes that he will be the next person..."[23]] or as a report of the belief that would be indirectly specified [i.e. "(3) J. L. B. believes the proposition that he would naturally express by saying, '*J. L. B.* will be the next person..."[24]]. Bermúdez argues that (1) and (3) are not equivalent because J.L.B. could be suffering from amnesia that prevents him from remembering his own name. In this scenario, although J.L.B. may believe someone's claim that J.L.B. will be the next person to receive a parking ticket in central Cambridge, J.L.B. fails to realise that he is in fact J.L.B. It follows that (1) is an incorrect description of J.L.B.'s belief, (3) is a correct description and (2) is a correct indirect report of both. Bermúdez notes that this creates a problem because—according to the aforementioned criterion of having thought content that constitutively involves the first person pronoun—only (1) is a genuine 'I'-thought and the distinction between (1) and (3) seems to be unintelligible at the level of an indirect specification of content.[25] Bermúdez solves this problem by appealing to the work of Hector-Neri Castañeda:

> Castañeda distinguishes between two different roles that the pronoun 'he' can play in *oratio obliqua* [i.e. indirect content specification] clauses. On

---

[21] *Ibid.*

[22] *Ibid.* Emphasis added. "A direct specification of content involves specifying what I would say in *oratio recta*, if I were explicitly to express what I believe. In contrast, an indirect specification of content proceeds in *oratio obliqua*." *Ibid.*

[23] *Ibid.*

[24] *Ibid.* Emphasis added.

[25] *Ibid.*, pp. 3-4.

> the one hand, 'he' can be employed in a proposition that the antecedent of the pronoun (i.e. the person named just before the clause in *oratio obliqua*) would have expressed using the first-person pronoun. In such a situation, Castañeda holds that 'he' is functioning as a *quasi-indicator*. He suggests that when 'he' is functioning as a quasi-indicator, it be written as 'he*'. Others have described this as the *indirect reflexive* pronoun (Anscombe 1975). When 'he' is functioning as an ordinary indicator, it picks out an individual in such a way that the person named just before the clause in *oratio obliqua* need not realize the identity of himself with that person. Clearly, then, we can disambiguate between (2) employed as an indirect version of (1) and (2) employed as an indirect version of (3) by distinguishing between (2.1) and (2).
>
> (2.1) J. L. B. believes that he* will be the next person to receive a parking ticket in central Cambridge.
>
> Proposition (2.1) is an example of the indirect reflexive, while (2) is not. So, we can tie up the definition of an 'I'-thought as follows.
>
> Definition An *'I'-thought* is a thought whose content can only be specified directly by means of the first-person pronoun 'I' or indirectly by means of the indirect reflexive pronoun 'he*'.[26]

Bermúdez maintains that, not only is the capacity to think about oneself held in common with all the different kinds of self-conscious states (see examples above), but this capacity also underlies all of these states. That is, he claims that the cognitive states under consideration can only be called forms of self-consciousness because these states all have contents that can only be specified directly with the first person pronoun 'I' or indirectly with the indirect reflexive pronoun 'he*' or 'she*' and terms these contents first person contents.[27]

Bermúdez distinguishes between two different types of first person contents that correspond to two different modes in which the first person can be used, as first noted by

[26] *Ibid.*, p. 4. Boldface omitted.

Ludwig Wittgenstein. The first type of first person content that Wittgenstein describes invokes the use of 'I' as object (versus 'I' as subject) and can be analysed in terms of more basic propositions. For example, if the thought 'I am Φ' involves the use of 'I' as object, it can be broken down into a predication component (i.e. 'a is Φ') and an identification component (i.e. 'I am a'). Bermúdez notes that the reason for making this distinction is to account for the possibility of error that Wittgenstein calls attention to. That is, one can simultaneously be correct that an individual is Φ and incorrect that Φ is oneself.[28] This distinction can be couched in terms of Sydney Shoemaker's immunity to error. That is, "First-person contents are immune to error though misidentification relative to the first-person pronoun."[29] "The point, then, is that one cannot be mistaken about who is being thought about."[30] Bermúdez contends that this kind of immunity to error through misidentification is too restrictive because it fails to account for justifications for beliefs and the evidence that such justifications are based upon.[31] "So, to take one of Wittgenstein's examples, my thought that I have a toothache is immune to error through misidentification because it is based on my feeling of pain in my teeth. Similarly, the fact that I am consciously perceiving you makes my belief that I am seeing you immune to error through misidentification."[32] In light of this, Bermúdez goes on to

---

[27] *Ibid.*, pp. 4-5.

[28] *Ibid.*, pp. 5-6.

[29] *Ibid.*, p. 6. "To say that a statement 'a is Φ' is subject to error through misidentification relative to the term 'a' means that the following is possible: the speaker knows some particular thing to be Φ, but makes the mistake of asserting 'a is Φ' because, and only because, he mistakenly thinks that the thing he knows to be Φ is what 'a' refers to. (Shoemaker 1968, 7-8)" *Ibid.*

[30] *Ibid.*

[31] *Ibid.*

[32] *Ibid.*

formalise the idea that the contents of first person thoughts are immune to error though misidentification relative to the first-person pronoun. He is careful to specify, however, that the possibility that these contents *can be mistaken* does exist but they nevertheless have some kind of prima facie justification due to the evidence upon which they are based and the fact that this evidence is closely linked to the fact that the contents are immune to error through misidentification.[33] Bermúdez then provides a cogent argument[34] for why any first person content subject to error through misidentification is ultimately anchored in other first person content that is immune to error though misidentification. He notes that this conclusion entails that there is a class of self-ascriptions that are identification-free. Bermúdez considers and rejects two possibilities for how identification-free self-reference (e.g. being in pain) is possible. These rejected possibilities both involve explaining what immunity to error through misidentification consists of via the class of predicates that feature in judgements that are identification-free.[35] After arguing that these possibilities fail to explain what they claim to, Bermúdez offers the following more credible alternative.

---

[33] *Ibid.*, pp. 6-7.

[34] From the preceding discussion of justification and evidence "It seems, then, that the distinction between different types of first-person content originally illustrated by Wittgenstein can be characterized in two different ways. We can distinguish between those first-person contents that are immune to error through misidentification and those that are subject to such error. Alternatively, we can discriminate between first-person contents with an identification component and those without such a component. For the purposes of this book I shall take it that these different formulations each pick out the same classes of first-person contents, although in interestingly different ways.

It will be obvious that these two classes of first-person contents are asymmetrically related. All first-person contents subject to error through misidentification contain an identification component of the form 'I am *a*'. Now, consider the employment of the first-person pronoun in that identification component. Does it or does it not have an identification component? If it does, then a further identification component will be implicated, of which the same question can be asked. Clearly, then, on pain of an infinite regress, at some stage we will have to arrive at an employment of the first-person pronoun that does not presuppose an

Bermúdez maintains that one can explain what immunity to error through misidentification consists in through the first person element (which is naturally expressed with the first person pronoun) of identification-free judgements. He notes that it is commonly and correctly held that the rules and practices that pertain to the first person pronoun determine what the reference of that pronoun will be whenever it is used. This is done according to the simple rule that the first person pronoun always refers to the person using it whenever it is used correctly. Thus, the correct use of 'I' ensures that 'I' has a referent and that this referent is the person who is using the word 'I'. Stated differently, it is impossible for the first person pronoun to refer to someone other than the person using it if it is correctly and genuinely used. Bermúdez notes that ensured reference and immunity to error through misidentification are very closely related: the latter (relative to the first person pronoun) is a function of the meaning rule (discussed above) for the first person pronoun.[36] This is not to say that all judgements with ensured reference map onto all judgements that are immune to error through misidentification, as there can be instances of the latter (that are expressible with the first person pronoun) that are not immune to error relative to the first person pronoun:

> ...a baritone singing in a choir hears a tuneful voice that he mistakenly judges to be his own when it is in fact the voice of his neighbour (also a baritone). He then judges, 'I am singing in tune'. This is a clear instance of misidentification relative to the first-person pronoun, because the baritone makes the mistake of assuming that the baritone whom he justifiably believes to be singing in tune is the baritone to whom 'I' refers. Nonetheless, this is not a counterexample to the suggested view, because

---

identification component." *Ibid.*, p. 7.

[35] *Ibid.*, pp. 7-8.

[36] *Ibid.* p. 9.

> the evidence base upon which the baritone judges is not one of the categories of evidence bases that generates judgements immune to error relative to the first-person pronoun.[37]

Bermúdez argues that this is so because the evidence base requires that the ensuing judgement must be analysed in terms of both an identification component (i.e. 'That baritone is singing in tune') and a predication component (i.e. 'I am that baritone').[38]

Bermúdez combines the very close relation between ensured reference and immunity to error through misidentification with his previous conclusions. That is, he proposes a deflationary account of self-consciousness that is defined in terms of the capacity to think 'I' thoughts that are immune to error through misidentification. This immunity to error is a function of the semantics of the first person pronoun. Thus, the deflationary account of self-consciousness seeks to explain what it is that makes first person thought contents immune to error through misidentification with reference to the semantics of the first person pronoun.[39]

Bermúdez further elucidates the deflationary account of self-consciousness by drawing attention to its three distinct claims:

> Claim 1 Once we have an account of what it is to be capable of thinking 'I'-thoughts, we will have explained everything distinctive about self-consciousness.
> ...
> Claim 2 Once we have an account of what it is to be capable of thinking thoughts that are immune to error through misidentification, we will have explained everything distinctive about the capacity to think 'I'-thoughts.
> ...

---

[37] *Ibid.* pp. 9-10.
[38] *Ibid.* p. 10.
[39] *Ibid.*

> [Lastly,] Semantics alone cannot be expected to explain the capacity for thinking such thoughts [that are immune to error through misidentification]. The point must be that all there is to the capacity to think thoughts that are immune to error through misidentification is the capacity to think the sort of thoughts *whose natural linguistic expression* involves the first-person pronoun, where this capacity is given by mastery of the first-person pronoun.
> ...
> Claim 3a Once we have explained what it is to master the semantics of the first-person pronoun, we will have explained everything distinctive about the capacity to think thoughts that are immune to error through misidentification.[40]

In temporarily defending this account of self-consciousness, Bermúdez explains how mastery of the semantics of the first person pronoun accounts for the distinction between first person thoughts that have contents that are immune to error through misidentification and first person thoughts that have contents that are not immune in this way. He notes that the first person pronoun is used to express both types of first person thought content. Bermúdez begins his explanation of how these types can be distinguished by recalling that first person thought contents that are immune to error through misidentification (i.e. those that use 'I' as object) must be broken down into an identification component and a predication component. Mastery of the semantics of the first person pronoun is only being called upon to explain the former component—which is immune to error through misidentification.[41]

Bermúdez gives further justification to the deflationary account by stressing that it accords extremely well with a persuasive school of thought in philosophy of mind. He notes that many philosophers hold that the ability to ascribe predicates to oneself is what

[40] *Ibid.*, pp. 10-11. Boldface omitted. Italics added.

is distinctive about self-consciousness. Cartesians hold that these predicates are psychological while non-Cartesians hold that they are both psychological and physical. If the assumption is made that these predicates have a constant sense, then everything distinctive about self-consciously grasping those predicates that apply to oneself (or others) is dependant upon the act of self-ascription. More specifically, doing so is dependant upon the first person pronoun by which the act of self-ascription is effected. Hence, the deflationary view of self-consciousness is the next logical step.[42] Moreover, since this view accords a serious role to mastery of the semantics of the first person pronoun, it meshes extremely well with an important principle that has greatly influenced the development of analytical philosophy:

> This is the principle that the philosophical analysis of thought can only proceed through the philosophical analysis of language. The principle has been defended most vigorously by Michael Dummett.
>
> ... Many philosophers would want to dissent from the strong claim that the philosophical analysis of thought through the philosophical analysis of language is the fundamental task of philosophy. But there is a weaker principle that is very widely held:
>
> The Thought-Language Principle The only way to analyze the capacity to think a particular range of thoughts is by analyzing the capacity for the canonical linguistic expression of those thoughts.[43]

Bermúdez notes that the thought-language principle dictates that one must both find the canonical linguistic expression for a particular range of thoughts and explain the linguistic skills that must be mastered for the use of that linguistic expression in order to

---

[41] *Ibid.*, pp. 11-12.
[42] *Ibid.*, p. 12.
[43] *Ibid.*, pp. 12-13. Boldface omitted.

understand how thinking the given range of thoughts is possible. Further, when this methodology is combined with Bermúdez's above consistency-of-sense thesis that states that predicates have a constant sense regardless of whether or not they apply to oneself or others (i.e. whether they are being used for first, second or third person ascriptions), the deflationary account of self-consciousness follows. The thought-language principle dictates that self-conscious thoughts can only be understood by understanding the linguistic expression of those self-conscious thoughts. Bermúdez notes that paradigm cases of self-conscious thought such as knowledge of oneself involve ascribing certain properties to oneself. This entails that the distinctive features of the linguistic means by which one is able to appropriately apply certain predicates to oneself must be attended to.[44] "Since the constancy-of-sense thesis means that the distinctive features cannot lie in the relevant predicates, we soon arrive at the deflationary theory."[45]

C) The Paradox

Bermúdez recalls attention to the third claim of the deflationary theory as stated above which, when considered together with the first two claims, entails that everything distinctive about self-consciousness can be explained by explaining what it is to master the semantics of the first person pronoun. He begins this explanation by first recalling his observation that the rule that pertains to the use of the first person pronoun specifies that

[44] *Ibid.*, p. 13.
[45] *Ibid.*

the first person pronoun refers to the person using it.[46] Bermúdez terms this token-reflexive rule 1 which is formulated thus: "When a person employs a token of 'I', in so doing he refers to himself."[47] Again, this rule is essential to understanding the semantics of the first person pronoun and explaining the mastery of it is essential to understanding the capacity to think thoughts that are immune to error through misidentification. Token-reflexive rule 1, however, is inadequate because it fails to distinguish between the direct reflexive 'she' or 'he' and the indirect reflexive 'she*' or 'he*', as discussed in the previous sub-section. The former use of the first person pronoun is inadequate because it allows for the possibility that the individual referring to her or himself with the first person pronoun might be using it without realising that she or he is doing so—which is impossible with the first person pronoun.[48] Thus, the latter use of the first person pronoun must be utilised and this leads to token-reflexive rule 2: "When a person employs a token of 'I', in so doing he refers to himself*."[49]

> This creates obvious problems of circularity, however, because we can only understand how a person can refer to himself* by understanding how a person can refer to himself by employing the first-person pronoun. The indirect reflexive in indirect speech needs to be explained through the first-person of direct speech, which is, of course, 'I'. This circularity appears damaging to the deflationary account of self-consciousness. Recall that I characterized a first-person content as one that can be specified directly only by means of the first-person pronoun 'I' or indirectly only by means of the indirect reflexive 'he*'. If, as the deflationary account suggests, we take the capacity to think thoughts with first-person contents to be what we are trying to explain, then it is viciously circular to suggest

[46] *Ibid.*, p. 14.
[47] *Ibid.*
[48] *Ibid.*
[49] *Ibid.*

> that this capacity can be explained by mastery of a rule that contains a first-person content.[50]

Bermúdez seriously considers three additional versions of the token-reflexive rule in an attempt to avoid this circularity, but these ultimately fail as well.[51] Thus, the first tine in the paradox of self-consciousness is termed explanatory circularity because "Any theory that tries to elucidate the capacity to think first-person thoughts through linguistic mastery of the first-person pronoun will be circular, because the explanandum is part of the explanans, either directly, as in version 2, or indirectly, as in [subsequent versions omitted here]."[52] Bermúdez is careful to specify that the token-reflexive rule is not circular in itself. Rather,

> It becomes circular only if it is adjusted to rule out the possibility of accidental self-reference either by replacing the direct reflexive pronoun with the indirect reflexive pronoun or by requiring that the utterer of a token of 'I' should know that he produced the relevant token. But neither of these *modifications* is required *for* an account of the meaning of 'I'. To hold that there is any such requirement is to confuse the semantics of the first-person pronoun with the pragmatics of the first-person pronoun. The problem of explanatory circularity arises because it is not sufficient for the deflationary account, or any compatible account of self-consciousness, simply to provide an account of the semantics of the first-person pronoun. [Again, t]he deflationary account needs to provide an account of mastery of the first-person pronoun, and this will have to include *both the semantics and the pragmatics*. Hence, *the modifications have to be made*, with the ensuing circularity.[53]

The second tine of the paradox of self-consciousness is termed capacity circularity. Bermúdez argues that this arises because the capacity to think thoughts with first person contents is being explained in terms of the capacity to master the semantics of

---

[50] *Ibid.*, pp. 14-15.
[51] *Ibid.*, pp. 15-17.

the first person pronoun. This involves the relationship between these two capacities insofar as they respectively underlie the thing being explained versus the explanation.[54] That is, "the capacity for reflexive self-reference by means of the first-person pronoun presupposes the capacity to think thoughts with first-person contents, and hence cannot be deployed to explain that capacity. In other words, a degree of self-consciousness is required to master the use of the first-person pronoun."[55]

Bermúdez argues that capacity circularity cannot be dismissed by holding that it simply represents the limits of explanation and reflects the fact that certain cognitive abilities form what Christopher Peacocke terms a local holism. This is because capacity circularity extends beyond a supposed interdependence of explanation. That is, if the abilities that found self-consciousness form a local holism that can only be explained in terms of one another, then it would be impossible to explain how these abilities can be acquired in the normal course of cognitive development.[56] Bermúdez makes this more explicit by noting a constraint that is present in all discussions of cognition:

> The Acquisition Constraint If a given cognitive capacity is psychologically real, then there must be an explanation of how it is possible for an individual in the normal course of human development to acquire that cognitive capacity.[57]

By "explanation," Bermúdez is not referring to the actual (psychology or physiology based) account that is offered to explain a particular cognitive capacity. Rather, he means

---

[52] *Ibid.*, p. 16.
[53] *Ibid.*, p. 17. Emphasis added.
[54] *Ibid.*, p. 18.
[55] *Ibid.*
[56] *Ibid.*, pp. 18-19.
[57] *Ibid.*, p. 19. Boldface omitted.

the metaphysical basis of that account; the facts that are characterised by it if it is true. So, the acquisition constraint is merely a negative test for the psychological reality of any posited cognitive ability. In this sense, if it is impossible for an individual to acquire a posited cognitive ability, then it cannot be psychologically real.[58] Bermúdez proceeds to outline a paradigmatic way in which the acquisition constraint can be satisfied:

> Every individual has an innate set of cognitive capacities that it possesses at birth. Let me call that S_0. At any given time *t* after birth an individual will have a particular set of cognitive capacities. Let me call that S_t. Now consider a given cognitive capacity *c* that is putatively in S_t. Suppose that for any time $t - n$ the following two conditions are satisfied. First, it is conceivable how *c* could have emerged from capacities present in S_{t-n}. Second, it is conceivable how the capacities present in S_{t-n} could have emerged from the capacities present in S_0. By its being conceivable that one capacity could emerge from a given set of capacities, I mean that it is intelligible that (in the right environment) the individual in question could deploy the cognitive capacities it already has to acquire the new capacity. If those conditions are satisfied, then we have a paradigm case of learning.
>
> It is precisely here that capacity circularity bites, for the following reason. If mastery of the first-person pronoun is to meet the Acquisition Constraint, then clearly there must be some time *t* when S_t includes the capacity for linguistic mastery of the first-person pronoun, and a corresponding time $t - n$ when S_{t-n} does not include that capacity but includes other capacities on the basis of which it is intelligible that an individual could acquire the capacity for linguistic mastery of the first-person pronoun. The implication of capacity circularity, however, is that any such S_{t-n} will have to include the capacity to think thoughts with first-person contents, and hence that any such S_{t-n} will have to include the capacity for linguistic mastery of the first-person pronoun. Clearly, then, there can be no such S_{t-n} in terms of which the Acquisition Constraint could be satisfied.[59]

Bermúdez acknowledges that some philosophers maintain that the *psychological* acquisition constraint or anything like it has nothing to do with a *philosophical* account of

[58] *Ibid.*

concepts and conceptual abilities. Although he is critical of this view in another work, rather than taking pains to refute it in the present work, Bermúdez offers a cogent argument[60] for why this view has no implications for Bermúdez's particular use of the acquisition constraint and does not support its being rejected. Bermúdez then goes on to consider a possible innatist solution to capacity circularity (omitted here) that also fails to avoid the circularity.[61] He offers a further argued for recapitulation (also omitted) that, although capacity circularity is generated from what might initially appear to be narrow philosophical concerns, it creates the substantial problem of making it impossible to understand how mastery of the first person pronoun could ever be learned; it requires one to master the first person pronoun by first mastering the first person pronoun.[62]

Bermúdez summarises the paradox of self-consciousness by listing its six incompatible propositions:

> 1. The only way to analyze what is distinctive about self-consciousness is by analyzing the capacity to think 'I' thoughts.
> 2. The only way to analyze the capacity to think a particular range of thoughts is by analyzing the capacity for the canonical linguistic expression of those thoughts (the Thought-Language Principle).
> 3. 'I'-thoughts are canonically expressed by means of the first-person pronoun.

---

[59] *Ibid.*, pp. 19-20.

[60] "For my present purposes, though, all that I need to point out is that, however strictly one does adhere to the distinction [between philosophical and psychological questions regarding concept possession], it provides no support for the rejection of the Acquisition Constraint. The neo-Fregean distinction is directed against the view that facts about how concepts are acquired have a role to play in explaining and individuating concepts. But this view does not have to be disputed by a supporter of the Acquisition Constraint. All that the supporter of the Acquisition Constraint is committed to is the principle that no satisfactory account of what a concept is should make it impossible to provide an explanation of how that concept can be acquired. The Acquisition Constraint has nothing to say about the further question of whether the psychological explanation in question has a role to play in a constitutive explanation of the concept in question, and hence is not in conflict with the neo-Fregean distinction." *Ibid.* pp. 20-21.

[61] *Ibid.* pp. 22-24.

[62] *Ibid.* p. 21.

> 4. Mastery of the first-person pronoun requires the capacity to think 'I'-thoughts.
> 5. A noncircular account of self-consciousness is possible.
> 6. Mastery of the semantics of the first-person pronoun meets the Acquisition Constraint (in the paradigmatic way defined earlier).
>
> What I have argued in this chapter is that propositions (1) through (6) cannot be maintained together. More specifically, neither proposition (5) nor proposition (6) can be maintained in conjunction with propositions (1) through (4).[63]

Thus, Bermúdez concludes that at least one of these propositions must be abandoned in order to restore logical consistency and this will determine how a solution to the paradox will be classified. He notes, however, that there are both intuitive and philosophical reasons[64] for why each of the six propositions is highly appealing. If, however, the paradox cannot be resolved, the foregoing arguments entail that the existence of self-consciousness in its myriad forms is a logical impossibility.

D) The Solution

Bermúdez solves the paradox by rejecting proposition (2) above. In subsequent chapters of his book, he offers independent positive arguments for the rejection of the thought-language principle and shows how its rejection solves the paradox. Again, I refer

[63] *Ibid.* p. 24.

[64] "Proposition (1) seems to be entailed by the thesis that self-ascribable psychological and physical predicates have a constant sense across first- and third-person uses. There are powerful epistemological reasons for not wanting to reject that thesis. Proposition (2) has been described as, and widely accepted to be, a fundamental principle of analytical philosophy. Propositions (3) and (4) are indisputable. Both propositions (5) and (6) are highly desirable. If (5) is not true, then it follows that the capacity to think 'I'-thoughts is unanalyzable, which is a highly undesirable result, while the alternative to (6) is to deny that linguistic mastery of the first-person pronoun is psychologically real in anything like the way we understand it." *Ibid.*, pp. 24-25.

the reader to the highly plausible arguments in Bermúdez's text.[65] For present purposes, I will simply state some of Bermúdez's extremely well argued for conclusions without explicating the arguments in support of them.

The capacity to think genuinely first person thoughts does not depend on any linguistic or conceptual abilities. More specifically, there are non-conceptual first person thought contents that neither require language, the concept 'I', nor mastery of them. Non-conceptual instances of first person thoughts constitute several different kinds of self-consciousness. First, Bermúdez shows that the very nature of perceptual experience (i.e. perceiving the outer environment in even the most basic of ways) entails the existence of non-conceptual first person thought contents or 'I'-thoughts. *Bermúdez does not comment on or consider the theoretical possibility that non-perceptual experience (i.e. perceiving the inner mental environment in either basic or non-basic ways) entails the existence of non-conceptual 'I'-thoughts because empirical evidence for such experience is not forthcoming and probably never will be. Nevertheless, his account leaves sufficient room for this possibility being true or false with equal probability. Considered as it stands, Bermúdez's account does demand that beings who have any sort of perceptual awareness whatsoever necessarily possess self-consciousness.* Furthermore, higher forms of non-conceptual, non-linguistic self-consciousness are found in beings who have certain navigational abilities and, further still, who exist within certain non-linguistic social contexts. The latter forms of self-consciousness include psychological awareness of other

[65] *Ibid.*, pp. 27-297.

minds at the non-conceptual, non-linguistic level. For example, not only do social animals such as mice and pre-linguistic human infants[66] possess the most basic form of non-conceptual self-consciousness, they also possess a much richer variety of non-conceptual self-consciousness that comes about by contrasting oneself with a physical environment that can be navigated within and, further still, with other minds.[67] Again, if the foregoing is not the case, Bermúdez has shown that self-consciousness does not exist—which is patently absurd.

After speaking of the most basic forms of self-consciousness in a purely descriptive way, Bermúdez makes the following rare normative suggestion:

> One implication of this is that it widens the scope of what might be termed the first-person perspective far beyond the domain of humans, and even the higher mammals. This is particularly significant for any philosopher who shares the plausible view that self-consciousness, even in its primitive forms, carries with it a degree of moral significance.[68]

In subsequent chapters, I will argue that this degree is that embodied by moral personhood. An implication of chapter five in particular will be that only beings who possess core self-awareness can be *empathised* with for the purpose of knowing truths that are relevant to moral status. For the purposes of this thesis, all forms of non-conceptual self-consciousness are termed core self-awareness.

---

[66] At least from the point of birth, pre-linguistic infants possess core self-awareness. In the next section, it will be shown that this is also true of all sentient beings including sentient human fetuses. At a certain age, pre-linguistic infants acquire navigational and social non-conceptual self-awareness. Most, if not all, adult non-human animals have this richer form of non-conceptual self-awareness owing to their navigational and social abilities.

[67] Bermúdez, *Op. Cit.*, pp. 27-297.

[68] *Ibid.*, p. 162.

SECTION III: Sentience and Core Self-Awareness

Sentience, broadly defined, is the capacity to experience with the senses of touch, taste, smell, hearing or sight. Sentience, however, is much more commonly defined in its narrow sense of involving the capacity to experience physical pleasure and pain. In reality, there are probably no beings who possess core self-awareness who are not also sentient. In theory, however, it is possible that a being could exist who possesses core self-awareness but not sentience (e.g. a hypothetical future computer), as noted in section two, sub-section D above. The reverse, however, is not true. As Bermúdez has shown above and as will be further shown below, all beings who possess sentience (i.e. one form of perceiving the outer environment in a basic way) necessarily possess core self-awareness. The theoretical possibility that there could be a being who has core self-awareness but not sentience is relevant because it has implications for the morality of painlessly killing beings who have core self-awareness. This issue will be discussed in chapter six. For present purposes, it is important to have an understanding of sentience and its relationship with core self-awareness. Although this thesis applies to *all* beings who possess core self-awareness – including the human variety – it is relatively unconventional insofar that it encompasses non-human animals. Owing to this relative unconventionality, it is worth while to place some emphasis on the core self-awareness and sentience of non-human animals, as they comprise the largest class of beings to whom this thesis applies.

Regarding the *sentience* of non-human animals, Francione observes that the generally accepted fact that they are sentient is non-controversial owing to the

neurological and physiological similarities between human and non-human animals. Moreover, both mainstream scientific organisations – and research institutions that conduct pain experiments – explicitly accept this fact. Francione also notes that the evolutionary theory of Charles Darwin unequivocally recognises the existence of non-human animal minds.[69] Due to the state of scientific knowledge in the seventeenth century and to the risk that the Catholic Church might excommunicate or kill anyone who blasphemed that humans and other animals are similar in morally relevant ways, Francione suggests that the view held by René Descartes and the seventeenth century mechanists that non-human animals were nothing more than machines[70] may have been excusable. Due to the overwhelming evidence against it, however, virtually *no one* any longer holds the view that non-human animals are not sentient.[71]

Regarding the *self-awareness* of non-human animals, Francione cites Harvard biologist Donald Griffin's book *Animal Minds* which concludes, among many other things, that non-human animals are necessarily self-aware as a result of their perceptual awareness. At a bare minimum, Griffin concludes that since non-human animals are aware of their own bodies and actions, they must be aware of whose body exists and what actions that body is undertaking. Moreover, Francione cites neurologist Antonio

---

[69] Francione, *Introduction to Animal Rights*, *Op. cit.*, pp. xxxvi-xxxviii.

[70] "Descartes maintained that animals are nothing more than automatons, or robots, created by God. According to Descartes, animals do not possess souls, which are required for consciousness, and therefore lack minds altogether and cannot experience pain, pleasure, or any other sensation or emotion." (*Ibid.*, p. 105.) "Descartes and his followers performed experiments in which they nailed animals by their paws onto boards and cut them open to reveal their beating hearts. They burned, scalded, and mutilated animals in every conceivable manner. When the animals reacted as though they were suffering pain, Descartes dismissed the reaction as no different from the sound of a machine that was functioning improperly. A crying dog, Descartes maintained, is no different from a whining gear that needs oil." (*Ibid.*, p. 2.)

Damasio's work with humans who have brain damage or who have had strokes or seizures. Damasio maintains that these humans have core consciousness, a state that provides the individual with a sense of self in the here and now that is independent of memory, language and reasoning. Core consciousness is also possessed by non-human animals and can be distinguished from extended consciousness, a state that requires memory and reasoning but not language. Since extended consciousness involves autobiographical details, it is described as representational and, according to Damasio's conclusion, is possessed by various species of non-human animals.[72]

Damasio's core consciousness maps onto Bermúdez's primitive non-conceptual, non-linguistic self-consciousness. Damasio's *extended* consciousness maps onto Bermúdez's more developed forms of *non-conceptual, non-linguistic* self-consciousness. Both Damasio's core and extended consciousness map onto my definition of core self-awareness; the capacity to experience oneself as existing in the absence of any other qualification such as having conceptual abilities. The moral implications of their being different degrees of core self-awareness, self-awareness generally and supposedly different degrees of sentience, will be discussed in chapter four, section two and in chapter six, section three.

Regarding the *relationship* between sentience and self-awareness, Francione argues that pain cannot exist as some sort of ethereal experience. As a matter of logic, a conscious self must perceive pain as happening to her or him in order for the pain to be

---

[71] *Ibid.*, p. 105.
[72] *Ibid.*, pp. 114-115.

perceived at all. Further, Francione illustrates how non-human animal behaviour cannot not be explained without reference to core consciousness.[73] "Moreover, the ethological evidence suggests that other mammals, birds, and even fish possess memory and some reasoning ability, which would suggest that many species of animals have some form of extended consciousness and some autobiographical sense of self..."[74] The most relevant conclusions to draw from this discussion are that core self-awareness as defined is both a logically consistent and philosophically justified notion, all the empirical evidence suggests its existence, all sentient beings possess it and most, if not all, non-human animals possess it.

[73] *Ibid.*, pp. 138-140.
[74] *Ibid.*, p. 140.

CHAPTER 3: Kantian Ethics

SECTION I: Axioms

As mentioned in chapter one, section one, *the sole purpose* of my discussing Kant is to show that "Kantian" ethics, as commonly used in the disciplines and fields of practical ethics, is compatible with my thesis that everyone who possesses the capacity for core self-awareness ought not to be used merely as a means. Kant, of course, would deny that claim. In the following section, I will attempt to undermine Kant's basis for that denial. *If* I succeed, it still might be objected that *if* there is no ultimate justification for Kantian moral claims, then one would not be justified in using "Kantian-like" claims at all. Contrast this objection, however, with utilitarianism: both Jeremy Bentham[75] and John Stuart Mill[76] explicitly admit that there is no ultimate proof for their axiom that pleasure or happiness is good in itself. Would it then be claimed that no one is justified in using utilitarian claims at all? This question also holds for other theories such as feminist ethics: if it cannot be absolutely proven that oppression is wrong, should feminist ethics be abandoned? In this thesis, I assume that the answer to these questions is no. It is beyond the scope of the thesis to argue why the answer is no, but it is hoped that the reader also assumes that widely accepted moral theories should not be abandoned because their axioms are not ultimately subject to abstract arithmetical proof. This is not to say that a moral theory must necessarily contain foundational axiom(s) in order to be valid.

---

[75] Jeremy Bentham, "An Introduction to the Principles of Morals and Legislation," in *Ethics: Selections from Classical and Contemporary Writers*, ed. by Oliver A. Johnson (Fort Worth: Harcourt Brace, 1994), p. 211.

[76] John Stuart Mill , "Utilitarianism," in *Ethics: Selections from Classical and Contemporary Writers*, ed. by Oliver A. Johnson (Fort Worth: Harcourt Brace, 1994), pp. 261-262.

Non-foundational moral theories, such as pragmatism or holism, may or may not be valid. The present point is merely that both axiomatic moral theories whose axioms are not subject to proof and non-axiomatic moral theories may nevertheless be valid and justifiable.

The further objection might be made that, if I succeed in showing that Kant's axiom that the rational good will is good in itself is not ultimately subject to proof, then perhaps it should still be retained as an axiom in the same way that other foundational moral theories and their axioms are retained. I maintain that this should not be the case for Kantian ethics due to the special nature and purpose of Kant's axiom. That is, Kant's whole moral theory is centered around his concept of the categorical imperative: "Act only according to that maxim by which you can at the same time *will* that it should become a universal law."[77] According to Kant, it is through this imperative and its equivalent formulations alone that one can determine the moral rightness or wrongness of actions. Moreover, Kant claims to arrive at his categorical imperative through an examination of the nature of reason itself. Kant argues that the capacity to act according to rational laws such as the categorical imperative "...is will. Since reason is required for the derivation of actions from laws, will is nothing else than practical reason."[78] Hence, Kant's axiom that the rational good will is good in itself is an essential component of his claim that the categorical imperative (*which is the only basis for determining the moral rightness or wrongness of actions*) is solely derived from reason and, as such, is

[77] Immanuel Kant , "Foundations of the Metaphysics of Morals," in *Ethics: Selections from Classical and Contemporary Writers*, ed. by Oliver A. Johnson (Fort Worth: Harcourt Brace, 1994), p.

universally and necessarily true. Accordingly, if I succeed in undermining Kant's view that the rational good will is good in itself, I have also necessarily succeeded in undermining Kant's categorical imperative, which is Kant's sole basis for claiming that any given action is morally right or wrong. Conversely, Bentham and Mill's admission that their axiom that pleasure or happiness is good in itself is not subject to *rational proof* (contrary to Kant's claim for Kant's axiom) does not serve to undermine utilitarianism as a whole. Likewise, Aristotle holds that his view that *eudaemonia*[79] is the absolute good cannot be demonstrated. Nevertheless, this is no reason to reject Aristotelian ethics, at least on Aristotle's own terms.[80] Thus, although the assumption that certain fundamental moral axioms that are not ultimately subject to proof (such as Bentham, Mill and Aristotle's) should nevertheless be retained is acceptable (as I suggest in the preceding paragraph), undermining the basis for and then retaining Kant's axiom that the rational good will is good in itself is neither acceptable nor even possible on Kant's own terms.[81] In other words, if the following rebuttal of Kant's argument for the truth of his axiom succeeds, then neither the axiom nor *Kant's* arguments in support of his theory can be retained.

---

198. Emphasis added.

[78] *Ibid.*, p. 192.

[79] Literally "good spiritedness," which is often translated as "happiness," but Aristotle describes it as "to live well," "do well" or be "happy." Perhaps a better understanding of eudaemonia would be "flourishing."

[80] Aristotle acknowledges that the truth of his theory as a whole is not ultimately subject to conclusive proof. Moreover, he states that "The reader, on his part, should take each of my statements in the same spirit..." These are the general terms of Aristotle's moral theory. (Aristotle, "Nicomachean Ethics," in *Ethics: Selections from Classical and Contemporary Writers*, ed. by Oliver A. Johnson (Fort Worth: Harcourt Brace, 1994), pp. 57-58.)

[81] Again, unlike Aristotle, Kant's general terms for his moral theory are that moral truths are contained within the nature of reason itself and are universally and necessarily true.

In section three below, however, I will nevertheless argue that although Kant's rational justification for his moral theory is questionable, the "practical" versions of his categorical imperative (i.e. its 2nd, 3rd and 4th formulations) should not be abandoned. More specifically I will argue that although the first formulation of the categorical imperative loses all of its content when deprived of its rational basis, the remaining formulations do not. I will then suggest that the content of these remaining formulations is valuable for moral decision making despite their being robbed of their rational basis. *Importantly, this will allow me to claim in chapter six that a principle of equal consideration of interests that applies to all beings who have core self-awareness is justified, in part, by Kantian principles.* Note that, in applied ethics, it is claimed that the principle of non-maleficence is justified by every traditional moral theory including Kantian ethics and is often applied to human babies and others who do not live up to Kant's definition of a rational being. I will now attempt to show that the basis for Kant's categorical imperative, namely that the rational good will is good in itself and bridges the logical gap between *a priori* and *synthetic* claims of moral truth, is questionable.

SECTION II: The Good Will

A) Brief Summary and Statement of Purpose

In the first section of the *Foundations of the Metaphysics of Morals*, "Transition from the Common Rational Knowledge of Morals to the Philosophical," Kant argues that the unconditional moral worth of actions lies "...in the principle of the will irrespective of the ends which can be realized by such action. For the will stands, as it were, at the

crossroads halfway between its a priori principle which is formal and its a posteriori incentive which is material."[82] From this claim and others that are related to it, Kant argues that the moral worth of actions consists "...only in the conception of the law itself (which can be present only in a rational being) so far as this conception ... is the determining ground of the will."[83] Kant then suggests that this law is the categorical imperative and further argues for and develops this point in the second section of the *Foundations*. Since I am only concerned with questioning Kant's view that the good will is good in itself and accordingly unconditionally determines the moral worth of actions, I will concentrate on the first section.

B) Exposition

Kant begins with the assertion that the good will is the only thing that can be conceived of that is good without qualification. For Kant, the good will is not good because it produces good consequences. Rather, its good is solely due to its willing. He points to several qualities, abilities and "gifts of fortune" that are good in many respects but that can become bad and harmful if they are possessed by an individual who lacks a good will. Even happiness leads to pride and arrogance without a good will. Moreover, Kant claims that the sight of someone who enjoys uninterrupted prosperity but who lacks a good will can never give pleasure to a rational impartial observer. Hence, Kant

[82] Kant, *Op. cit.*, p. 189.
[83] *Ibid.*, pp. 189-190.

concludes that the good will is a necessary condition for such a happy or prosperous individual to be worthy of his or her happiness.[84]

Kant raises the possible objections that his claim that the good will is good in itself has no basis and that his assumption that nature appointed reason as the ruler of the will is misguided. He responds by asserting the axiom that any being that is suitably adapted to life will only have organs that are the fittest and best adapted to their respective purposes. If, however, nature endowed a being who has reason and will with those qualities for the purpose of the being attaining happiness (i.e. contentment with one's condition, preservation and welfare), then Kant asserts that reason and will would be ineffective at being the executors of that purpose. Kant maintains that instinct would be much more effective for the purpose of attaining happiness than reason or the will could ever be. Moreover, he claims that if, in addition to instinct, a being who is structured by nature to attain happiness also possessed reason, the sole purpose of that being possessing reason would be for that being to contemplate the happy constitution of her or his nature. That is, nature would not have constituted this being such that reason, with its "weak and elusive guidance" and "with its weak insight," could enable her or him to think out the plan of happiness and the means of attaining it. Kant supports this view with the further contention that as reason devotes itself to happiness, true contentment is reduced. For example, Kant holds that those who are most experienced in the use of reason to acquire various advantages (e.g. common luxury and the sciences which seem

[84] *Ibid.*, pp. 184-185.

to them to be a luxury of the understanding) eventually discover that reason has brought them trouble and not happiness.[85] That is, those who use reason to attain happiness do not attain their desired end and they eventually envy "the common run of men who are better guided by mere natural instinct and who do not permit their reason much influence on their conduct."[86] Moreover, Kant claims that those who temper or refute the view that nature has constituted them for the purpose of attaining happiness and provided them with reason to achieve this purpose do not necessarily possess a morose attitude or ingratitude to "the goodness with which the world is governed." Rather, these individuals maintain that nature has not constituted them for the purpose of attaining happiness. They maintain that the purpose of their existence is the necessary condition for all other subjective purposes; the good will. Again, Kant asserts that reason is not competent to govern the will towards the purpose of attaining its objects and the satisfaction of needs. Rather, nature would have achieved this purpose far better through instinct.[87]

Kant further asserts that nature endows humans with reason and this reason is a practical faculty because it is structured by nature to have an influence on the will. Moreover, he holds that the natural purpose of reason is to produce a will that is good in itself in the same way that the natural purposes of other capacities (or organs) are suited to the functions that they perform. This is because reason is absolutely essential to there being a will that is good in itself. Kant asserts that this good will is unconditional and is the highest good that is the condition for all other goods, including the desire for

[85] *Ibid.*, pp. 185-186.
[86] *Ibid.*, p. 186.

happiness. Moreover, the view that the cultivation of reason restricts or eliminates happiness is compatible with the wisdom of nature. Kant says that this is so because nature structures reason such that its purpose is to produce a good will, not to produce happiness. The purpose of reason to create a good will is only capable of a contentment of its own kind, namely a contentment that results from the production of a good will.[88]

At this point, Kant proceeds to develop or – more accurately stated – *describe* the concept of a will that is good in itself. He does this by considering the concept of duty, which itself contains the concept of a good will along with certain subjective restrictions and hindrances that nevertheless do not conceal the good will. When defining the concept of duty, Kant does not include actions that are opposed to duty because it is impossible that these actions could be done *from* duty due to the fact that they conflict with duty. He also excludes actions that are in accordance with duty but that are not performed by an individual who has a direct inclination to do these actions. That is, such an individual would not be impelled to do actions that are in accordance with duty *by* duty itself, but would rather be impelled to the actions by some other inclination. Regarding actions that are in accord with duty, Kant notes that it is easy to distinguish between actions that are done from duty and those that are done for some selfish purpose. When actions that are in accord with duty are done by someone who has a direct inclination to do them, however,

---

[87] *Ibid.*
[88] *Ibid.*, pp. 186-187.

it is difficult to distinguish between actions that are done from duty and those that are done for some selfish purpose.[89]

For actions that are in accord with duty that are not done from duty, Kant first gives one example of an action that is done by someone without a direct inclination to do it and one example of an action that is done by someone who does have a direct inclination to do it. The former example is of a merchant who, in accordance with duty, does not overcharge inexperienced customers. Nevertheless, the merchant's action is done neither from duty nor from direct inclination but rather is done solely for the merchant's own self-interest. The latter example is of most individuals who, in accordance with duty, preserve their own lives. Moreover, everyone has a direct inclination to preserve her or his own life. Nevertheless, the action of most individuals preserving their lives is not done from duty. This truth is exposed when these individuals kill themselves in response to unavoidable painful adversities and sorrow. If, however, an individual in such circumstances silently wishes for death but nevertheless preserves his or her life without loving it, only then is the action done from duty. Moreover, only then does the maxim for this individual's action have moral import.[90]

Kant offers a few more examples of actions that are done from duty and those that are not. He asserts that the actions of most individuals who are inclined to help others out of kindness in the absence of selfish motives and who get inner satisfaction from doing these actions are not done from duty and have no true moral worth. Kant claims that these

[89] *Ibid.*, p. 187.
[90] *Ibid.*

actions are similar to actions that are done as a result of other inclinations, such as the inclination to honour. The maxim of this action does not come from duty. Kant then distinguishes actions that are done from inclinations such as kindness and honour from actions that are done from duty alone. He gives two examples of individuals who help others but who nevertheless have no sympathy for others nor inclination to help them. The helpful actions of these individuals are done from duty and do have true moral worth.[91]

Kant maintains that there is at least an indirect duty to secure one's own happiness. He is careful to stipulate, however, that the *inclination* towards securing one's happiness does not play any role in the *duty* to secure that happiness. He argues that happiness is at least an indirect duty because a state of unhappiness could tempt one to transgress her or his duties. Kant contrasts this happiness-as-indirect-duty with actions that secure happiness that comes from inclinations rather than duty. He asserts that all inclinations are summed up in the idea of happiness. It is impossible, however, for individuals who attempt to secure happiness from sources other than duty to form a definite and certain concept of the sum total of all inclinations (i.e. happiness). Indeed, when these individuals attempt to form an idea of happiness, they find that some of their inclinations are nevertheless thwarted. Kant notes that this is the reason why a single inclination can, for some, take precedence over the more general (fluctuating) idea of happiness or the sum total of all inclinations. This happens, for example, when someone

[91] *Ibid.*, pp. 187-188.

with gout calculates that drinking alcohol in the present moment in order to gain enjoyment in that moment takes precedence over the more general, perhaps groundless, expectation of happiness that is supposed to reside in good health. If, however, this individual with gout did not refrain from drinking alcohol because the inclination towards happiness (and the idea that health supposedly brings happiness) determined his will, but rather abstained out of duty alone, Kant maintains that his action would have true moral worth. Likewise, the command to love both one's neighbour and enemy does not come from inclination because love as an inclination cannot be commanded. Similarly, beneficence from duty is practical love, resides in the will and principles of action and can be commanded—even if it is opposed by a natural and unconquerable aversion. Beneficence impelled by inclination, however, is pathological love, resides in the propensities of feeling and tender sympathy and cannot be commanded.[92]

From the forgoing description of the concept of duty, Kant concludes that the first proposition of morality is that actions must be done from duty in order to have moral worth. The second proposition, also derived from the forgoing description and examples, is that an action done from duty gets its moral worth in the maxim by which the action is determined. An action done from duty does not get its moral worth from the purpose that is to be achieved through it. Thus, the moral value of an action done from duty solely depends upon the principle of volition by which the action is done. It does not depend upon the reality of the object of the action and this action is done without any regard to

[92] *Ibid.*, pp. 188-189.

the objects of the faculty of desire. Again, from the foregoing description of the concept of duty, Kant concludes that actions do not get unconditional and moral worth from the purposes of those actions or from the effects of those actions-as-ends-and-incentives-of-the-will. That is, this worth does not reside in the will in relation to its hoped for effect. Rather, actions get their unconditional and moral worth from the principle of will that is unrelated to the ends that can be realized by these actions. Again, this is because of Kant's view that the will stands halfway between its formal *a priori* principle and its material *a posteriori* incentive. Given that the unconditional and moral worth of actions must be determined by something, Kant argues that if these actions are done from duty, their unconditional moral worth must be determined by the formal principle of volition in itself. This is the case because every material principle has been withdrawn from the action done from duty.[93]

From the preceding two propositions of morality, Kant infers the third proposition that "Duty is the necessity of an action done from respect for the law."[94] He argues that it is impossible to have respect for an object-as-an-effect-of-a-given-action because it is merely an effect and not an activity of the will. It is possible, however, to have an inclination towards such an object. If one has an inclination then one can, at most, approve of it. If someone else has an inclination then one can, at most, "love" it, or regard it as something that promotes one's self-interest. One cannot, however, respect any inclination. A thing cannot be respected if it is connected to the will only insofar as it

[93] *Ibid.*, p. 189.
[94] *Ibid.*

serves as the will's consequence, as in the case of inclinations. Kant maintains that the only thing that can be respected is the thing that is connected to the will only insofar as it serves as the will's "ground." This thing, called the law, does not serve one's inclinations. Rather, the law overpowers or ignores one's inclinations. Since the law itself is the only thing that can be an object of respect, Kant concludes that it is a command. Moreover, the law is the only thing that can objectively determine the will because actions that are done from duty completely overpower or ignore inclinations (and therewith every object of the will). Once inclinations and the objects of the will are excluded, nothing remains to objectively determine the will except the law. Similarly, nothing remains to subjectively determine the will except the pure respect for the practical law.[95] (Note that Kant specifies that "This subjective element is the maxim[96] that I should follow such a law even if it thwarts all my inclinations."[97]) Hence, Kant concludes that the expected effects of actions (or principles that borrow their motives from this expected effect, such as the principle of utility) do not give those actions their moral worth. He says this is also true because the highest and unconditional good can only be found in the will of a rational being and the expected effects of actions (which may produce happiness) could be caused as a result of other factors than the will of a rational being. Therefore, Kant concludes that the unconditional moral worth of actions consists in the conception of the law itself (which is the determining ground of the will) and nothing else. Again, this conception of

---

[95] *Ibid.*

[96] Kant defines a maxim as "the subjective principle of volition. The objective principle (i.e. that which would serve all rational beings also subjectively as a practical principle if reason had full power over the faculty of desire) is the practical law." *Ibid.*

[97] *Ibid.*

the law can only be present in rational beings. The rational being who acts according to this conception of the law has unconditional moral worth regardless of the result of her or his action.[98]

Kant notes that he has defined the will such that it does not include "impulses which could come to it from the obedience to [just] *any* [conception of the] law..."[99] Since nothing else remains to describe the will, Kant states that the conception of the law which must determine the will (and act as the will's principle without reference to the expected result of the action that is willed) is universal conformity of the action that is willed to the law in itself. In other words, the law states that one should never act unless one could (consistently) will that the maxim for one's action be a universal law. The principle of the will is mere conformity to this law in the absence of assumptions about any other particular law that applies to certain actions. Kant notes that this is necessarily true unless the concept of duty is a delusion. He states that mere conformity to the law must serve as the principle of the will and that the reason that humans use in their judgments constantly has this principle of the will in view.[100] He offers a brief example of an action that cannot be consistently willed as a universal law and then states "that the necessity of my actions from pure respect for the practical law constitutes duty. To duty every other motive must give place, because duty is the condition of a will good in itself, whose worth transcends everything."[101] In his closing remarks to the first section of the *Foundations*, Kant observes that, for the average rational being, the rational good will and

[98] *Ibid.*, pp. 189-190.
[99] *Ibid.*, p. 190. Emphasis added.

the duties that issue from its universal law come into conflict with the propensity of those beings to seek the sum total of their inclinations; happiness. If the latter prevails in a given individual, Kant notes that the actions of that individual cannot be morally good according to common practical reason.[102] In the second section, Kant further explains the principle of universal law called the categorical imperative, offers some examples of its implementation and specifies equivalent ways in which it can be stated.[103]

C) Critique

Kant's suggestion that the good will is good in itself does not follow from his claim that various qualities, abilities, natural gifts and happiness can be bad and harmful depending on the circumstances, or whether or not there is a good will present. Kant responds to this objection by formulating a second objection whose answer leads him to a reply to the first objection stated above. The second objection is, contrary to Kant's assumption, nature did not appoint reason as the ruler of the will. Kant's reply that if nature appointed the pursuit of happiness as the ruler of the will, this would be an ineffective arrangement (contrary to the effective arrangements of organs and their purposes that are found in nature), does not entail the conclusion that nature appointed reason as the ruler of the will. While it may or may not be true that reason has some sort of influence on the will, Kant's view regarding the particular form that this alleged influence takes is questionable. Note that this is where Kant's reply to the first objection

[100] *Ibid.*, pp. 190-191.
[101] *Ibid.*, p. 191

stated above becomes relevant. Kant's claim that nature appointed reason to the purpose of producing a will that is good in itself does not follow from his claim that a condition for their being a will that is good in itself is the presence of reason. In other words, even if Kant's unargued for assumptions that a) reason has an influence on the will and b) reason is absolutely essential to their being a will that is good in itself are true, it does not follow from this that nature appointed reason to the purpose of producing a will that is good in itself. One alternative possibility that is widely accepted by those who study nature[104] is that nature appointed reason to the purpose of efficiently and extensively replicating the genes of rational beings and ensuring that these genes, in turn, are replicated in future generations.[105] Countless other possibilities exist. Kant, however, does not defend his particular version of natural history. Even if his questionable observation that those who are most experienced in the use of reason to acquire "happiness" (as defined by Kant) do not acquire it, it does not follow from this that the purpose of reason is to produce a will that is good in itself. Again, his assertion that the good will is the highest good that is the condition for all other goods (e.g. positive qualities, abilities, natural gifts and happiness) does not entail that this will is good *in itself*. Kant merely assumes that the good will is good in itself without substantive argument.

It might be objected that when Kant says that reason is "absolutely essential" to there being a will that is good in itself and that the good will is the condition for all other goods, he is not simply referring to necessary conditions. Rather, Kant is asserting that a)

---

[102] *Ibid.*, pp. 191-192.
[103] *Ibid.*, pp. 192-207.

reason is both a necessary and sufficient condition for a will that is good in itself and b) the good will is both a necessary and sufficient condition for all other goods. If these claims were true, it would follow that a) reason can exist *if and only if* there is a will that is good in itself[106] and b) qualities, abilities, natural gifts, happiness and so on are good *if and only if* there is a good will. This, in turn, suggests that the good will is good in itself. In the portion of the *Foundations* critiqued so far, however, Kant does not offer any argument for why reason is necessary and sufficient for there being a will that is good in itself or why a good will is necessary and sufficient for their being any other goods. This argument is only offered when Kant describes the concept of a will that is good in itself further on in the first section of the *Foundations*.

A rough outline of Kant's argument that reason is necessary and sufficient for there being a will that is good in itself and that a good will is necessary and sufficient for there being any other goods is as follows:

P1) Duty contains the concept of a will that is good in itself.

P2) Only actions that are done exclusively from duty are properly called dutiful.

P3) The absolute or unconditional moral value of an action done from duty (read "done from a will that is good in itself") solely depends on the principle of volition by which the action is done.

P4) This principle is the formal principle of volition itself (i.e. the law) which states that one should never act unless one can *consistently* will that the maxim for one's

---

[104] i.e. evolutionary biologists.

[105] Richard Dawkins, *The Selfish Gene*, New ed. (Oxford: Oxford University Press, 1999).

action be *universal*. In other words, duty, *or a will that is good in itself that is contained within duty*, produces actions that are correspondingly absolutely and unconditionally good in themselves due to the principles of consistency and universality alone; i.e. *reason*.

C1) Therefore, reason is necessary and sufficient for there being a will that is good in itself and a good will is necessary and sufficient for the possibility of their being any other goods (e.g. various actions that correspond to positive qualities, abilities, natural gifts and happiness).

C2) The good will is good in itself.

I will now argue that Kant does not sufficiently substantiate the truth of the above premises. Hence, it will be shown that his conclusion that the good will is good in itself is not justified.

Kant defines the concept of duty without arguing for the truth of his definition. For example, when he excludes from the concept of duty all actions that are both in accordance with duty *and* that are done by those who have direct inclinations to do these actions, Kant does not argue why this is or should be the nature of duty. Kant's example of most individuals having a direct inclination to preserve their own lives but nevertheless not doing this action from duty because they would not, in the event of a painful terminal illness, silently wish for death and then refrain from killing themselves merely restates Kant's claim that actions that only accord with and are inclined towards duty are not

[106] Note that, for Kant, immoral actions (i.e. those that are not influenced by a good will) are necessarily irrational.

dutiful. Rephrasing a claim in the form of an example is not the same thing as drawing attention to an example of an occurrence and then comparing that occurrence with a general principle or definition that the occurrence fits. Kant does the former. When he asserts in his example that when a painfully and terminally ill individual preserves her or his life without loving it (or, presumably, even when strongly hating it), only then is this action done from duty and only then does its maxim have any moral import whatsoever, Kant is merely making an assertion without arguing for its truth. This is also the case for Kant's examples of kind actions, done "without any motive of vanity or selfishness,"[107] which are not dutiful unless those who perform them do not do so as a result of the least bit of sympathy or inclination to help others. Likewise, for his example of the healthy action of an individual with gout abstaining from alcohol, Kant asserts that this action is dutiful and has moral worth if and only if it is not undertaken to seek health or the happiness that health might bring. Kant does not, however, offer any reasons for holding this view. Lastly, with respect to love, Kant claims that loving one's neighbour (presumably "without any motive of vanity or selfishness") out of feeling and tender sympathy is pathological, lacks moral worth and is not done from duty. This, in addition to the claims found in all of the other examples Kant uses, is not substantiated with any reasons.

Recall that Kant's purpose in defining duty is really to define a good will that is good in itself—which is supposedly contained within the concept of duty. Perhaps the

[107] Kant, *Op. cit.*, p. 187.

question of whether or not a good will is good in itself can be better understood by referring to a couple of more examples. Imagine an individual who only acts from Kant's conception of duty. Out of duty and duty alone, she protests an unethical war and is unjustly taken as a political prisoner. She remains imprisoned for 30 years and everyone outside of the prison has forgotten about her. Her health is failing and it can reasonably be said that she will die in prison. During the time that is left in her life, does this prisoner have a good will *that is good in itself*? Many individuals might be tempted to answer "yes." Is this answer, however, due to the faint hope or possibility that the prisoner might be, against all odds, rescued? Consider a second example. As punishment for her supposed crime, the same political prisoner is anesthetised, shackled, caged and placed in a rocket that is destined for the sun. The nations of the Earth cannot spare the massive resources that would be required to attempt a rescue. Ten minutes before she burns up, it can reasonably be said with absolute certainty that, for the rest of her short life, the prisoner is completely powerless to undertake any action done from duty. Kant states:

> Even if it should happen that ... this [good] will should be wholly lacking in power to accomplish its purpose, and even if the greatest effort should not avail it to achieve anything of its end, and if there remained only the good will (not as a mere wish but as the summoning of all the means in our power), it would sparkle like a jewel in its own right, as something that had its full worth in itself.[108]

This is a beautiful metaphor. *If* "God" existed, this would no doubt describe God's will.[109] Kant certainly seems to have absolute faith in its truth. Kant, however – despite

[108] *Ibid.*, p. 185.

[109] Notwithstanding that "God" is usually defined as having omnipotent will. The present point is merely that, if it existed, the will of God would have absolute worth or good in itself.

his personal deep seated religious convictions – claims to be positing a wholly secular moral theory that is founded in *a priori* reason and *a priori* reason alone. In the preface to the *Foundations*, he stipulates that in order to discover the supreme principle of morality, one must examine reason as it functions in the guidance of one's conduct.[110] "The essence of reason is consistency and the test of consistency is universal validity."[111] Kant's claims of devotion to secular objectivity and reason alone, however, are belied by his pronouncements that are devoid of rational argument. As has been shown above, he does not ground his understanding of duty and good will in the principles of *a priori* reason; consistency and universality. The foregoing examples of the political prisoner are not intended to suggest that the prisoner's will – considered in itself – is sad or worthless. Rather, the second example in particular suggests that, although the prisoner's will is good, it is not necessarily good *in itself* and this is so largely because Kant fails to justify this claim.

From the foregoing critical analysis of Kant's conception of duty and good will, it follows that Kant's first two propositions of morality are unfounded. In particular, his views that actions must be done from duty – as Kant defines it – in order to have moral worth and that actions done from duty cannot get their moral worth from their intended purposes are not substantiated by rational argument. Since the principle of volition by which actions are done (i.e. a good will that is good in itself) does not appear to be a rationally defensible one, Kant's view that actions get their moral worth from that

[110] *Ibid.*, p. 183.
[111] *Ibid.*

principle is questionable. Accordingly, Kant's third principle of morality that follows from the first two is likewise questionable. That is, his view that actions that are necessarily performed out of respect for the law (i.e. the formal principle of volition itself) constitute duty is suspect given that the principle of volition (i.e. a will that is good in itself) is itself suspect. As such, Kant's claim that this law objectively and subjectively determines the will is moot. Moreover, the conception of the law as universal conformity of the action that is willed to the law is rendered irrelevant if the law cannot be said with certainty to determine the will. Thus, the conclusion that duty (or a will that is good in itself) produces actions that are good in themselves due to their universally conforming to the law of reason is unsound. This suggests that the view that reason is necessary and sufficient for there being a will that is good in itself, or that a good will is necessary and sufficient for the possibility of their being any other goods, is also unsound. In short, Kant has not shown that the good will is good in itself.

I agree with Kant that "Mere conformity to the law as such … must serve as [the principle of the will] if duty is not to be a vain delusion and chimerical concept." Given the preceding, however, I disagree with his claim that "The common reason of mankind … is in perfect agreement with this"[112] chimera[113] "and has this principle constantly in view."[114] Kant's view that the good will is good in itself and accordingly unconditionally

[112] *Ibid.*, p. 190.

[113] By the word "this," Kant is referring to the previous sentence which states "Mere conformity to the law as such serves as the principle of the will and it [i.e. mere conformity to the law as such] must serve as such a principle [i.e. of the will] if duty is not to be a vain delusion and chimerical concept." Thus, the "this" Kant is referring to is mere conformity to the law as such, which is alternatively described as the principle of the will. I argue that this principle is indeed a chimera.

[114] Kant, *Op. cit.* Emphasis added.

determines the moral worth of actions does not conform to the principles of reason that Kant employs; consistency and universality. Moreover, from my discussion and questioning of Kant's examples, it is not clear if the common reason of humanity has a concept of a will that "sparkles like a jewel" in its own right constantly in view. Since this view of the good will is an essential component of Kant's claim that his categorical imperative is solely derived from reason and, as such, is universally and necessarily true, the basis of Kant's entire moral theory appears to be questionable.

D) A Classic Objection Explained

The foregoing critique sheds light on a common objection that is often lodged against Kantian ethics. It is objected that Kantian ethics requires one to blindly follow the dictates of the categorical imperative without regard to consequences, no matter how severe those consequences are. For example, it is 1944 Germany and a Nazi soldier knocks on the door and asks if anyone is hiding in the attic. According to the first formulation of the categorical imperative, it would be immoral to lie by saying "no." This is because, if universalized, the maxim that "When I believe the lives of my guests to be in danger from an answer of 'yes' to the question of an executioner, I will answer 'no,' although I know this not to be the case." would necessarily contradict itself. To paraphrase Kant, the universality of a law which says that anyone who believes others to be in mortal danger from a certain answer to a posed question could say what she or he pleased with the intention of not answering correctly would make the answer itself and the life-saving end to be accomplished by it impossible; no questioner would believe

what was answered but would only laugh at any such assertion as vain pretence.[115] Likewise, an answer of "no" would be to use the Nazi soldier merely as a means to saving the lives of others and to withhold information that would prevent the Nazi from making an autonomous decision; thus violating the second and third formulations of Kant's categorical imperative respectively.[116] The fourth formulation would be violated because one would be issuing a self-contradictory law to someone who is a subject within the realm of ends, thus using that individual merely as a means and imposing a price on that individual rather than respecting her or his dignity.[117]

The savvy Kantian will reply that answering "no" to the Nazi who knocks on one's door is not at all required. One could simply refuse to answer the question. It could be argued that the mere withholding of information to an agent whose motive or intention is to use another merely as a means does not violate any positive duty to help that agent, or respect her or his autonomy, because this would amount to a duty to respect someone as an end in her or himself while simultaneously using someone merely as a means, which is self-contradictory. Moreover, it might be argued that an answer of "yes" would be prohibited because it would entail using one's guests in the attic merely as means to helping the Nazi. Hence, the savvy Kantian might argue that the correct response is to refuse to answer or tactfully avoid the Nazi's question. Any question avoidance, however, could not involve any intentional deception whatsoever if the categorical imperative is not to be violated.

---

[115] *Ibid.*, p. 199.
[116] *Ibid.*, pp. 202-204.

This reply is highly inadequate. In all reasonable likelihood, doing anything in response to a question asked by the armed representative of an oppressive totalitarian state other than directly answering that question would result in severe *consequences*; the Nazi would arrest the individual who answered the door, search the house and send the rest of its occupants to concentration camps. Given this reality, the maxim for one's action is accurately stated thus: "When I believe the lives of my guests to be in danger from any other response than 'no' to the question of an executioner, I will give this response although I know it not to be the case." As shown above, this personal maxim violates all four formulations of Kant's categorical imperative. Moreover, Kant's theory in general and his realm-of-ends formulation of the categorical imperative in particular do not admit to conflicts between those who are regarded as ends in themselves. For instance, the "yes" answer that this case requires *under Kantian theory* does not use one's guests in the attic merely as means to helping the Nazi. This is because in answering "yes," one is solely acting in accordance with the positive duty to help the questioner; this is one's only intention or motivation and it consists of willing the maxim for that action alone. *If this was not so*, then one would be undertaking act A (answering the Nazi's question) while considering act B (saving the lives of one's guests). *Relative to act A* – i.e. the action that one is formulating one's maxim for – act B is solely a consequential consideration. As such, it is wholly irrelevant to Kantian theory. Thus, in this way, Kantian ethics does not allow for any conflicts between different individuals who are

[117] *Ibid.*, pp. 205-206.

regarded as ends in themselves. One's guests must simply bite the bullet for the sake of one's adherence to Kant's categorical imperative.

Given the above analysis of the good will, we are now in a position to understand the often raised objection to Kantian theory that it requires one to coldly follow an abstract directive regardless of how severe the consequences are. If the good will of every being who has one (or the capacity for one) is absolutely good in itself irrespective of anything else, then no one of these billions of "jewels" can be harmed *for any reason whatsoever*; reason itself precludes this. Moreover, since all of these jewels are hard and exist in isolation, the perfect beauty of one cannot detract from that of any other and there can be no conflicts between them; they all coexist in perfect harmony within the realm of ends; reason itself requires this. The beauty and grandeur of Kant's moral theory inspires one to metaphor. Unfortunately for Kant, however, its foundation – the good will that is good in itself – has been shown to be unstable. This should come as welcome news, both to harboured refugees hiding within oppressive states and to feminist ethicists[118] who advocate a more realistic conception of human nature and autonomy.

E) Implications for "Non-Rational" Beings

Recall that, according to Kant, the moral worth of actions solely consists in the conception of the law itself, which is in turn the determining ground of the will. Moreover, this conception of the law can only be present in rational beings and rational

[118] Sally Sedgwick, "Can Kant's Ethics Survive the Feminist Critique?" in *Feminist Interpretations of Immanuel Kant*, ed. by Robin M. Schott (University Park, Pennsylvania: Pennsylvania State University

beings get their moral worth from acting (or the possibility of acting) according to this conception of the law. This conception of the law, the categorical imperative, has four different formulations. One of these is that "every rational being exists as an end in himself and not merely as a means to be arbitrarily used by this or that will."[119] Thus,

> Beings whose existence does not depend on our will but on nature, if they are not rational beings [who are defined as being capable of conceiving the categorical imperative], have only a relative worth as means and are therefore called "things"; on the other hand, rational beings are designated "persons," because their nature indicates that they are ends in themselves, i.e., things which may not be used merely as means.[120]

Since the assumptions about the good will that underlie this view of rational nature have been undermined, the possibility remains open that beings such as human babies, humans who are severely mentally challenged, humans who suffer from severe dementia such as that found in end-stage Alzheimer's and AIDS patients, non-human animals and hypothetical future computers who could have core self-awareness (all of whom may be incapable of asking themselves if the maxim for their actions can be consistently willed as universal law) are persons who, according to Kant's definition of that term, are ends in themselves and should not be used merely as means. Contrary to Kant's theory, these human and non-human beings may have inherent dignity and not merely a relative price. This possibility will be argued for in the following pages.

Press, 1997); Susan Sherwin, *No Longer Patient: Feminist Ethics and Health Care* (Philadelphia, Temple University Press, 1992), pp. 137-140, 142, 156-157.

[119] Kant, *Op. cit.*, p. 202.

[120] *Ibid.*, pp. 202-203.

SECTION III: Salvaging "Kantian Ethics"

A) Applied Ethics

Kantian moral theory places its emphasis on duty, intention or motivation and the interests of the individual. Utilitarianism emphasizes consequences and the interests of every individual considered collectively. The objection against Kantian ethics in section two, sub-section D, above, has its counterpart in utilitarianism: it is often argued that utilitarianism tramples over the lives and interests of individuals in service of maximizing good consequences. In response to the seemingly irresolvable tensions between these two theories, Beauchamp and Childress developed the theory of principlism, the ethical theory that utilizes a set of principles to determine the ethical course of action in any given situation. The major principles used are those of autonomy, beneficence, non-maleficence, and justice; all of which are ethically binding

> ...but on any given occasion one principle may eclipse another with which it conflicts. So we say that a given principle is binding *prima facie*, or "at first blush," but that in the final analysis, all things having been considered, the pull of another principle might turn out even stronger.[121]

In the first editions of *The Principles of Bioethics*, Beauchamp and Childress give the impression that particular ethical judgments are justified by rules, which are justified by the principles, which are in turn justified by ethical theories such as Kantian ethics or utilitarianism.[122] In later editions of their book, Beauchamp and Childress shy away from this method and return to a more prima facie grounded approach. It might be objected

[121] John D. Arras and Bonnie Steinbock, "Moral Reasoning in The Medical Context," in *Ethical Issues in Modern Medicine*, 4th ed., ed. by John D. Arras and Bonnie Steinbock (Mountainview, California: Mayfield, 1995), p. 35.

[122] *Ibid.*, p. 37.

that the latter approach is flawed because it fails to provide a "defensible framework for settling conflicts between competing principles."[123] It also might be objected that the former approach is flawed because it fails to provide a procedure for deciding between conflicting theories.

Wilfrid J. Waluchow argues that one should allow for both fixity and flexibility when considering the merits of different ethical theories.

> The fixity is provided by acknowledging that moral conflicts need not, and perhaps should not, be resolved within a moral vacuum, and that the application of an ethical theory with which one is not entirely happy can nevertheless shed light on the issues in dispute. ... Flexibility arises in acknowledging that competing theories and approaches may well offer insight as well... Reasonable flexibility may even lead us judiciously to exact rules, principles, or values from competing systems as determined by their apparent relevance to the case in question. It may be true that sometimes Mill [a utilitarian] provides a better answer than Kant—and that the tables are reversed other times. ... A single, unified theory would no doubt be preferable. But till such time as one becomes available, it would be imprudent to ignore the existing theories altogether, or subscribe to one and forget about the other(s). ... We must not let our failures to achieve completeness, or our failures to appreciate in all cases the full range of factors at play in particular contexts, blind us to the incremental gains in knowledge that have been made. Perhaps we would do well to heed Aristotle's caution that "precision is not to be sought alike in all discussions. We must be content, in speaking of such subjects [as ethics and politics] to indicate the truth roughly and in outline."[124]

It still might be objected that both Waluchow's analysis and Beauchamp and Childress's principlism lack justification. That is, just as W.D. Ross's ethical theory "is very controversial among philosophers, who are generally suspicious of 'self-evident

---

[123] *Ibid.*

[124] Wilfrid J. Waluchow, "Ethical Resources for Decision-Making," in *Readings in Health Care Ethics*, ed. by Elisabeth Boetzkes and Wilfrid J. Waluchow (Peterborough, Ontario: Broadview Press, 2000), p. 34. Internal citation omitted.

principles' and 'intuition,'"[125] Beauchamp and Childress's appeal to "prima facie" principles and Waluchow's appeal to "fixity and flexibility" could be thought to be philosophically suspect. Hence, if Kantian ethics and the other theories that will be discussed in chapters four and five are not ultimately justified, one is left with the prospect of moral nihilism or moral egoism. In answering this objection, I will abide by Waluchow's advice and briefly consider the answer offered by Aristotelian ethics.

B) A Note on Aristotle

The question of whether Aristotelian ethics could be made to be compatible with this thesis regarding all beings who possess core self-awareness will not be considered here. Suffice it to say that Aristotle's claim that the only good that is the proper subject of moral philosophy is the good that can only be possessed by human, male, citizen, (privileged) philosophers – and his subsequent conclusion that both non-human animals and certain humans are "natural slaves" – begs the question and should accordingly be rejected. After presenting major feminist criticisms of Aristotle's ethics, Ruth Groenhout attempts to salvage the positive elements out of his theory and consolidate them with a feminist ethic of care.[126] Presently, I merely discuss Aristotelian ethics for the limited purpose of answering the above objection and finding a justification to salvage Kantian ethics in light of my critique of it.

---

[125] *Ibid.*, p. 22.

[126] Ruth Groenhout, "A Feminist Critique of Aristotelian Ethics," in *Feminist Interpretations of Aristotle*, ed. by Cynthia A. Freeland (University Park, Pennsylvania: Pennsylvania State University Press, 1998), pp. 171-194.

In section two above, I undermine the basis for Kantian ethics. Moreover, as noted in section one above, both Aristotle and classical utilitarians explicitly admit the bases for their theories are not ultimately subject to proof. In formulating a response to this state of affairs with Aristotle's help, it is useful to contrast how Kantian ethics, utilitarianism and Aristotelian ethics would respond to two archetypical cases. The following "lifeboat examples" are admittedly extreme, but they serve to call attention to the relevant ideas and, given recent and past history, are not that far fetched.

The first example is one of terrorism. A terrorist, accompanied by one innocent hostage, approaches a large city in an aircraft containing a highly transmittable and lethal chemical weapon. Unless the terrorist is stopped, 30 million humans will die. Assume that the only way to stop the terrorist is to destroy the aircraft and its two occupants. Also assume that, if the terrorist is not stopped, both the terrorist and the hostage will survive the ordeal unharmed. Now, consider a second example:

> Every first year law student has read *Regina v. Dudley & Stephens*, a case involving cannibalism. Dudley and Stephens, together with Brooks and Parker, were shipwrecked in a storm that claimed the lives of the remainder of the crew. The four young men were afloat in a small boat that had survived the storm, but the boat had no water and only two small cans of turnips, and the nearest land was over a thousand miles away. After having no food for nine days or water for seven days, Dudley and Stephens killed parker without the latter's consent. They then drank Parker's blood and ate his body. Four days after Parker was killed, a passing ship rescued the men, and Dudley and Stephens were tried for the murder of Parker. ... the jury found specifically that at the time of the murder, Parker was in a much weaker physical condition than the other three men, that it was likely that Parker would have died before the other three men even if he had not been murdered [although his murder did shorten his life], and that there had been no reasonable prospect that the men would be saved.

> Nevertheless, the court found that the defendants' actions were not justifiable as "necessary."[127]

Regarding the first example of terrorism, a pure Kantian would no doubt conclude that the innocent hostage must not be killed. To do so would be to use her or him merely as a means.[128] Hence, the Kantian would permit 30 million humans to die. A pure act utilitarian, however, would probably conclude that the terrorist and innocent hostage should be killed in order to save those 30 million humans. To do so would minimize disutility. Some act utilitarians might object, perhaps arguing that a failure of the state to protect kidnapping victims would result in a general distrust of the government's ability to protect victims of crime and the safety of the nation as a whole would thus be undermined. Hence, all things considered, disutility would be minimized through the single act of not destroying the aircraft and its occupants. Regardless of the merits of this argument, it can be side-stepped by specifying that only a few individuals in a given government agency are aware of the crisis and they can be trusted to keep it secret without having their faith in the safety of the nation undermined. If an act utilitarian introduces further considerations in response to this, more qualifications can be specified to nullify these considerations. Thus, in the end, a pure act utilitarian would conclude that the innocent hostage should die in order to save the 30 million humans in the city below.[129]

---

[127] Francione, *Animals, Property and the Law*, *Op. cit.*, p. 21. Internal citations omitted.

[128] See section two, above for a detailed explanation.

[129] This claim will not be further defended here, as I do not discuss utilitarianism as a whole in this thesis.

Regarding the second example of *Dudley and Stephens*, a pure Kantian would conclude that Parker's murder was immoral.[130] Hence, had the ship that rescued Dudley and Stephens arrived in fourteen days rather than four, a Kantian would have deemed their deaths (as a result of *not* eating Parker) to be morally acceptable under the circumstances. A pure act utilitarian, however, would conclude[131] that it was morally acceptable for Dudley and Stephens to kill Parker in order to save their own lives.

The pure Kantian and pure act utilitarian analyses of these two examples present a quandary. Regarding the first example of the terrorist, the Kantian would sacrifice the interests of the many to protect the interests of the few. This conclusion arguably *conflicts* with the "prima facie" responses of most humans who would maintain that Beauchamp and Childress's principle of beneficence eclipses that of non-maleficence. This prima facie response, however, would *accord* with the opposite utilitarian conclusion.

Regarding the second example of *Dudley and Stephens*, the Kantian would again sacrifice the interests of the many to protect the interests of the few. This conclusion arguably *accords* with the prima facie responses of most humans who would maintain that the principle of non-maleficence eclipses that of beneficence. This prima facie response, however, would *conflict* with the opposite Kantian conclusion.

The quandary described in the above four paragraphs is visually represented in the following table. The conclusions marked in boldface represent what arguably accords

[130] See section two, above for a detailed explanation.

[131] This conclusion is qualified in the same manner as I qualified the first example of the terrorist. Any considerations that an act utilitarian introduces to avoid this conclusion can be side-stepped by introducing further qualifications to the example.

with the prima facie responses of most humans—which conflict with one another depending on the moral theory being appealed to.

	Terrorist Example	*Dudley and Stephens Example*
Kantian Ethics	Let the 30 million humans die	**Let Dudley and Stephens die**
Act Utilitarianism	**Kill the innocent hostage**	Kill Parker

Perhaps Aristotle can resolve this quandary. After carefully and thoughtfully deliberating about the relative merits of the options and ranking those options in a principled and responsive manner[132], an Aristotelian ethicist might conclude that, in the first example, a coward or deficiently intemperate individual would choose to let the terrorist kill millions of humans while a brash or excessively intemperate individual would choose to destroy the aircraft without reasonably ensuring that the chemical weapons would not harm anyone as a result. A courageous or temperate individual, however, would choose to destroy the aircraft and its two occupants with careful consideration and precision in order to save the 30 million humans. Likewise, an Aristotelian ethicist might conclude that, in the second example, Dudley and Stephens were cowards or deficiently intemperate due to their killing and eating Parker. If, however, Dudley and Stephens were brash or excessively intemperate, they would have thrown away their two cans of turnips and wasted their energy and internal water supply by continuously speaking about how confident they were about being rescued within the hour. If Dudley and Stephens had been courageous or temperate, they would have simply left Parker alone and retained a healthy hope for rescue. If an Aristotelian ethicist argued

[132] Aristotle, *Nicomachean Ethics* (Indianapolis: Bobbs-Merrill, 1962), 1106 b36-1107 a2.

for these two conclusions, she or he would ultimately argue that the individuals who chose to undertake the relevant actions would have the specified virtues because those virtues lead to *eudaemonia*.[133] Regarding the question of how one truly knows that this is the case (after carefully and thoughtfully deliberating about the options and, in so doing, ranking them in a principled and responsive way) Aristotle states:

> ... nothing but a good moral training can qualify a man to study what is noble and just—in a word, to study questions of Politics. For the undemonstrated fact here is the starting-point, and if this undemonstrated fact be sufficiently evident to a man, he will not require a "reason why." Now the man who has had good moral training either has already arrived at starting-points or principles of action, or will easily accept them when pointed out.[134]

Waluchow remarks:

> On Aristotle's account, there is a kind of indeterminacy in moral judgments when it comes to deciding on particular courses of action. The variable contexts of moral life prevent us from fashioning hard-and-fast rules or procedures for settling what we ought to do. The best we can do is rely on [practical wisdom], our virtuous dispositions, and the examples set by paragons of virtue. ... Whether this is a weakness in Aristotle's account of moral life is a good question. Perhaps this indeterminacy better reflects moral reality and the perplexing dilemmas with which we are often faced, than theories which purport to provide ready-made answers which fail to emerge when we seek to apply the theories to concrete circumstances. Is it any more helpful to be told that one must maximize utility, or seek to treat humanity as an end in itself, than it is to be told that one must seek a mean between deficiency and excess? In explicitly acknowledging that moral theory can provide only a limited amount of help, Aristotle's theory may in fact be the more honest one.[135]

Thus, in Aristotle, we have good reasons to say that Dudley and Stephens acted immorally. This is consistent with the Kantian emphasis on duty, intention or motivation

---

[133] *Ibid.*, 1095 a16-20.

[134] Aristotle, "Nicomachean Ethics," *Op. cit.*, p. 59.

and the interests of the individual. Indeed, the importance of these concepts is highlighted and they are given additional justification by an Aristotelian analysis. Moreover, the second, third and forth formulations of Kant's categorical imperative are consistent with the above Aristotelian analysis of *Dudley and Stephens*.

C) Salvaged Content of the Categorical Imperative Allows for Beings who Possess Core Self-Awareness

Since there seems to be some justification for keeping the latter three formulations of the categorical imperative, it would be intemperate to abandon them. For example, the content of Kant's second formulation of the categorical imperative that one must "Act so you treat humanity, whether in your own person or in that of another, always as an end and never as a means only."[136] is perfectly comprehensible and applicable without the supposedly determinate axiom of a will that is good in itself. Note however, that its human-centric, "rational" bias has been shown to be misguided in section two, sub-section E, above. The third formulation is as follows: "This principle [i.e. the supreme ground for duty or imperative] I will call the principle of *autonomy* of the will in contrast to all other principles which I accordingly count under *heteronomy*."[137] The content of this formulation also remains intact. That is, the fact that a will is not good in itself does not prevent that will from acting autonomously or respecting the autonomy of another. Note, however, that Kant's claim that the opposite of autonomy, "heteronomy," does not

[135] Waluchow, *Op. cit.*, p. 26.
[136] Kant, *Op. cit.*, p. 198.

contain factors that can have a certain degree of influence on the will has been shown to be misguided in section two, sub-section D, above. This accords with feminist analysis that maintains that autonomy is contingent upon caring relationships of interdependence.[138] Lastly, the content of the fourth formulation of the categorical imperative (which Kant says follows from the first[139]), "...a whole of all ends in systematic connection, a whole of rational beings as ends in themselves..."[140] can also be salvaged. That is, everyone who is an end in him or herself belongs to a "realm" or community in which the moral worth of each end is given the same high consideration. Note, however, that this realm of ends is not absolute and does not necessarily only apply to rational beings, as argued in section two, sub-sections D and E, above. Since content of the first formulation solely consists of the concepts of reason and a will that is good in itself which have been shown to be unfounded in section two, it cannot be salvaged. In chapter six, I will argue that the fourth formulation of the categorical imperative is compatible with a principle of equal consideration of interests that includes all beings who possess core self-awareness. As such, each individual within the realm of ends would remain subject to the second formulation of the categorical imperative. In other words, all beings who possess the capacity for core self-awareness ought not to be used merely as a means.

---

[137] *Ibid.*, p. 205.
[138] Sherwin, *Op. cit.*

CHAPTER 4: Utilitarianism

SECTION I: Bentham

A) Statement of Purpose

The sole purpose of my discussing Bentham is to show that his theory of act utilitarianism is compatible with the "rule" that everyone who possesses the capacity for core self-awareness ought not to be used merely as a means. Therefore, with this purpose in mind, I will not address Bentham's detailed arguments in support of this theory as a whole. Rather, I will restrict myself to addressing the elements within his moral theory that are relevant to the aforementioned purpose.

B) Exposition

Bentham defines the principle of utility as one that approves or disapproves of actions based upon their promoting or opposing happiness.[141] Utility is the property of an object that produces "benefit, advantage, pleasure, good, or happiness, (all of this in the present case comes to the same thing) or (what comes again to the same thing) to prevent mischief, pain, evil or unhappiness *to the party whose interest is considered...*"[142] Regarding the term "interest," Bentham states that "Interest is one of those words, which not having any superior *genus*, cannot in the ordinary way be defined."[143] Nevertheless,

---

[139] Kant, *Op. cit.*, p. 205.
[140] *Ibid.*
[141] Jeremy Bentham, "An Introduction to the Principles of Morals and Legislation," in *Ethics: Selections from Classical and Contemporary Writers*, ed. by Oliver A. Johnson (Fort Worth: Harcourt Brace, 1994), p. 210.
[142] *Ibid.* Emphasis added.
[143] *Ibid.*, note at p. 210.

he maintains that "A thing is said to promote the interest, or to be *for* the interest, of an individual, when it tends to add to the sum total of his pleasures: or, what comes to the same thing, to diminish the sum total of his pains."[144] Bentham defines "happiness," "unhappiness," "pleasure," "pain" and the aforementioned like terms as coming from four sources; physical, political, moral and religious. He calls these sources sanctions because each "is a source of obligatory powers or *motives*: that is, *pains and pleasures*; which ... are the only thing which can operate as *motives*."[145] Regarding the four sources of happiness and unhappiness or pleasure or pain, Bentham gives the example of someone's property, *or even his or her life*, being consumed by fire. If this action happened by accident, then its source is physical. If it happened as a result of the sentence of a political magistrate (who does not necessarily issue the sentence for moral reasons), its source is political. If it happened as a result of the punishment of a neighbour who punishes for moral reasons, its source is moral. If the fire happened as a result of the punishment of God who punishes sinners, its source is religious. Bentham maintains that the physical source of happiness or unhappiness or pleasure or pain is the foundation for and is contained within, all the other sources.[146] He further describes the nature of pleasure and pain in terms of the circumstances that determine the "value"[147] of this pleasure or pain. These circumstances are the intensity, duration, certainty or uncertainty, propinquity or

---

[144] *Ibid.*, p. 210.
[145] *Ibid.*, note at p. 213.
[146] *Ibid.*, pp. 214-215.
[147] Bentham seems to hold that the value of a pleasure or pain consists of its relative weight to other pleasures and pains with respect to a) one individual or b) the sum total of all individuals effected by an action.

remoteness, fecundity, purity and extent of the pleasure or pain.[148] The first four "are to be considered when estimating a pleasure or pain considered each of them by itself."[149] The fifth and sixth circumstances are considered when estimating the value of a pleasure or pain with respect to a given act. All seven circumstances are considered with respect to the total number of individuals affected by an action.[150] So, although Bentham defines happiness and unhappiness in terms of pleasure and pain and does not directly provide a definition of pleasure and pain, the foregoing points that he raises can be used to gain a fairly good understanding of what he means by pleasure and pain. To determine the morality or immorality of an action, Bentham suggests that one should undertake a calculus in which the value of each pleasure and pain produced by the action for each individual and the total value for all individuals taken together are considered. This will allow one to find the overall utility of an action. Bentham holds that the action with the greatest utility or least disutility is the moral action.[151]

From the preceding brief sketch of Bentham's moral theory alone, it should be clear that a being who possesses core self-awareness is to be treated as one "individual" within the utilitarian calculus. In chapter six, I will suggest that even beings who have the capacity for core self-awareness and no other capacity may have an interest in their own continued existence. Bentham allows the utilitarian calculus to be undertaken in "whatever shape" pleasure and pain appear: "whether it be called *good* ... or *profit* ... or *convenience*, or *advantage, benefit, emolument, happiness*, and so forth: to pain, whether

[148] *Ibid.*, p. 216.
[149] *Ibid.*

it be called *evil* ... or *mischief*, or *inconvenience*, or *disadvantage* or *loss*, or *unhappiness*, and so forth."[152] One of the circumstances that determine the value of the good, advantage, benefit or loss in the existence of a being who has core self-awareness and no other relevant quality would be its duration. The duration of the advantage (or interest) of merely existing in a state of core self-awareness in some cases may be limited to the present instant. The sum total of these instances is the duration of the being's existence. Perhaps the other circumstances that Bentham says determine the value of the advantage resulting from an action would apply to such a being as well.[153]

Regardless of whether or not a being who has core self-awareness and no other relevant quality counts for one in Bentham's moral theory, it is clear that sentient beings who have core self-awareness do. As noted in chapter two, section two; all beings who are sentient necessarily possess core self-awareness. Bentham maintained that sentient non-human animals, for example, count for one in the utilitarian calculus:

> A full grown horse or dog is beyond comparison a more rational, as well as a more conversable animal, than an infant of a day, or a week, or even a month old. But suppose the case were otherwise, what would it avail? the question is not, Can they *reason*? Nor, Can they *talk*? But, Can they *suffer*?[154]

---

[150] *Ibid.*

[151] *Ibid.*, p. 217.

[152] *Ibid.*

[153] Bentham says that intensity, fecundity and purity depend upon the individual. Fecundity is the chance that the advantage of an action has of being followed by sensations of the same kind; i.e. a continued state of core self-awareness. Purity is the chance that the advantage of an action has of *not* being followed by sensations of the opposite kind; i.e. a state of non-core-self-awareness. For a being who merely has core self-awareness, the propinquity would be instantaneous. That is, there would be no remoteness.

[154] Jeremy Bentham, *The Principles of Morals and Legislation* (Amherst, New York: Prometheus, 1988), c. XVII, § IV (1781), pp. 310-311, note 1. Citation omitted.

Moreover, Bentham held that to deny that sentient beings are subject to the principle of utility would be to wrongfully degrade them "into the class of *things*"[155] and forsake them "without redress to the caprice of a tormentor."[156] Thus, Bentham's utilitarianism includes sentient beings. Moreover, as mentioned above, Bentham's very broad conception of suffering as consisting of "evil," "disadvantage," "loss" and so on allows for the possibility of including beings who merely have core self-awareness. Since it is uncontroversial that Bentham includes these beings as individuals within the utilitarian calculus, I will now proceed to the more controversial claim that his theory of *act* utilitarianism is compatible with the rule that they ought not to be used merely as a means.

C) Bentham's Mistake

The title of this sub-section is taken from a chapter in Francione's *Introduction to Animal Rights*. Francione, a rights theorist, suggests that "The rule-utilitarian position is ... at least a distant cousin of the rights view because rule-utilitarianism, like rights theory, requires that we follow a general rule even if the consequences of doing so in a particular case would be undesirable."[157] He correctly notes that Bentham is generally regarded as an *act* utilitarian.

Although Bentham claimed that any given human interest could be ignored if the positive consequences of doing so outweighed the negative, he nevertheless completely

[155] *Ibid.*, p. 310. Citation omitted.
[156] *Ibid.*, pp. 310-311., note 1. Citation omitted.

rejected the institution of human slavery. With respect to that institution, Francione argues that Bentham was at least a rule utilitarian because he even rejected "humane" forms of slavery. Moreover, Francione notes how Bentham explicitly compared the treatment of human slaves and non-human animals. In particular, Bentham hoped that non-human animals would acquire the basic *legal* rights that humans currently enjoy in the future. Nevertheless, unlike with humans, Bentham did not question the legal property status of non-human animals.

Francione explains this discrepancy by noting that although Bentham held that both human and non-human animals have an interest in not suffering, he also held that the former have an additional interest in continued existence while the latter do not. This is because Bentham maintained that non-human animals are not "better" or "worse" for being killed due to their lacking "long protracted anticipations of future misery which we have." For Bentham, this supposed qualitative distinction is not relevant to the treatment of other animals as things with respect to their capacity to suffer. He nevertheless held that the distinction is relevant with respect to their lives and deaths. Francione objects that the claim that non-human animals are sentient but have no anticipations of the future is conceptually problematic.[158] Moreover, "we cannot apply the principle of equal consideration if humans have an interest in not suffering at all from their use as resources

[157] Francione, *Introduction to Animal Rights*, *Op. cit.*, p. 132.
[158] See chapters two and six in this thesis.

and animals have no such interest. The result is that [Bentham's] theory ... landed us in exactly the same place as the views he purported to reject[159]."[160]

The arguments in chapters two, six and seven show that the two serious flaws in Bentham's approach that Francione discusses are genuine and present serious inconsistencies to Bentham's position that the legal property status of non-human animals is morally acceptable. Hence, this position of Bentham's should be abandoned. If this is done, and given the foregoing explication of Bentham's act utilitarianism, it follows that his moral theory is highly conducive to the rule[161] or *secondary principle*[162] that all beings who possess core self-awareness ought not to be used merely as a means.

It might be objected that Bentham was neither a rule utilitarian nor accepted the concept of *moral* rights in any instance. That is, he was only opposed to human slavery (*and in light of the aforementioned flaws would have been opposed to both "humane" and "non-humane" non-human slavery*) because, for the vast majority of circumstances and individual actions, the *institution* of slavery *as a whole* is at extreme variance with the principle of utility. Francione responds by arguing that the possibility exists that certain forms of "humane" institutionalised human slavery might maximise overall utility. This possibility suggests that Bentham's opposition to human slavery is based upon his principle of equal consideration (i.e. "each shall count for one and none more than one"),

---

[159] See chapter seven in this thesis.

[160] Francione, *Introduction to Animal Rights*, *Op. cit.*, pp. 131-134. Internal citations omitted.

[161] The term "rule of thumb" is not used due to of its oppressive, sexist connotations. The origin of the term "rule of thumb" was the common law rule that a husband could beat his wife without legal sanction if he used a rod no thicker than his thumb. Davidson, "Wifebeating: A Recurring Phenomenon Throughout History," in *Battered Women: A Psychosociological Study of Domestic Violence*, ed. by M. Roy (1977), pp. 18-21.

his acknowledgment that humans have a similar interest in not being treated as things and his view that the institution of slavery does not maximise overall utility. Francione argues that utilitarianism is inconsistent because it both claims to ensure that all human interests are given equal moral significance and allows for the possibility that the interests of some humans will be valued at "zero" or completely ignored such that they are excluded from the moral community. Francione concludes that utilitarianism should reject slavery regardless of consequences in order to preserve its emphasis on equal moral consideration.[163]

It might be objected that Bentham would remain consistent and, in the tremendously unlikely event that a "humane" form of human *or* non-human slavery would accord with the principle of utility, maintain that such slavery is morally justified. Moreover, Bentham's view that his principle of equal consideration (which includes the consideration of not being "reduced to the class of *things*") is not necessarily inconsistent with allowing certain forms of slavery—in the unlikely event that this slavery was conducive to the principle of utility. That is, in the utilitarian calculus, each individual counts for "one." The interest of each of these individuals in not being used as a thing or a mere means is considered equally, or accorded equal weight. *If*, for example, when weighing the interest of one particular individual in each town not to be "humanely" used as a mere means by having some of her blood forcibly stolen once a year (when doing so is not done for her benefit) against the interests of those in need of blood (and, in turn,

---

[162] This is Mill's term, which reaffirms Bentham's position. See below.
[163] Francione, *Introduction to Animal Rights*, *Op. cit.*, pp. 133., note 9.

those who benefit from those in need of blood, and so on), it turned out that utility would be maximized by overriding the interests of the former, then the former individuals would not count for "zero." In other words, the individuals whose interests were overridden would count for "one," but the other more numerous "ones" would take precedence in order to maximize utility. The individuals whose interests were overridden would still be given equal weight as "one" within the calculus. Bentham may well have both accepted this aspect of his theory and condemned slavery as a general matter, if only because both individual instances of slavery and the institution of slavery as a whole are unlikely to maximize utility under almost any set of circumstances. This accords with Bentham's stated view:

> A measure of government (which is but a particular kind of action, performed by a particular person or persons) may be said to be comfortable to or dictated by the principle of utility, when in like manner the tendency which it has to augment the happiness of the community is greater than any which it has to diminish it.
>
> When an action, or in a particular measure of government, is supposed by a man to be comfortable to the principle of utility, it may be convenient, for the purposes of discourse, to imagine a kind of law or dictate [such as one prohibiting slavery, or using beings who are subject to the principle of utility merely as means], called a law or dictate of utility: and to speak of the action in question, as being comfortable to that law or dictate.[164]

The above objection and corresponding view of Bentham's are consistent with this thesis. In the vast majority of instances, Bentham's utilitarianism would conclude that exploiting or killing those who are subject to the principle of utility is immoral. As such, Bentham's position above entails that the general institution of slavery is immoral.

---

[164] Bentham, "An Introduction to the Principles of Morals and Legislation," *Op. cit.*, pp. 210-211.

Nevertheless, the possibility remains that certain forms of slavery could be justified by the principle of utility—however unlikely that possibility may be. *For the purposes of this thesis, I neither accept nor reject this utilitarian conclusion.* The reason why this thesis (that it is immoral to use core self-aware beings merely as a means) is not – considered in itself – *necessarily* committed one way or the other to the above utilitarian conclusion (which theoretically permits such use as a mere means) is that it avoids the seemingly irresolvable tensions between Kantian ethics and act utilitarianism by ultimately appealing to the method offered in Aristotelian ethics. In chapter three, section three, sub-section B, I argued that both Kantian ethics and act utilitarianism have serious intrinsic problems that can only be circumvented by an Aristotelian rejection of the absolute truth of their respective axioms. Once this argued for rejection is made and the remaining content of the theories is justifiably salvaged, the theories and their divergent conclusions cannot be appealed to with absolute rigidity. Therefore, the utilitarian conclusion that certain "humane" forms of slavery could – despite the extreme unlikelihood of this – be justified by the principle of utility can be rejected while simultaneously and consistently accepting the truth of this thesis. Thus, the foregoing discussion of Bentham shows that his theory of act utilitarianism is compatible with the secondary principle that everyone who possesses the capacity for core self-awareness ought not to be used merely as a means.

D) Implications for Beings Who Possess Sentience or Core Self-Awareness

As argued above, Bentham counts each sentient being as "one" within the utilitarian calculus. The arguments in chapter two show that all sentient beings necessarily also possess core self-awareness. It will be argued in chapter six that beings who possess core self-awareness but who lack sentience may nevertheless have interests (that would be recognised by the principle of utility[165]) regardless of whether or not they can, in Bentham's words, have "long protracted anticipations of future misery." Moreover, as shown in sub-section C above, there are two serious flaws in Bentham's argument that, when corrected in a manner that is consistent with his theory as a whole, necessitate that beings who possess sentience, or even just core self-awareness, should not be used as mere means according to Bentham's act utilitarianism.

SECTION II: Mill

A) Socrates and The Pig

Mill's essay *Utilitarianism* commences with an almost verbatim reaffirmation of Bentham's theory. As such, much of what was said of Bentham in section one above can also be said of Mill. The most important difference between Bentham and Mill's utilitarianism, both in general and with respect to this thesis, is that Mill was unwilling to accept Bentham's view that all pains and pleasures are qualitatively equivalent. Nevertheless, after concluding that the principle of utility "both in point of quantity and

[165] See sub-section B, above.

quality"[166] is the end of human action and is necessarily "the" standard of morality, Mill argued that the principle of utility is defined as:

> the rules and precepts for human conduct, by the observance of which an existence such as has been described might be, to the greatest extent possible, secured to all mankind; *and not to them only*, but, so far as the nature of things admits, *to the whole sentient creation.*[167]

Now, just how far does the nature of things admit this? Also, as a separate matter, do beings who are subject to the principle of utility who are capable of a "high" degree of quality of happiness count for "one" and those who are capable of a lesser degree count for "less than one" in Mill's utilitarian calculus? Regarding the first question, Mill states:

> It is better to be a human being dissatisfied than a pig satisfied; better to be Socrates dissatisfied than a fool satisfied. And if the fool, or the pig, is of a different opinion, it is because they only know their own side of the question. The other party to the comparison knows both sides.[168]

Mill maintains that "A being of higher faculties requires more to make him happy, is capable of probably more acute suffering, and is certainly accessible to it at more points, than one of an inferior type..."[169] For Mill, a mind that has higher faculties is "A cultivated mind—I do not mean that of a philosopher, but any mind to which the foundations of knowledge have been opened, and which has been taught, in any tolerable degree, to exercise its facilities—finds sources of inexhaustible interest in all that surrounds it; in the objects of nature, the achievements of art, the imaginations of poetry, the incidents of history, the ways of mankind past and present, and their prospects for the

---

[166] John Stuart Mill, "Utilitarianism," in *Ethics: Selections from Classical and Contemporary Writers*, ed. by Oliver A. Johnson (Fort Worth: Harcourt Brace, 1994), pp. 266-267.
[167] *Ibid.*, p. 267. Emphasis added.
[168] *Ibid.*, p. 265.
[169] *Ibid.*

future."[170] In anticipation of an objection that would undermine his stance, however, Mill says that the *mere possibility* of having "genuine private affections, and a sincere interest in the public good," present in "unequal degrees" alone consists of higher faculties that are sufficient to be called "enviable."[171]

Thus, Mill maintains that the observance of the principle of utility should be secured by every sentient being but this observance will be constrained by the ability of particular sentient beings to enjoy the aforementioned higher pleasures and suffer from the corresponding pains. Although different sentient beings may be capable of different degrees of pleasure and pain, Mill nevertheless suggests that all beings who possess any degree of sentience (or, as argued in section one above, core self-awareness) ought not to be used merely as means. That is, given the analysis of Bentham's utilitarianism – *which corresponds to Mill's insofar as the application of the principle of utility to all sentient beings through secondary principles is concerned* – the principle of utility should be applied equally regarding the equally similar interests of those concerned. For example, *if* a pig and Socrates both have a similar interest in not suffering as a result of being forcefully confined in a cage, then – *according to the analysis in section one and Mill's reaffirmation of the theory that this analysis is based upon* – both Bentham and Mill's utilitarianism can be correctly said to entail that these interests should be given equal consideration by the principle of utility. Similarly, *if* a pig and Socrates both have a similar interest in not being painlessly killed for the purpose of medical experimentation,

[170] *Ibid.*, p. 268.
[171] *Ibid.*

the position of Bentham and Mill entails that these interests should be given equal consideration within the utilitarian calculus. Since, as argued in section one, the principle of utility considers the interest of all beings who have sentience (or core self-awareness) in not being used merely as means as a secondary principle that always or nearly always maximizes utility, then this interest should always or nearly always be protected.

It might be objected that Socrates' interest in not being forcefully confined or killed is not sufficiently similar to a pig's interest in not being forcefully confined or killed because the former can experience "more acute suffering," is "accessible to it at more points" and has more to lose.

Steve F. Sapontzis responds to this objection as follows:

> Now, feelings are not particularly human nor particular to human-like animals. Both behavioural and physiological evidence indicate that feelings are part of the psychology and worlds of a wide variety of nonhuman animals, including fish and reptiles as well as birds and mammals. Furthermore, there is no reason to believe that intellectually sophisticated beings have feelings to a quantitatively or qualitatively greater degree than do intellectually unsophisticated beings. Jeremy Bentham, who maintained that all moral values derive from contributions to or detractions from happiness, noted seven dimensions to the value of feelings: intensity, duration, certainty, extent, fecundity, purity, and propinquity. So, even if intellectually more sophisticated beings can enjoy a wider variety of feelings, those who are intellectually less sophisticated *can compensate for and even overcome this deficit* through greater intensity, duration, purity, extent, etc., of their feelings. Next time you go to the beach or the park, take a look around and see who is happiest and enjoying the day to the fullest. Is it the intellectually sophisticated human adults, or is it the children and the dogs?[172]

[172] Steve F. Sapontzis, "Aping Persons – Pro and Con," in *The Great Ape Project: Equality Beyond Humanity*, ed. by Paola Cavalieri and Peter Singer (New York: St. Martin's Griffin, 1993), p. 272. Emphasis added.

In *Utilitarianism*, Mill anticipates this reply.[173] Rather than debating the issue further, the present objection can be elucidated and overcome with the aid of the following example:

> If I give a horse a hard slap across its rump with my open hand, the horse may start, but it presumably feels little pain. Its skin is thick enough to protect it against a mere slap. If I slap a baby in the same way, however, the baby will cry and presumably does feel pain, for its skin is more sensitive. So it is worse to slap a baby than a horse, if both slaps are administered with equal force. But there must be some kind of blow—I don't know exactly what it would be, but perhaps a blow with a heavy stick—that would cause the horse as much pain as we cause a baby by slapping it with our hand.[174]

In this example, both the adult horse and the baby humany can be said to have a similar interest in not feeling a similar degree of *physical* pain, although they do not have a similar physical interest in avoiding a hard slap with an open hand.[175] Regarding *mental* pain, both Socrates and a pig can be said to have a similar interest in not being psychologically frustrated, although they *might*[176] not have a similar interest in not being forcefully confined to a cage. Mill, however, would maintain that Socrates can be psychologically frustrated to a greater degree than can a pig because Socrates experiences that frustration differently. For instance, Socrates may conceptually understand the spurious reasons for the Athenian state confining him to a cell and suffer righteous

---

[173] Mill, *Op. cit.*, pp. 264-267.

[174] Peter Singer, *Animal Liberation: A New Ethics For Our Treatment of Animals* (New York: Avon Books, 1975), p. 16.

[175] *Ibid.*

[176] "Sometimes an animal may suffer more because of his more limited understanding. If, for instance, we are taking prisoners in wartime we can explain to them that while they must submit to capture, search, and confinement they will not otherwise be harmed and will be set free at the conclusion of hostilities. If we capture a wild animal, however, we cannot explain that we are not threatening its life. A wild animal cannot distinguish an attempt to overpower and confine from an attempt to kill; the one causes as much terror as the other." (*Ibid.*, p. 17); *Supra*, 172.

indignation as a result, whereas a pig may not understand the profit-motivated reasons for a farmer confining her to a tiny, cement floored, metal barred stall and accordingly experience no indignation at these reasons. From this, one can conclude that – *in cases of true emergency or unavoidable conflict of interests* – the interest of Socrates in not experiencing a greater degree of psychological frustration may outweigh the interest of a pig in not experiencing a *somewhat lesser* degree of psychological frustration—*i.e.* the only degree to which the pig may be capable of experiencing. The same is true of their other interests that are considered by the principle of utility, such as their interest in not being killed or otherwise used as a mere means. The meaning of what constitutes a true emergency or unavoidable conflict will be addressed in chapter six, section three. In that chapter and section, it will be argued that the interests of sentient beings who are capable of different degrees of pleasure and pain should never be accorded greater or lesser weight except in cases of true emergency. The question of whether sentient beings who have different capacities for sentience count for one within the utilitarian calculus is addressed directly below.

B) The Fool

Recall that Mill states that the higher mental faculties are present in humans whose minds have been open*ed* to the foundations of knowledge and *taught* to exercise its own facilities. He also suggests, however, that the possibility of the existence (to a lesser degree with respect to some humans) of the higher faculties of private affections

and an interest in the public good is present in human animals[177] but not present in non-human animals[178]. Again, Mill deems the possession of this possibility to be "enviable."

From the above, one might reach the erroneous conclusion that sentient beings who are capable of having a *degree*-of-affections-and-interest-in-public-good that is the same as the degree possessed by the human who has this capacity to the smallest extent are the only ones who fully count for "one" in Mill's utilitarian calculus. Note that the only instances of sentient beings who have this capacity are human. If this conclusion were sound, the objection could be made that stipulating the *degree*-to-which-the-human-who-has-the-relevant-capacity-to-the-smallest-extent as the degree that is sufficient to have full moral standing under utilitarianism is arbitrary and unfounded. The same arbitrariness and corresponding lack of justification is true of the extent to which the aforementioned capacity must be present in order to be labelled as "enviable," an inherently vague term. Moreover, as a matter of fact, some humans (such as those who, due to brain surgery, for example) have zero capacity for genuine affections[179] while other humans (such as those who are severely mentally challenged) have zero capacity to have a sincere interest in the public good. From the above arbitrary stipulations and matter of fact, two possibilities present themselves: a) both sentient humans and sentient non-humans count for "one" or b) Socrates counts for "one," the "average" human counts for "7/10," human "fools" count for "2/5," humans with affective disorders and severe mental challenges count for "zero" and pigs fit somewhere in between. If the latter

[177] *Ibid.*, pp. 268-269.
[178] *Ibid.*, pp. 263-264.

possibility were true, then any difficulty that utilitarians encounter in attempting to perform their calculus would be infinitely[180] magnified. Although Mill acknowledges the objection that the utilitarian calculus may be difficult to perform and offers a fairly good response to this objection[181], it is hard to conceive of what his reply would be to one that includes a calculation for the diverse capacities for utility possessed by myriad individuals. Indeed, this is a classic objection to Mill's utilitarianism. Thankfully, however, neither possibility (b) above nor the general conclusion that the present objection is based upon are suggested by Mill's text.

Mill, in responding to another unrelated objection, acknowledges that *most* humans are "incapable" of the possibility of having higher mental faculties:

> Capacity for the nobler feelings [i.e. the higher pleasures or feelings for everything noble] is in most natures a very tender plant, easily killed ... and in the majority of young persons it speedily dies away ... Men lose their high aspirations as they lose their intellectual tastes, because they have not time or opportunity for indulging them; and they addict themselves to inferior pleasures...[182]

This loss and addiction can be permanent in some cases, as suggested by Mill's use of "incapable," "killed," and "dies away." Again, the absence is permanent in the case of the "natural fool" for whom there is no possibility of higher mental pleasures. Mill states that most instances of disease, for example, are removable and those that are presently not (e.g. being severely mentally challenged and having an affective disorder that would

---

[179] Daniel Goleman, *Emotional Intelligence* (New York: Bantam Books, 1995), pp. 15, 52-53.

[180] The different capacities to which various individuals are capable of ("higher" and "lower") pleasures and pains are dependent upon multiple factors (e.g. character, upbringing, constitution of the nerves and so on) that are subject to infinite variations (e.g. Someone is "slightly" less pious than Socrates or is "slightly" more sensitive to being tickled).

[181] Mill, *Op. cit.*, pp. 275-277.

preclude Mill's minimum condition for having the capacity for higher pleasures) will probably be curable in the future. "And every advance in that direction *relieves* us from *some* ... [diseases] which deprive us of those ['chances' or circumstances] in whom our happiness is wrapt up."[183] Until the day, however, that these sorts of conditions are removed, would Mill contend that the individuals who suffer from them count for less than "one" in his utilitarian calculus? Would he contend that a 96 year old human who has irrevocably lost her or his taste for the higher mental pleasures in her or his teens, or whose mind was never "opened to the foundations of knowledge and taught to exercise its higher facilities" in the first place and is now permanently closed to them, does not have full moral standing within utilitarianism?

Mill does not directly answer these questions, but his progressive social justice activism and essay *On Liberty* at least indirectly strongly suggest that human "fools" who do not or cannot make use of the higher mental faculties as Mill arbitrarily[184] defines them are regarded as equals. Moreover, following Bentham, Mill maintained that there are many "secondary principles," "subordinate principles" or "intermediate generalizations" to the principle of utility that should be appealed to in order to make the utilitarian calculus less difficult for the majority of cases. These rules consist of (legal) rights and duties and presumably include a rule against slavery.[185] As with Bentham, Mill

---

[182] *Ibid.*, pp. 265-266.

[183] *Ibid.*, p. 269.

[184] Why, for example, do reading poetry and having genuine affections, passions or emotions constitute "higher" pleasures while the pleasures of "swine" such as making love (in what could be described as a poetic manner) not? Note that non-human animals possess a full range of emotion; see: Donald R. Griffin, *Animal Minds* (Notre Dame, Indiana: University of Notre Dame Press, 1995).

[185] Mill, *Op. cit.*, pp. 275-277.

states that the principle of utility should be appealed to directly "only in these cases of [presumably unavoidable] *conflict* between secondary principles..."[186] Moreover, feminist ethicist Susan Moller Okin notes that Mill's liberal philosophy is firmly solidified by his stated conclusion "that the only purpose for which power can be rightfully exercised over *any member* of a civilized community, against his will, is to prevent harm to others."[187] Thus, humans who are severely mentally challenged count for "one" within Mill's utilitarianism *due to their sentience alone*. From this, it follows that all sentient (and, as argued in both this and the previous section, core self-aware) beings have equal standing within utilitarianism. Therefore, in light of the overlap between Bentham and Mill's utilitarianism discussed above and as a contingent matter of fact that is tremendously unlikely to change, these beings ought not to be used merely as means according to any plausible articulation of Mill's utilitarian theory.

[186] *Ibid.*, p. 277. Emphasis added.
[187] Susan Moller Okin, "Mill's Liberal Feminism and Utilitarianism" in *Feminist Interpretations of*

CHAPTER 5: Ethical Empathism

In forming a new moral theory that accords all beings who possess core self-awareness[188] full moral standing, Sztybel first posits a strongest possible case against the view that non-human animals have such standing. Sztybel makes substantial efforts to avoid a "straw person" argument and shows how this argument can be used to refute all of the positions of those who maintain that non-human animals have full moral standing to date. Due to its powerful logic and ability to persuade, Sztybel dubs this strongest possible case against non-human moral standing "Juggernaut." Juggernaut is only refuted by the formulation of a new moral theory that simultaneously takes account of the good and the true in ethics.

Sztybel makes a list of 20 qualities that have traditionally been cited by philosophers as qualities that contribute to richness of life or quality of being. These include artistic or creative endeavour, autonomy, self-awareness, intelligence, language, moral agency, spirituality and so on. Taken together, Sztybel notes that all of these characteristics contribute to one's quality of being and refers to this quality as "Q."[189] Juggernaut is as follows:

> Since this is a practical ethic, I will assume the practically universal (or at least widespread) idea that all beings who have Q also have moral standing. Here, then, is the argument:
>
> 1. Q is not only relevant but also <u>sufficient</u> for assigning moral standing, since all those who possess Q also have moral standing.

John Stuart Mill (University Park, Pennsylvania: Pennsylvania State University Press), p. 211. Emphasis added.

[188] i.e. "consciousness," in Sztybel's words.

[189] David Sztybel, *Empathy and Rationality in Ethics* (Toronto: University of Toronto Press, 2000) pp. 19-24.

2. Q alone is relevant to determining moral standing, since morally, it is the very best such criterion that one could choose amongst all of the competing criteria, and this is true for the following reasons:
(a) That which is best is that which has the most good.
(b) That which has the most good is richest.
(c) Therefore what is richest is best.
(d) Each aspect of Q is good, for it seems better to have than to lack such things.
(e) So Q is richer than any more modest criterion of moral standing such as being alive, sentient, or a subject of a life.
(f) Ethics is a pursuit of the good, or "the good life," and aspires to what is best.
(g) Therefore, morally, we should aspire to holding Q as the best criterion for moral standing.

3. So Q is necessary for having full moral standing.

4. Since Q is both necessary and sufficient for full moral standing, it follows that those who have only some of the criteria do count for something, since they exemplify some riches, but they will have less of a moral claim than those who fully embody all of Q.

5. Non-human animals either lack Q, or might only have a more or less impoverished realization of it, such as in the case of whales, [non-human] apes, and dogs.

6. Non-human animals—as well as plants, rocks, ecosystems, etc.—which utterly lack Q have no moral standing.

7. Those non-human animals who have some Q, such as self-awareness, advanced intelligence, sentience, etc., have a degree of moral standing, but in many cases it might be so limited that it only constitutes a minor ethical consideration.[190]

Sztybel notes that this argument implies that the interests of beings, if any, should be ranked in part by how rich or "superior" they are in awareness, sensitivity and other

[190] *Ibid.*, pp. 19-24

aspects of quality of being.[191] He then goes on to elucidate the argument, its meaning and implications, answer objections and use the argument to refute existing views in favour of non-human animal moral standing.[192] Rather than discuss this further, I refer the reader to Sztybel's text.

Sztybel begins his refutation of his Juggernaut argument by observing that it assumes an "objectivist" view of reality. After a thorough analysis and explanation of the difference between what is both metaphysically and epistemologically objective and subjective[193], Sztybel argues:

> In taking empathy seriously, we acknowledge the full reality of subjects, and not just a universe of objects. Others' points of view are a reality which we must acknowledge, as surely as our own points of view are real (as any casual introspection will reveal). When we empathize with another, we acknowledge the absolute reality of another point of view in another, and try to imagine what it is like to be that other. Merely seeking to surround ourselves with the superficialities of another's view is not enough. We must aim to be considerate of the other's values, emotional dispositions, attitudes, experiences, and so forth, insofar as this is possible. This considerateness makes it possible for one to be and to act *with* others, in an important sense, and not apart from them, merely observing them impartially, from an objective point of view.[194]

Sztybel provides a cogent response[195] to the objection that this sort of empathy is very difficult and overly presumptive. He then goes on to argue that if one does not

---

[191] *Ibid.*, p. 25.
[192] *Ibid.*, pp. 25-78.
[193] *Ibid.*, pp. 107-116.
[194] *Ibid.*, p. 118.
[195] "We are often unsuccessful in such would-be-empathetic imaginings, but we are likely to be more successful in trying to empathize in this way, with attention to the evidence of others' mental states and also their situations, than we are if we make no imaginative effort at all. Making no such effort results in a kind of default perspective which, of course, acknowledges the other's body, and may even register a list of certain mental attributes observed "from the outside," as it were (e.g., irritation, faith in God, etc.), but does not try to know what it is to be that other, from his or her own perspective (even if that, itself, is *abstractly*—hence objectively accepted as existing)." *Ibid.*, p. 118.

attempt to empathise with others, a nullity exists in which knowledge of crucial aspects of reality would and should have been. The absence of this knowledge leads to the objectivist view that treats individual subjects as objects that are composed of a list of mental attributes including "a subjectivity" that is in some way part of this objective list. The list of attributes is described from a neutral perspective that does not allow for a better approximation of the individual to whom they belong.[196] When the other is more or less conceived of as an object, this is a false conception because the other is, in fact, a subject. Sztybel notes that someone who merely lists a number of terms that pertain to some of another's mental states without identifying with those mental states cannot adequately understand the other's experience. For example, a computer that has no conscious point of view that, as such, lacks imaginative empathy could also produce a list of someone's mental attributes if it was outfitted and programmed to respond to certain behavioural and verbal cues. The computer, however, cannot know what it is like to be the other because it lacks subjective, imaginative empathy. Sztybel observes that when an objectivist fails to imaginatively identify with others, she or he unnecessarily and unrealistically differentiates her or himself from these others. The objective "distance" involved in this failure views others as objects or collections of objects and constitutes a

[196] This is not to say that objective facts about subjects are immaterial to their subjective point of view that can be empathized with. For example, if I can imagine the loss and suffering associated with the death of someone I love, it is easier for me to do so for another subject. Without the objective knowledge of a subject's love for her close friend, it would be extremely difficult for me to identify with the subject's experience of the friend's death. Similarly, without the objective knowledge of a subject's capacity for feeling pain, it would be extremely difficult for me to identify with the subject's being cut. As Sztybel argues, however, the mere existence of *objective knowledge* (e.g. an individual's love or capacity for pain) that is inherent to Juggernaut results in unrealistically ignoring other subjective points of view. Empathizing with another's *subjective perspective*, however, results in a more accurate view of reality.

neutral stance in relation to them. Again, objective differentiation between oneself and another point of view completely prevents one from knowing (or at least trying to understand) what it is like to be in that other point of view. Since the objectivist perspective fully precludes the possibility of identifying with other points of view, it is unrealistic. Since Juggernautians differentiate themselves from the ever-constant fact of other points of view, they neutrally regard others as objects and grade them according to their real or imagined having or lacking of Q. *This prevents Juggernautians from recognising what is good or bad for someone from her or his point of view* and leads to them imposing their own values, like objects, on everyone. Again, this results from their necessarily ignoring fundamental aspects of reality; other points of view. Note that, all things being equal, the point of view of subjects includes what is good for them or in their subjective interest. Just as fully identifying with oneself involves fully identifying with one's own good, the same is true of fully identifying with others and their good.

In attempting to gain a more adequate description of reality than that provided by Juggernaut, Sztybel notes that the only other feasible possibility is to empathise with others and their points of view. He argues that these points of view must be important if our own point of view is the only thing that we know directly. Although empathy is an imperfect exercise insofar that it requires imaginative subjectivity, it – unlike the objectivist view of reality assumed by Juggernaut – nevertheless leaves room for the possibility of discovering the value that others have in themselves rather than merely tallying their objectively known value to others. If an attempt is made at empathy, then there will be an approximation, or at least a much more adequate idea, of what it is like to

be another from her or his perspective. "Identifying with another is a categorical act, ideally, in aiming for a full sense of another's experience of reality. It is only done by degrees, in as much as we cannot perfectly identify with others." If the attempt is not made at all, then failure to understand the crucial aspects of reality is guaranteed. Although an attempt at empathy leaves open the possibility for error, the results of such error can be no worse than failing to identify with others at all and treating them as objects as a result of one's *necessarily* faulty view of reality. In empathising with another, one's view of her or his perspective is not necessarily faulty because it allows for the possibility of accurately discovering that point of view to a certain degree. Moreover, certain aspects of another's point of view are more relevant than others for the purpose of identifying with their perspective – including their good – and generating a set of values that is appropriate to a given circumstance. For example, although my identification with a severely mentally challenged human's suffering or imminent death might be incomplete or biased, it is probably much more accurate than my identification with his dream experience of the previous night. The fact that certain degrees of difficulty present themselves in discovering the various aspects of other points of view (including their good) does not entail that one should ignore those realities outright. Otherwise, failure is guaranteed as is the unrealistic objectification of valuing others that Juggernautians engage in. Of course, when engaged in imaginative empathy, care, caution and wisdom should be exercised to the furthest extent possible.

Identification with another's experience inevitably means experiencing things to be of value and, in this way, leads to a firm set of values. Sztybel offers a cogent

argument[197] for why is it is impossible to empathise with someone's causing undue harm except at a bare level of intellectual understanding of motives and so on. *Identification with another necessarily entails identification with his or her good.* "If one truly identifies with another, with a view to his or her own good, then one is 'on-side' with that other, and will not allow impositions, such as harm to that being's good, without just cause." This identification with a view to the other's good is not deterred by the objective fact that a being has or lacks a certain amount of Q. It constitutes a fundamental requirement to gaining knowledge of reality.

Sztybel acknowledges the objection that one's choice[198] to empathise or not

[197] Sztybel's argued-for principle that one should identify with each individual (in a given circumstance) with a view to his or her good requires one to differentiate oneself from those who do not have a view to either their own good or the good of others. Otherwise, a "moral paralysis" would exist in which everyone is accepted just as they are, including their harmful aspects that contradict the original principle and intent of identifying with each individual with a view to his or her good. (*Ibid.*, p. 120) "One cannot empathize with someone's causing undue destruction, except perhaps at a bare level of intellectual understanding of motives, etc., but one can be 'with' him or her in a way that lends itself to advocacy, say, of his or her basic well-being." (*Ibid.*, p. 121) Moreover, Sztybel argues that one can empathise either too little or too much. If others (or oneself) are over-empathised with, then empathising with oneself (or others) is neglected. (*Ibid.*, p. 144) This, as previously argued, would entail having a less accurate view of reality. In the same way, it could be argued that over-empathising with someone intent on causing undue harm necessarily involves neglecting to empathise with those who would be harmed. "One cannot rightly *assume* the point of view of an oppressor without abandoning morality itself, or *contradicting* one's ideal." (*Ibid.* Emphasis added.) In other words, it is contrary to a given individual's good to allow or condone the infliction of harm and, when empathizing with this individual, one attempts to take on that individual's subjective perspective—including his or her good, an aspect of which involves not being harmed. Hence, it would be contradictory to simultaneously fully empathise with the good of a would-be victim and the state of mind of a would-be oppressor. Such identification might be psychologically impossible and, as Sztybel argues, would involve a "moral paralysis." Nevertheless, Sztybel argues that one can momentarily empathise with an oppressor's motives, etc., but not with his or her wrongful perspective insofar that this perspective is mistaken or "undue". (*Ibid.*) The reason why the contradiction of simultaneously empathising with potential victims and potential oppressors should be resolved in favour of the former is ultimately found in the Aristotelian method described in chapter three. In other words, fulfilling the short-term motives of someone who is intent on causing undue harm will probably be accorded less significance relative to the interests of everyone else who is affected by the situation *if each individual involved is fully empathised with to the same extent.* Empathy involves knowledge of and identification with the *interests* of the individual who is empathised with. Again, the moral reasons for favouring some interests against others in cases of conflict are found in chapter three, section three, sub-section B.

[198] Although acts of empathy are often habitual and sub-conscious responses for many individuals,

should depend on how rich or poor a being is in terms of goodness. He responds by noting that ethics is equally concerned with the good and the true. "Identification with others is always required in attempting to have an adequate view of reality. Truth or reality does not alter, depending on how good or convenient or pleasing or rich or useful or rewarding it is for anyone or anything." Hence, the truth of other points of view, including their good, must be adequately acknowledged if one's goal is to have a fuller view of the truth.

> Just as someone who identifies fully with himself or herself and his or her own good would not tolerate being used as a mere means, so one who truly identifies with others in this way would not tolerate similar treatment of others. To subordinate another as a mere means is to objectify him or her by ceasing to (fully) identify with him or her as a subject (supposing what may be unlikely, that a subordinator identified with him or her in the first place). One who identifies with another accepts that being as he or she is at that moment. Only by differentiating one's self from the other, and

---

they nevertheless constitute choices. *An act of empathy involves one imagining what it is like to be another from his or her perspective and identifying with that perspective.* For example, if I look into the eyes of someone who is in pain, I might consciously or sub-consciously recognize that her or his facial expressions convey being in a state of pain. From my past experience or present inference, I can identify with this pain and any corresponding emotions and imagine what it is like. The fact that this identification and imagination constitute a choice is illustrated by the alternative option in this circumstance. After recognizing the individual's facial expressions as painful, I can choose (or strongly motivate or predispose myself to choose) to think of the individual exclusively as an automaton, enemy, one deserving of pain and so on. If any of these characterizations exclusively constitute my thoughts regarding the individual, then there will be no room for imagining what her or his painful perspective is like. For example, many privileged individuals walk by homeless individuals without empathizing with them in the least; from the perspective of these privileged, the homeless are merely obstructions on the sidewalk. Yet, for other privileged individuals this is not the case; they choose (or are predisposed to choosing) to react to their objective observations of a given homeless individual by imagining what it is like to be him or her and identifying with his or her subjective perspective. This is only possible because they have allowed their minds to include much more of an understanding of the individual than "obstruction" or "vagrant." A more extreme example underscores this point. In wartime, soldiers have sometimes been trained to view their enemies as "Jerries," "Gooks" or monsters who deserve no thoughtful consideration. Slaughterhouse employees are trained to view the animals whom they kill as unfeeling automatons. These actions are chosen at some point and serve to block oneself from identifying with others and imagining what it is like to be them from their perspective. Fortunately, most of us do not make such choices most of the time. This is not to say that everyone who can be empathized with is capable of making autonomous choices. Rather, for those individuals who are capable of making autonomous choices, their acts of empathy most often constitute choices or predispositions to make choices.

> grading him or her as an object from the outside, could one not accept that other as he or she is, and instead require him or her either to become something else, or to take lower priority in relation to others—or perhaps even to cease to exist altogether. It is only too easy to turn away from an object, or to be neutral toward it.

Sztybel argues that placing individuals on a hierarchical values scale such as Q – which necessarily entails that one is differentiating oneself from them from an "outside" perspective – lacks rational justification. In particular, the objective way of viewing reality excessively focuses on differences without reason and, as such, exaggerates and misconstrues the moral significance of these differences.[199]

In the foregoing way, Sztybel argues that all beings who can be empathised with ought not to be used merely as a means. It might be objected that this conclusion does not necessarily follow from Sztybel's argument that empathy leads to a firm set of values, including a full identification with the good of the being who is being empathised with. Although the good of a given individual might consist in not being used as a mere means, the fact that this good *is* fully identified with may not entail that the good *ought* to be pursued. Sztybel responds to this objection by acknowledging that identifying with others does not, in itself, suggest what one ought to do. He nevertheless argues that identifying with others with a view to their good has normative implications because doing so highlights the worth of different choices. Importantly, Sztybel recalls that identification with others refutes Juggernaut—a view that completely fails to identify with others with a view to their good. This refutation shows that Juggernaut commits the classist fallacy; the view that a willingness to harm another being is morally justified because she or he is

different in some way. In order to avoid this fallacy, which Sztybel argues is inherent to all forms of oppressive thinking, one must not be willing to harm another because this other is different insofar that she or he lacks Q or some other quality.[200] "Non-negotiable identifying with others with a view to their good, and accepting them as they are, results in a kind of respect for individuals which does have normative implications."[201] In a subsequent chapter of his thesis, Sztybel argues that this moral respect constitutes absolute rights-based protection, as opposed to a utilitarian account.[202]

Rather than addressing Sztybel's strong account of rights, I will simply recommit myself to the weaker position argued for in chapters three and four: *genuine and unavoidable* conflicts between respecting the interests (or "good") of different core self-aware beings (who can be empathised with) in not being used as a mere means should be resolved with an Aristotelian method that takes into account the differing interests of those beings, if any. Sztybel provides additional justification to his argument that just as fully identifying with oneself and one's own good entails that one would not tolerate being used as a mere means, the same is true of fully identifying with others and their good:

> Not having an utterly serious regard for an animal's good, nor assuming that being's good as a good for oneself in one's own choices, even after trying to identify with that animal, only indicates a *failure* to identify, whether or not one's attempt was made sincerely. For that result is simply not consistent with identifying with the other's good. A contaminating, objectifying differentiation has crept in, or has failed to creep out, somewhere along the line. You can be sure that the other values his or her

[199] Sztybel, *Op. cit.*, pp. 118-121.
[200] *Ibid.*, p. 130.
[201] *Ibid.*
[202] *Ibid.*, pp. 249-346.

> well-being, or ought to, if he or she healthily identifies with him or herself [and accordingly has the capacity to do so]. With *claims* that one is identifying with others, then, mere rhetoric does not stand up very well. The proof is in the pudding.[203]

For Sztybel, the pudding is that those who are fully empathised with are not treated merely as a means. He then goes on to provide further argumentation and examples in support of this claim.[204]

It still might be objected that it is epistemologically problematic to morally respect beings who can be empathised with by not treating them merely as a means because empathy might not yield any clearly defined boundaries between oneself and everything else. For example, since both a tomato and I can be crushed, perhaps it is possible to attribute a subjectivity to the tomato so that the prospect of its destruction can be identified with. Also, since both I and the air that surrounds my body are subject to alteration, perhaps it is possible to attribute a subjectivity to the surrounding air so that the prospect of its constituent molecules and atoms being rearranged or dispersed can be identified with. If taken to its logical extreme, this attribution of subjectivity could be extended to the whole of reality itself and any distinction between different 'parts' of reality would be illusory. For practical purposes[205], however, an argument can be made

---

[203] *Ibid.*, p. 131.

[204] *Ibid.*, pp. 131-133.

[205] Certain Eastern philosophies maintain that reality is ultimately composed of thought and exists as one, unbroken whole (or else, cannot be linguistically described at all) in which individual egos or distinctions between self and not-self are mere delusions that are dispelled with an enlightenment experience; a realisation of the true nature of reality. These philosophies, however, nevertheless advocate an ethic of respect or non-violence and compassion to all those who suffer under the delusion of ego or self (awareness) versus not-self. Practically speaking, socially constructed classifications such as there being distinct human races, human and non-human animal species, colours of the rainbow and other aspects of reality can be useful for various human purposes (at least for those of us who are not enlightened—assuming that the claim of the aforementioned Eastern philosophies that their validity can be tested simply

that distinct beings who have core self-awareness are the only beings who can be empathised with.

Sztybel argues that core self-awareness[206] is necessary and sufficient for having the capacity to be empathised with. That is, only core self-aware beings have a subjective point of view that can be identified with.[207] For example, although this is not Sztybel's view, non-existence might be considered "bad" in an *objective* sense for a tomato considered in itself.[208] The tomato, however, has no *subjective* point of view; it has no experience of good or bad things happening to it as a subject. In other words, I as a core self-aware being may be able to subjectively experience being crushed, but it is impossible for a tomato to have this subjective experience. I cannot empathise with a quality in another being (such as the subjective bad experience of being crushed) if that being does not possess that quality. Hence, only core self-aware beings can be empathised with. This is entirely different from it being difficult, for instance, for a human to empathise with a fish's perception of changes in water pressure. In this case, there are two subjects (who, as such, both have a subjective perspective) and one of them is having

---

by following their prescribed methods is true). For example, although race distinctions are not founded in biology, it might be useful for the purposes of conducting HIV/AIDS research to distinguish between groups that have a natural immunity to the disease and those that do not. Somewhat similar points are true of species distinctions. *See:* Richard Dawkins, "Gaps in the Mind," in *The Great Ape Project: Equality Beyond Humanity*, ed. by Paola Cavalieri and Peter Singer (New York: St. Martin's Griffin, 1993), pp. 80-87.

[206] Sztybel uses the term "consciousness," by which he means something akin to what *I* have defined as core self-awareness. He rejects *his* definition of self-awareness, which includes a conceptual awareness of oneself as a psychologically continuous ego, as a criterion for moral standing.

[207] Sztybel, *Op. cit.*, pp. 134-140.

[208] That is, a tomato has a certain structure that entails that it will behave in certain ways within the natural environment that tomato plants have come to exist. This behaviour, induced by the tomato's natural structure, will be thwarted if it (including its seeds) is crushed. In this very limited sense, it might be claimed that it is objectively "bad" for a tomato to be crushed.

difficulty imagining what it is like to be the other in a given circumstance. Conversely, in the case of a human not empathising with a tomato, there is one subject and one object (that, as such, does not have a subjective perspective) and it is impossible for the subject to imagine what it is like to be the object without falsely imputing a supposed subjectivity on the object—a subjectivity that does not exist. Just as Juggernaut's treatment of subjects as objects misrepresents reality, treating objects as subjects also involves a gross misrepresentation of reality.

Thus, it is not at all epistemologically problematic to morally respect all beings who have core self-awareness – as subjects who can be empathised with – by not treating them merely as a means. Rather, it is epistemologically necessary to treat subjects as subjects and objects as objects. The view that only core self-aware beings are subjects who have interests (or a "good") is given further justification in chapter six, section three below. If this is the case, and the interest of core self-aware beings in not being used merely as a means is fully identified with from their perspective and with a view to their good, Sztybel concludes that the result is that this interest is respected.

CHAPTER 6: The Principle of Equal Consideration of Interests

SECTION I: Synopsis

In chapter one, the thesis that all beings who possess the capacity for core self-awareness ought not to be used merely as a means was presented. Chapter two defined and defended the definition of core self-awareness as the bare capacity to experience oneself as existing. Chapters three through five argued that the thesis is supported by Kantian claims, classical utilitarianism and the theory of ethical empathism respectively. In this chapter, I will present a principle of equal consideration of interests that is compatible with the aforementioned moral theories and argue that it offers further grounds for the truth of this thesis.

SECTION II: Explanation and Argument

Ingmar Persson assumes as an axiom that "normal" humans have full moral standing. He argues that non-human great apes have equal moral standing with their human cousins. I apply his arguments to all beings who possess core self-awareness. Persson begins by discussing *human beings*. He presents the principle of equal consideration of interests, as described in this case by Peter Singer, which states that one ought to give equal moral consideration to the similar interests of all those who are affected by one's actions. Singer derives this principle from his interests utilitarianism; the view that one should act to maximize the fulfillment of the interests of all those affected by the action.

Persson notes that Singer calls attention to situations in which acting in accord

with the principle of equal consideration of interests widens rather than narrows the gap between two humans who are experiencing different levels of well-being. Singer gives the example of one human who has already lost a leg and is in danger of losing a toe on the remaining leg and another uninjured human who is in danger of losing a leg. There are only sufficient resources to treat one of these individuals. Since the interest in not losing a leg is stronger than the interest in not losing a toe, the principle of equal consideration of interests indicates that the resources should be used on the uninjured individual. Although increasing the difference in health status between these two humans in this way may be morally acceptable, Persson argues that it might not be acceptable to do so in every situation involving scarce resources. For instance, if one human is hungry and another is slightly less so, then the principle of equal consideration of interests alone indicates that scarce food resources should be given to the former—regardless of the fact the former is rich and the latter is impoverished. Persson notes that situations such as this are largely resolved by the principle of declining marginal unity; the fact that the more one has – the more fulfilled one's interests are – the more difficult it is to increase one's interest fulfillment. Hence, in the forgoing example, the slightly less hungry impoverished human should be given the scarce food resources. In this way, the principle of declining marginal unity tends to mitigate aspects of the principle of equal consideration of interests such that scarce resources will be given to the worse off because doing so maximizes interest fulfillment and narrows the gap between those who are experiencing different levels of well-being. Persson, however, is critical that this good result is not guaranteed by the principle of equal consideration of interests alone.

Moreover, Persson notes that the natural inequalities or factual differences[209] between various humans entail that advancing the interests[210] of some more than others may further the utilitarian ideal. For example, those who are capable of making a greater contribution to the overall good such as scientists, inventors, artists, charitable donors and so on may need to be encouraged by rewards in order for them to use their capabilities to the fullest extent possible. Likewise, those who detract from the overall good by committing crimes may need to be discouraged by the prospect of imprisonment in order to further the utilitarian ideal. Hence, the factual differences between humans entail that treating some humans differently by advancing their interests more than others is justified by the principle of utility. This conclusion regarding rewards and imprisonment is contingent upon whether or not capitalism and the institution of criminal incarceration do, in fact, maximize utility. Regardless, the interests utilitarian justification for the principle of equal consideration of interests remains the same; interest fulfillment is the basis for the distribution of goods.

Persson summarizes the two points discussed so far as follows. Those who are healthier, stronger and so on will tend to live lives that, *to them*, contain more interest fulfillment and the principle of equal consideration of interests only insures that the

[209] Factual differences between individuals are entirely different from any normative claims about different individuals being treated equally or unequally.

[210] At this point, Persson does not use the language of interest advancement. Rather, he states that the natural inequalities among various humans entails that some are more "valuable" than others regarding their capacity to contribute to the utilitarian objective. Since this claim of Persson's is derived from his previous commentary on the principle of equal consideration of interests, and given Persson's text that follows it, it is clear that Persson is referring to interest advancement when he speaks of "value," "well-being" and "quality of life." To avoid ambiguity, I employ the former term.

interests of the worse off will be advanced when it is supplemented with the principle of declining marginal unity. Furthermore, since the factual differences between individuals entail that their capacities to contribute to the *overall* maximization of interest fulfillment will differ, treating them differently in certain situations may be justified by the principle of equal consideration of interests. This leads Persson to entertain an objection that questions the interests utilitarian justification for the principle of equal consideration of interests.

Persson posits the objection that those who contribute to the overall advancement of interests more than others (e.g. doctors) deserve to have their interests furthered more than others and those that detract from the overall advancement of interests (e.g. criminals) deserve to have their interests curbed. Although Persson notes that the concept of moral desert is linked to that of justice (because a thing that an agent deserves is proportionate, in terms of interest advancement, to the act that the agent accomplishes), it need not be for the purposes of this discussion. That is, *if* the interests utilitarian justification for the principle of equal consideration of interests (i.e. interest fulfillment is the basis for the distribution of goods) is called into question by the concept of desert, this entails that one should *not* necessarily give equal moral consideration to the similar interests of all those who are affected by one's actions. This is the primary moral consideration under discussion and it is not inherently linked to any one account of justice. The objector to the principle of equal consideration of interests on the basis of desert may or may not wish to formulate an argument in favour of a desert-based account of justice, but the merits of any such account will not impact the objector's conclusion

that the concept of desert – in itself – undermines the principle of equal consideration of interests. Although Persson assumes that the objector possesses a desert-based account of justice, I make no such assumption. Accordingly, rather than adopting Persson's language regarding whether or not (desert-based) justice demands that humans be treated unequally, I will merely consider the more basic underlying question of whether the principle of equal consideration of interests is undermined by the concept of desert in itself.

Persson begins his response to the present objection regarding desert by citing Singer's rejection of the ideal of equal opportunity because it rewards the lucky who have inherited dispositions to behave in socially useful ways and penalises the unlucky who have not. This leads Persson to conclude that no one deserves to have her or his interests advanced more or less than anyone else because everyone's contributions to the state of the world are ultimately the result of factors that are beyond their control and responsibility. Persson's argument in favour of this broad conclusion is questionable but if the scope of the conclusion is narrowed to merely include contributions that are directly linked to wholly uncontrollable factors, then the argument in favour of it becomes sound. Accordingly, although I do not accept Persson's argument outright, I will outline it below and then modify it to fit my more modest conclusion.

Persson argues that moral credit or blame that is proportionate in terms of interest advancement to an event or state cannot be attributed if:

1 through action or inaction, one made no causal contribution to it *or*

> 2 if one made such a contribution, one either (a) could not have foreseen one's making this contribution or (b) could not have avoided making this contribution (even if one had the requisite foresight).

Persson maintains the truth of this claim because moral desert is commonly ascribed for what one intentionally causes. I will add strength to this claim with the following argument.

If the overall level of interest fulfillment connected with a state of affairs (which was caused by a detraction or contribution to the interests of affected individuals) is offset or compensated for by the distribution of burdens or benefits to a particular individual, then an ascription of desert is being placed upon that individual. This is because a particular individual can only deserve to receive burdens or benefits that are equal in terms of interest fulfillment to a state of affairs *if* that particular individual brings about that state of affairs; the concept of desert—as applied to a given state of affairs—by definition necessarily applies to an individual who deserves to receive burdens or benefits for the existence of that state of affairs. It is, however, impossible to offset or compensate for the overall level of interest fulfillment that was caused by a detraction or contribution by distributing burdens or benefits to a particular individual who did not intentionally make that detraction or contribution; any such offsetting or compensation would necessarily be in response to the factors that actually caused the detraction or contribution rather than the individual in question. Therefore, it is impossible to attribute desert to an individual who does not intentionally cause a state of affairs.

Persson further argues that desert is ascribed to what one intentionally causes because intentionally caused actions are foreseen and avoidable. More specifically, if an

individual does not make any contribution to a state of affairs (that advances or hinders the overall fulfillment of interests of everyone), then that individual does not deserve to have her or his interests advanced or hindered more or less than anyone else, as the state of affairs does not originate from her or him. In cases such as this, it follows that a failure to accord equal moral consideration to the similar interests of affected individuals cannot be justified on the basis of desert. Similarly, if an individual does make a contribution to a state of affairs, but does so unavoidably or without being able to foresee the outcome, failure to apply the principle of equal consideration of interests cannot be justified on the basis of desert. In all of these cases, Persson maintains that (desert-based) justice cannot require that some individuals have their interests furthered more than others. Again, as argued above, Persson's assumption that the present objection involves a desert-based account of justice need not be accepted or rejected.

Persson proceeds to argue that every intentional action ultimately results from conditions that the agent did not make any causal contribution to. He notes that intentional actions result from motivational states such as desires, decisions, intentions, and certain capacities or skills. Although these motivational states may have been caused by earlier intentional actions of the agent, Persson argues that they ultimately result from properties that are determined by genetic or early environmental factors that the agent did not make any causal contribution to. Hence, moral desert cannot be used to justify (and Persson's assumed account of justice cannot require) abandoning the principle of equal consideration of interests. Persson defends this argument:

> I have presupposed determinism, but it changes little if we suppose that intentional actions are occasioned by some condition – say, a decision that lacks a sufficient cause. For, to the extent that this decision is causally determined, it is ultimately due to causes to which one has not contributed, and to the extent that it is undetermined, it is, definitionally, out of reach of all (causal) contributing. ... So regardless of whether the world is completely determined or partially undetermined, the concept of desert lacks application: nobody deserves anything.

Persson concludes[211] that the principle of equal consideration of interests should be retained (and this is what his assumed account of justice requires). That is, benefits and burdens (which advance or hinder interests) should be distributed such that everyone leads lives that, to them, are equal in value (i.e. interest fulfillment) as possible. Persson assumes that no circumstances could require that the interests of some be furthered more than others except for those related to desert—which he also rejects for the foregoing reasons. Persson acknowledges that he has not proven that no other circumstances aside from those related to desert could undermine the principle of equal consideration of interests and suggests that this claim cannot be proven. Persson, however, asserts that until it is shown that circumstances other than those related to desert require that the interests of some be furthered more than others, his argument stands and provides justification for the principle of equal consideration of interests. In particular, no one deserves to have her or his interests advanced or hindered as a result of a state of affairs that she or he did not intentionally cause. With respect to such states of affairs, the non-

[211] Persson responds to a possible objection: "Perhaps some thinkers are tempted to hold that it could be just that some individuals are better off than others because, according to some institutions or conventions - for example, the current institution of property - they are entitled to larger resources. However, this would be just only if the institutions themselves are just, and the latter appears to throw us back on the notion of desert: for instance, the relevant institution of property seems to be just only if everyone deserves the fruits of their labour and has a right to dispose of them as they see fit." (Persson, *Op.*

application of the concept of desert results in the interests of everyone being given equal consideration.

At this point, several objections may be considered. First, Persson's argument that an (undetermined) intentional action that is occasioned by a decision is not connected with any causal contribution because this undetermined decision, as such, lacks a sufficient cause, may be questionable. For example, although an intentional action may be occasioned by a decision that lacks a sufficient cause, it may nevertheless comprise a free choice that is caused by the agent. That is, if free will exists, the sufficient cause of a decision is that free will and – although the existence of this will may ultimately be due to conditions that the agent could not have made any causal contribution to – it is nevertheless capable of making freely chosen decisions that are caused by the agent. This view of soft determinism entails that the agent does deserve to receive burdens or benefits as a result of those choices. If, however, one accepts this soft determinist view and accordingly begs to differ with Persson's response that, to the extent that a decision or intentional action is undetermined it is by definition out of reach of all causal contributing, this has no bearing upon the validity of Persson's argument as it relates to other attributes that do not involve decisions, intentions or will. As such, these attributes are not freely chosen and desert cannot be attributed to them. For example, humans of differing races, sexes, sexual orientations, ages, physical and mental abilities and so on do not deserve to have their interests hindered or advanced as a result of these states of

---

cit., pp. 188-189.)

affairs to the extent that these phenomena are not freely chosen.[212] Since these humans do not deserve to have their interests hindered or advanced in this way, the only logical alternative is that their interests are given equal consideration.

Related to the above objection is the view that intentionally causing a state of affairs is not the only condition for ascribing desert. For instance, individuals are often deemed to deserve to have their interest in freedom impinged upon as a result of their undertaking negligent or reckless actions. Moreover, individuals with wholly inherited talents, such as certain athletes, are often held to deserve their enormous salaries. Persson's two criteria for the non-attribution of desert, however, account for these contingencies.

Negligent individuals must have at least some degree of foresight (per Persson's criterion 2a) if they are to be properly deemed negligent. For example, if a motorist neglects to look at a traffic light before proceeding to drive through it, then she or he is aware of or foresees—on some level—the *possibility* that this action could lead to a very negative state of affairs; a traffic accident. If, however, the motorist was genuinely *incapable* of foreseeing that running red lights carries risks, then she or he would

[212] Race, defined in terms of physical characteristics that are expressions of an individual's genetic composition, is not freely chosen. The same is true of biological sex, which is defined in terms of at least five (not two) X and Y chromosome combinations. The degree to which sexual orientation is or is not freely chosen is a subject of debate and may differ greatly from individual to individual. To the extent that it is freely chosen for some individuals, it is a state of affairs that does not detract from overall interest fulfillment and therefore cannot involve ascriptions of negative desert. The degree to which physical and mental abilities are or are not freely chosen also differs from individual to individual, depending upon whether or not the abilities were inherited, caused by early environmental factors, were the result of accident and so on. One's chronological age at any given moment is obviously not freely chosen. To the extent that all of these characteristics are related to biological classifications, they admit to degree and are ultimately arbitrary. To the extent that the characteristics are socially constructed, a further degree of arbitrariness is introduced.

probably be deemed incompetent to make driving-related decisions and not held liable for the action, although she or he might be prevented from driving in the future for the sole purpose of protecting lives.

Reckless individuals must have at least some capacity to avoid making a causal contribution (per Persson's criterion 2b) if they are to be properly deemed reckless. For example, if a motorist foresees the risk of consuming alcohol before driving and decides to drink and drive anyway, she or he could have avoided the ensuing negative state of affairs. If, however, the motorist was tied down, had alcohol involuntarily funnelled down her or his throat and was forced to drive at gun-point, then she or he would not be deemed liable for the action. Although this motorist may have made an intentional decision to acquiesce to the demand to drive, Persson's criteria 2a and 2b allow for intentionally made causal contributions.

Individuals with inherited talents must make at least some causal contribution (per Persson's criterion 1) to the state of affairs that is embodied, in the case of athletes, by athletic performance. Athletes with wholly inherited talent must cultivate and hone their natural talent. This takes a lot of time, work, effort, patience, perseverance and energy. Rather than deserving to be compensated for the inherent talent itself, these athletes are compensated for both the cultivation and the exercising of their talent. This does not entail that unskilled and unaccomplished athletes deserve to be compensated for attempting to cultivate and exercise athletic performance. *Wholly inherited*, high level athletic performance can exist if and only if a) talent and b) cultivation and execution of talent exist. That is, one of these two necessary conditions for wholly inherited, high level

athletic performance cannot be present without the other. Since the cultivation and execution *of talent* is not present in unskilled and unaccomplished athletes, such athletes do not bring about high level performance and therefore make no causal contribution to any such performance. Likewise, if skilled and accomplished athletes did not cultivate and, more so, did not exercise their talent, they would not have made any causal contribution. That is, without the cultivation and execution of a talent, there would be no athletic performance that people would be willing to pay money for. In both the case of the skilled athlete with wholly inherited talent and in the case of unskilled athlete without any inherited or cultivated talent, complete lack of causal contribution entails that the individuals do not deserve to be compensated for the states of affairs that they are associated with. It is also highly questionable if individuals with wholly inherited talent exist at all as it is commonly accepted that both nature *and* nurture have a role in the manifestation of talent.

It is informative to note that the conception of desert assumed by the present objection regarding inherited traits corresponds to Sztybel's Juggernautian account of the good. This account could be described as ontological desert while Sztybel's empathy-based account could be described as allowing for teleological desert.[213] As Sztybel's refutation of Juggernaut has shown, ontological desert that is based upon having certain qualities (e.g. genetically inherited athletic talent) involves a misrepresentation of reality.

[213] Private communication from Dr. David Sztybel: Department of Philosophy, University of Toronto (Subsequently at Department of Philosophy, Queen's University), February 1999.

Thus, athletes with wholly inherited talent do not deserve to have their interests furthered merely as a result of their having this talent.

In the first two examples above of negligence and recklessness, the individuals would not be deemed to deserve to have the interest fulfillment in their lives decreased by having their interest not to be imprisoned overridden. In the third example of wholly inherited talent, the individual who lets her or his talent waste away would likewise not be deemed to have the interest fulfillment in her or his life increased by having her or his interest in being paid go unfulfilled. Accordingly, Persson's position on desert takes negligence, recklessness and inherent talent fully into account.

It also might be objected that Persson's argument that the principle of equal consideration of interests should be accepted unless circumstances other than those related to desert can be mustered to justify its rejection is invalid. This objection would be applicable if there were no independent justification for the principle of equal consideration of interests. This principle, however, is ultimately justified by utilitarian theory. Moreover, as will be argued in the next section, the principle is also justified by Kantian ethics. Persson himself argues later in his article that the principle of equal consideration of interests (or justice) is not the only moral principle that should be appealed to. Hence, Persson's argument does not take the invalid form of refuting a negative claim and then inferring the truth of a positive claim.

Persson originally states that the principle of equal consideration of interests is ultimately justified by interests utilitarianism.[214] Nevertheless, he later states that his principle of justice (i.e. that supporting the principle of equal consideration of interests due to the inapplicability of the concept of desert) has a source that is independent of utilitarianism.[215] Presumably, this is because the non-application of the concept of desert is not based in utilitarian theory but nevertheless results in the interests of everyone being given equal consideration.

Persson argues that his principle of justice (or giving similar interests equal moral consideration because no one deserves otherwise) is not the only ethical principle because, if it were,

> it would be morally indifferent whether we equalise the value of lives by raising the value of some or lowering the value of others [say, by paralyzing all able bodied individuals so that their lives are equal in value to the currently physically challenged].[216]

Hence, in order to avoid this absurd result of a principle of pure equality, Persson resolves that some sort of ethical principle is needed to introduce considerations of benevolence and support advancing the interests of individuals rather than violating them. Persson suggests interests utilitarianism can serve this purpose. If this is done, Persson notes that considerations of equality may conflict with those of benevolence or utility in certain circumstances. A method to resolve such conflicts has already been suggested in chapter three, section three.

---

[214] *Ibid.*, p. 184.
[215] *Ibid.*, p. 189.
[216] *Ibid.*

At this point, Persson has only discussed equality among "normal" human beings. It is clear, however, that the principle of equal consideration of interests applies to all beings who have interests or – in Persson's words – to all those who are capable of having value (understood in terms of interest fulfillment) in their lives. For example, Persson observes that just as it is contrary to the principle of equal consideration of interests to violate the interests of the mentally challenged because of their conditions (states of affairs that they did not intentionally cause), it is contrary to the principle of equal consideration of interests to violate the interests of non-human great apes because of their genetic make-up, degree of rational capacity, and so on (states of affairs that are completely beyond their control).[217] Therefore, equality requires "that both groups be so treated that the value of their lives to them becomes as equal as possible to the value to others of their lives."[218]

Persson observes that this does not imply that all beings who are capable of having value (or interest fulfillment) in their lives should be treated the same in every respect. For example, Persson's conclusion does not imply that chimpanzees should have a right to vote; since they lack the mental capacities that are suited to the electoral process, distributing this 'benefit' to them would not further their interests. It also does not imply that beings who lack core self-awareness such as humans whose only brain function is in their brain stems, plants and inanimate objects are subject to the principle of equal consideration of interests. This question will be addressed in section four below. I

[217] *Ibid.*, pp. 189-190.
[218] *Ibid.*, pp. 191.

will now argue that, in addition to being grounded in utilitarian theory, the principle of equal consideration of interests is also justified by Kantian ethics.

SECTION III: Kantian Basis for Equality

The principle of equal consideration of interests is not linked to any one moral theory. Although Persson derives this principle from interests utilitarianism, it can also be found in Kantian ethics. In order to illustrate this point, I will briefly summarise the key points of Persson's argument regarding desert and the equal consideration of interests and show how they accord with Kantian theory.

The three versions of Kant's categorical imperative that were salvaged in chapter three, section three are all concerned with the interests of the individual; either the interest in not having one's autonomy violated or, stated with different emphasis, the interest in not being used as a mere means. The final version of the categorical imperative in particular protects the interest of *everyone* within the moral realm in not being so used. In other words, everyone is given this same equal moral consideration which protects their basic interests. Persson's conclusion that, in situations involving scarce resources, the principle of equal consideration of interests only insures that the interests of the worse off will be advanced when it is supplemented with the principle of declining marginal unity conforms to Kantian ethics. To use the example that Persson cites, the Kantian positive duty to help others would arguably require scare medical resources to be given to someone who is in danger of losing a leg rather than to someone who is in danger of losing a toe. The Kantian positive duty expresses the imperative to treat everyone as ends

in themselves; it is the flip side of not treating individuals as a mere means. Permitting the loss of a leg is a more serious violation of the duty to treat the affected individual as an end in him or herself because the interest associated with this instance of duty is stronger than the interest associated with the duty not to permit the loss of a toe. Likewise, Persson's conclusion that the factual differences between individuals entail that the interests of some should be advanced more than others in certain contexts is compatible with Kantian ethics. Kant would maintain that duty alone – not external incentives – should be the motivation to treat everyone in the moral realm as ends in themselves. This, however, does not necessarily prohibit an individual from autonomously choosing to further the interests of some more than others by paying them for products or services that further the individual's own interests—as Kant's example of the merchant who charges the same price for all regardless of market experience partially suggests. Moreover, imprisoning those who use others merely as a means accords with the positive duty to prevent such harmful actions from taking place.

Following Persson's argued for conclusions, the point is not that some individuals *deserve* to have their interests advanced more than others—at least when doing so is in response to states of affairs that they did not intentionally cause. Rather, desert is not applicable in such circumstances and this results in the application of the principle of equal consideration of interests. The non-application of the concept of desert and its principled result has its own internal logic and, as such, it is not intrinsically based within utilitarianism or Kantian ethics. Nevertheless, in order for the principle of equal consideration of interests to be supported with positive arguments, appeal to

utilitarianism and Kantian ethics is required. With respect to the latter, this support is related to the final version of the categorical imperative, as argued above. Stated simply, the principle of equal consideration of interests requires that similar cases, and similar interests, be treated similarly and this principle is found in both utilitarian and Kantian theory. I will now argue that this principle does not apply to beings that lack core self-awareness because such beings do not have interests.

SECTION IV: Core Self-Awareness as Necessary and Sufficient for Being Subject to The Principle of Equal Consideration of Interests

If the interests of a being are advanced or hindered, he or she must be capable of experiencing the things that cause this interest advancement or hindrance. That is, a being without experiences cannot experience the things that would advance, hinder or cause no change to her or his interest fulfillment. Moreover, a being (regardless of whether or not it has a sense apparatus) that lacks core self-awareness cannot experience anything, let alone the things that would advance or hinder interests. The following examples will illustrate this point.

When a ray of infra-red light (heat) strikes an air particle and consequently moves it, the particle does not experience the light. Likewise, when the light is turned off, a strip of metal bends in response to the change in temperature, and that metal connects a circuit in a thermostat that turns a furnace on, neither the metal, thermostat, nor furnace experience the light being turned off. Similarly, when the infra-red light is turned back on and transmitted in specific frequencies that are arranged in certain patterns which pass

through the lens of a video camera, the light impulses are converted by a computer program into a series of zeros and ones, and those numbers are fed into the computer program that switches on a circuit that is attached to a toy car, neither the camera, computer program, nor toy car experience the light being turned on. In the same way, when the light is turned off again and an extremely complex chain of stimulus-response actions takes place that culminates in a plant's leaf curling, neither the leaf, other plant parts, nor the plant as a whole experience the light being turned off.

In all of these examples, it is true that the materials involved undergo change. This change, however, is not properly termed "experience." Through their illustration of cause and effect, the above examples suggest that the air particle does not experience the light, but is simply bumped by it. Likewise for the plant, except the series of events in that case are more complex. That is, beings that lack core self-awareness are incapable of experiencing anything[219], which includes the things that advance and hinder interests. Since it is necessary for a being to experience such things in order to experience a loss in interest fulfillment, beings that lack core self-awareness cannot experience any loss (or gain or no change) in interest fulfillment. Therefore, beings that lack core self-awareness do not possess lives or existences that contain interests. Accordingly, the principle of equal consideration of interests does not apply to beings that lack core self-awareness that, as such, do not have any interests.

Conversely, beings who have core self-awareness are by definition aware of themselves, the things they experience, and themselves in relation to the things they

experience, including the experiences (thoughts, events, and so on) that advance or hinder their interests. Again, a being without experiences cannot experience things that would advance, hinder or cause no change to her or his interests. Without the subjective *experience* of interests, there can be no subjective interests that the principle of equal consideration of interests requires. Hence, core self-awareness is a necessary condition for having interests that are subject to the principle of equal consideration.

The above argument does not assume that experiences are the only things that advance or hinder interests. Rather, it concludes that the capacity for experience is a necessary condition for a being to have interests *in his or her life* in the first place. It might be objected that a vegetable, for example, has an 'interest' in not being pulled from the ground despite the fact that vegetables lack core self-awareness. Moreover, although vegetables may further the interests of the beings who have core self-awareness who eat them, this does not negate the claim that vegetables may have 'interests'; circumstances can be conducive or disruptive to their biological functions.

Vegetables, however, do not have interests *in themselves*. That is, vegetables are incapable of experiencing anything—relative to *them*. As such they cannot, *in themselves*, experience an increase, decrease or no change in interest fulfillment. *This is the only sense of the term "interest" that is relevant to the principle of equal consideration of interests*: recall that, since no one deserves to have the interests *in their lives* decreased as a result of a state of affairs that they did not intentionally cause, the principle of equal

[219] See chapter two in this thesis.

consideration of interests demands that benefits and burdens be distributed such that everyone has lives that, *to them*, are as equal as possible to the lives of everyone else.

Unlike the previously discussed particle, thermostat, computer-guided toy car and plant or vegetables, core self-aware beings are capable of having experiences[220] and their having certain sorts of experience necessarily entails that they have interests. For example, a core self-aware being who is sentient has an interest in not being tortured because torture is painful and pain is an inherently negative experience. Similarly, a core self-aware being has an interest in not being painlessly killed because death involves the destruction of all of the actual or potential positive experiences in that being's life. All other things being equal, it is patently clear that both painful experience and the annihilation of the capacity for positive experience itself are contrary to the interests of those who can experience.

Even the most basic and theoretical form of core self-awareness (i.e. that without sentience) may be a sufficient condition for having an existence that contains at least one interest. That is, this form of core self-awareness itself might constitute an experiential interest. Is it in the interest of a simple being who possesses core self-awareness and no other relevant characteristic to continue existing as a being who has core self-awareness?

Recall that a being must be capable of experience in order for the being to experience an increase, decrease or no change in interest fulfillment. Moreover, as argued and further explained in chapter two, beings who possess core self-awareness *experience*

[220] *Ibid.*

themselves as having this core self-awareness. Do beings with core self-awareness and no other relevant characteristic have an interest in this experience?

Just as *continuing to live a life* that is relatively free of pain is in the interest of a more complex being, it is possible that *continuing to exist in itself* is in the interest of a being who merely has core self-awareness without sense experience or other attributes. Perhaps the latter being's very existence could be described as being the sole positive[221] aspect of its existence. Despite the fact that this simple being at time 2 does not remember its past that occurred at time 1, the core-self at time 2 may be in the interest of that same core-self as it exists at time 2. Although this being who merely has core self-awareness *may not conceptually understand that it has an interest in continued existence*, it is possible that the existence of this core self-aware being is nevertheless in that being's interest. That is, the subjective awareness of *an interest in* continued existence is not necessary for the experience of that existence to be in the interest of the being who experiences it. By analogy, the pleasant experiences of an individual who does not undertake the process of reflecting upon his or her experiences or interests may nevertheless have an interest in having those pleasant experiences without the individual having conceptual or non-conceptual knowledge of this. More specifically, humans in comas, infants and humans who suffer from temporary depression may nevertheless have an interest in continued life because they can either have positive experiences in the present, in the future or both. Thus, subjective awareness of interests is not necessary for

[221] If, however, a core self-aware being had mostly negative and unavoidable experiences, as in the case of some terminal illnesses, it might or might not be argued that this being does not have an interest in

the subject to have interests in the first place—although the subject's determination of what her interests are and what constitutes an acceptable level of their fulfillment may be subjective. Accordingly, all beings who have core self-awareness should be assumed to have the capacity for living a life that contains the capacity for interests and are consequently subject to the principle of equal consideration of interests. Beings that lack core self-awareness, however, cannot experience anything and accordingly cannot experience an increase, decrease or no change in interest fulfillment. Again, without the subjective experience of interests, there can be no subjective interests that the principle of equal consideration of interests requires.

Regardless of whether or not core self-awareness is a sufficient condition for having interests or if a simple being who merely possesses core self-awareness has interests that can be violated, it is certain that most if not all incompetent human patients have interests that are subject to the principle of equal consideration. Even an incompetent patient in a coma, for example, can have his or her interests hindered by being subjected to pain or can have the capacity for having those interests extinguished by being robbed of all his or her dream-state experiences (i.e. killed). Likewise, as noted in chapter two, most if not all non-human animals have experiential interests that are subject to the principle of equal consideration of interests. Human patients in Permanent Vegetative States (PVS) whose higher brain function is destroyed and whose bodies are kept alive by their brain stems, however, do not possess core self-awareness. As such,

---

continued life. The capacity for experience is merely a condition for having an interest continued existence in the first place. Extremely negative experiences, or other factors, might mitigate against this condition.

they cannot have interests and are not subject to the principle of equal consideration of interests.

As previously argued, the principle of equal consideration of interests is ultimately supported by higher moral theories. Likewise, Persson's assumption that "normal" human beings have full moral standing is also supported by higher moral theories. If the view that all "normal" humans have full moral standing were rejected, it would be pointless to apply the principle of equal consideration of interests to them or anyone else. If, however, all "normal" human beings are accepted as having full moral standing, then the principle of equal consideration of interests must be applied to all beings who are capable of having interests; core self-aware beings.

It might be objected that human babies, severely mentally challenged human adults, human patients in comas and some species of non-human animals have a smaller capacity to experience things that advance or hinder interests than "normal" human beings do. This would be due to the claim that their degree of core self-awareness (or sentience) is smaller than that possessed by others. Accordingly, it might be argued that all beings who possess core self-awareness are not subject to the principle of equal consideration of interests because that principle cannot apply equally to them due to their unequal capacities for experiencing things that are in their interest.

Persson acknowledges that his principle of justice (i.e. the principle of equal consideration of interests that results from the inapplicability of the concept of desert) is constrained in practice by the natural inequalities between individuals. As previously noted, equality requires that the interests of some should be advanced more than others if

circumstances dictate that doing so would be in accord with a higher moral theory. One such circumstance might be a "life-boat" situation in which saving the life of one individual necessarily entails that another individual will die. Francione gives the example of a human child and a dog being trapped in a burning house where there is only time to save one. He varies this example to involve a human stranger and human family member, a very old human and a young human and a virtuous human and a mass murderer. Francione's point is that, in situations of true emergency, there might be many good reasons to choose the human child over the dog, the familiar over the unfamiliar, the very old over the younger and the virtuous over the vicious. What must be avoided, however, is using extreme emergency situations as a foundation for general moral principles.[222] More importantly, one must not throw individuals into the burning house and then ask whose life should be saved after the fact. That is, Francione argues that it is disingenuous and morally unacceptable to arbitrarily label some individuals as those who can be used merely as a means and then supposedly "balance" the interests of those individuals against others who have been deemed to be individuals who cannot be used merely as a means. The outcome of such biased balancing is predetermined.[223] For example, Nazi vivisectionists regarded mentally challenged humans and others as mere means to the advancement of medicine. They supposedly balanced the basic interests of the mentally challenged and others against the interests of future medical patients. The Nazis held that the former made "sacrifices" so that the latter would not die of

[222] R. M. Hare, *Moral Thinking: Its Levels, Method, and Point* (Oxford: Clarendon Press, 1981), pp. 47-48.

hypothermia and other ailments. This, of course, is ludicrous; their victims were simply exploited as tools for medical advancement. Hence, true emergency situations aside, beings who have a smaller capacity to experience things that advance or hinder interests than "normal" human beings do are nevertheless subject to the principle of equal consideration of interests.

Persson observes that an implication of the principle of equal consideration of interests is that burdens and benefits must be distributed to those who, through no fault of their own, have been born with different characteristics such that *their lives* are, *as much as possible*, equal in terms of interest fulfillment to the lives of everyone else. Individuals who have a smaller capacity to experience the things that advance interests than others do not cause themselves to have these capacities. So, for example, the level of interest fulfillment in the life of any *specific* severely mentally challenged human being should, *as much as possible*, be made to be equal to the level of interest fulfillment in the lives of human beings who are not severely mentally challenged. Again, as argued above, if the principle of equal consideration of interests is applied to "normal" human beings, it follows that severely mentally challenged human beings and all other beings who have core self-awareness are subject to it. As such – as will be concluded in the next section – they ought not to be used as a mere means. In situations of true emergency, however, individuals who are subject to the principle of equal consideration of interests and who have a greater capacity to experience interest fulfillment may have their interests balanced

[223] Francione, *Introduction to Animal Rights*, *Op. cit.*, pp. 151-160.

against the interests of others who are also subject to the principle of equal consideration of interests but who have a lesser capacity to experience interest fulfillment. The aforementioned "life-boat" situations qualify as examples of true emergencies whereas a situation of somewhat scarce resources within a hospital in an affluent nation does not. When attempting to determine whether or not a given situation constitutes a true and unavoidable emergency, it must be remembered that all of beings who have core self-awareness are subject to the principle of equal consideration of interests.

This aspect of the principle of equal consideration of interests will be repugnant to many of those who are sympathetic to Kantian ethics. Indeed, it is morally repugnant to me as well. For those who react this way, I respond by drawing attention to the only other logical option in situations of true emergency: if, in Francione's burning house, there are two beings who possess core self-awareness, there is only time enough to save one and the possibility of success is equal in both cases, then one should decide by flipping a coin. This is true regardless of who is in the burning house; a "normal" human being versus a severely mentally challenged one or a human child versus a dog. I am happy to accept this conclusion but I cannot defend it with reference to the principle of equal consideration of interests. For, in a true emergency in which all external circumstances are equal, the only remaining criteria to resolve the dilemma are internal qualities. Otherwise, a coin must be flipped. In any case, the resolution of dilemmas such as these is beyond the scope of this thesis.

SECTION V: Implication

If the moral claim that all "normal" human beings ought not to be treated merely as means is true, the principle of equal consideration of interests demands that the same is true of all beings who possess core self-awareness. For example, if my interest in not being used merely as a means to the advancement of medical science by being subjected to pain, forced confinement and death is sufficiently similar[224] to a rat's interest in not being so used, these interests must be given equal moral consideration.

[224] See chapter two, section two, sub-section A and chapter six, section three in this thesis.

CHAPTER 7: Core Self-Awareness and Personhood

SECTION I: Legal Personhood

In chapters three through six, I have defended the thesis that all beings who possess core self-awareness ought not to be used as a mere means. This is a moral claim. In chapter one, section two, I argued that moral claims of this sort must be enforced with legal prohibitions. It is impossible to legally enforce the moral claim made in this thesis with any status other than legal personhood.

Francione contrasts the notion of legal personhood with that of legal property and notes that the latter has generated a rich and highly complex philosophical and jurisprudential literature.[225] After discussing the reasons for this complexity, Francione notes that for his purposes (and mine), problems related to the complexity of the concept of legal property can be avoided because they are largely concerned with real property land estates and not personal property.[226] Francione is concerned with the latter because human and non-human animal properties (i.e. slaves) are characterized as personal property. Francione shows that the fact that non-human animals are currently legal property and have been throughout recorded history is unquestionable. He does this by

[225] Francione, *Animals, Property and the Law*, *Op. cit.*, p. 33.

[226] "Indeed, as one scholar has correctly argued, the confusion between 'property' and 'ownership' is particularly important 'if property is in the earth. ...It has been argued that 'the' modern era of ownership emerged in England, at least, in the seventeenth century, and part of the evidence for that thesis concerns new powers given to persons holding various sorts of estates in land.' Although discussions of 'property' and 'ownership' consume large sections on real property, discussions of these topics in books on personal property are usually quite succinct. This, of course, is not to say that the jurisprudential issues that concern personal property, as opposed to estates in land, are insignificant or unimportant. It is only to say that these issues may be simplified when—as in the case of much, but certainly not all, personal property—the *object* of property is readily identifiable and the incidents of ownership, including the identification of an 'owner,' do not involve the complexities that arise when estates in land, including the innumerable legal distinctions that allow ownership to be divided among many different people sometimes living over many generations,

citing the various pre-eminent legal scholars who have written on the subject and by drawing attention to aspects of the history of law.[227] Moreover, non-human animals are both the object of property and the incidents of ownership that constitute the property relationship.

In arguing that the ownership of non-human animal property is no different from the ownership of other sorts of personal property or chattels, Francione notes that (according to most legal theorists) property cannot have rights and legal relations and reciprocity cannot exist between property and persons. Property does not have any recognised inherent interests that must be respected. Francione further notes that the question of whether or not a being is property that is, as such, not considered to be a carrier of interests or is a person who is considered to be a carrier or interests is not an empirical one. Rather, it is a moral question that is answered with moral reasoning. The characterisation of non-human animals as property necessarily results in the doctrine that Francione describes as legal welfarism.[228]

When describing legal welfarism as applied to both human and non-human animals, Francione summarizes the thesis of his book *Animals, Property and the Law* as follows. Just as the restrictions on the use of different kinds of property in general do not establish rights for that property or impose duties upon property owners that are ultimately directed to the well-being of the property itself (independent of any benefit for owners or other stakeholders), restrictions on the use of non-human animal property such

---

are involved." *Ibid.*, pp. 33-34.

[227] *Ibid.*, p. 34

as anti-cruelty laws or laws regulating non-human animal experiments likewise do not establish rights for non-human animals or impose duties upon humans that are ultimately directed to the well-being of the non-human animals. Instead, laws that restrict the use of non-human animal property require that the interests of human and non-human animals be balanced against each other in order to determine if a particular use of non-human animal property causes "unnecessary" suffering or is "inhumane." Francione, however, notes that human interests are protected by rights whereas the interests of non-human animals as property are not. Moreover, humans have a right to own property.

Since non-human animals are property, have no legal rights and are the objects of the exercise of human property rights, the result of the balancing process between human and non-human animal interests that is required by welfare laws is necessarily predetermined: the interests of non-human animals never outweigh those of humans regardless of the relative triviality of the latter and the gravity of the former. Francione argues that legal welfarism is a normative theory that is implicit in the law and contains assumptions that are scarcely recognised and discussed in case law or academic comment. A core assumption of legal welfarism is that legal terms (when applied to other animals) such as "unnecessary" pain and "inhumane" treatment are interpreted within a context in which non-human animals are legal property, the institution of property ownership has immense cultural and social importance and legal doctrine generally serves to protect and maximise the value of property. Both this context and the aforementioned underlying .

[228] *Ibid.*, pp. 34-36.

normative view obscure any purported difference between non-human animal property and other forms of personal property. Actions that involve "unnecessary" suffering and "inhumane" treatment are implicitly defined by the law as those that fail to exploit non-human animals in the most economically efficient manner possible. The only actions that are legally prohibited are those that do not generate social benefit or decrease overall social wealth, such as inflicting gratuitous pain. If, however, the infliction of pain or death is part of an institutionalized exploitation of non-human animals for a socially accepted benefit such as taste enjoyment, the advancement of medical science or other profitable enterprises, the activity is necessarily legally sanctioned. Francione concludes that this state of affairs is a logical consequence of the fact that humans are persons who have rights and non-human animals are property.[229]

> Our legal system is quite adept at making it appear as though disenfranchised groups receive legal protection. By directing our attention to issues that are often quite tangential, legal discourse steers clear of the more important fundamental moral and economic assumptions upon which the legal system ultimately rests. One need only read cases from the eighteenth and nineteenth centuries concerning slavery; these cases read with the same formality as cases just decided just yesterday by the United States Supreme Court and solemnly discuss the same issues of due process and rights. Nevertheless, the slave cases avoid completely the issue of the justice of the institution of slavery and assume that the legal system functioned to provide adequate legal protection to those who were enslaved.[230]

The situation is exactly the same in the case of non-human animals as legal property. Francione concludes that the law produces the illusion that a vulnerable group is provided with adequate protection despite the fact that this vulnerable group is treated primarily, if

[229] *Ibid.*, pp. 4-6.

not exclusively, as a means to the ends of others as a matter of law. This apparent contradiction is due to the fact that the law never examines the fundamental assumptions that comprise the basis of institutionalised non-human animal exploitation.

Throughout the remainder of *Animals, Property and the Law*, Francione provides excellent analyses of the relevant legal concepts, cases, acts and statutes, and presents an unassailable argument that non-human (or human) animals as legal property (contrasted with legal persons) can never be accorded any meaningful legal protection, let alone have their interest in not being used as a mere means be respected. Francione further adds to this argument in *Introduction to Animal Rights*.[231] Regarding legal personhood, Francione concludes:

> [Human] slaves were regarded as chattel property. ... For a while, we tried to have a three-tiered system: things, or inanimate property; persons, who were free; and, depending on your choice of locution, "quasi-persons" or "things-plus"—the slaves. But, as we saw[232], that system could not work. We eventually recognized that if slaves were going to have morally significant interests, they could not be slaves anymore. We recognized that the moral universe [that ideally underlies the positive law] is limited to only two kinds of beings: persons and things. "Quasi-persons" or "things-plus" will necessarily risk being treated as things because the principle of equal consideration cannot apply to them.
>
> ... If we are going to apply the principle of equal consideration to animals and treat animal interests in not suffering [or being killed] as morally significant, then we must extend to animals the basic right not to be treated as our resources.
>
> ...the extension of this one right to animals will profoundly affect our use and treatment of animals. We will no longer be able to justify our institutional exploitation of animals for food, biomedical experiments,

---

[230] *Ibid.*, p. 6.
[231] Francione, *Introduction to Animal Rights*, *Op. cit.*, pp. 86-90, 98-101, 126, 206-208, 219-220.
[232] *Ibid.*

> entertainment or clothing. All of these uses *assume* that animals are resources and have no moral status.[233]

Thus, if the view argued for in the foregoing chapters that all beings who possess core self-awareness (e.g. human and non-human animals) morally ought not to be treated merely as means and this status morally ought to be enforced by the law is correct, then all beings who possess core self-awareness morally ought to be legal persons.

It might be objected that moral personhood cannot exactly map onto legal personhood because the latter has a large pragmatic or policy-oriented component that does not apply to all core self-aware beings. The conclusion that it is morally wrong to use core self-aware beings as a mere means can map onto legal personhood, however, does allow for the fact that the pragmatic or policy-oriented component of legal personhood may contain many specific applications that do not apply to all classes of legal persons. If the *possibility* that legal and moral notions such as personhood can map onto one another[234] *as concepts* (versus how they are applied to particular legal persons) did not exist, then both of these concepts would collapse. As argued for in chapter one, this is because a) not being used as a mere means is the minimum condition for being a member of the moral community and b) this condition morally ought to be enforced by the law. The concept of personhood incorporates both of these ideas. The fact that legal personhood is sometimes used for pragmatic reasons such as making policy about whether two men or two women can get married, for example, is merely one instance of

[233] *Ibid.*, pp. 101-102.
[234] See chapter one in this thesis.

deciding whether a given state of affairs (e.g. a policy prohibiting same-sex marriages) uses certain homosexual couples as a mere means (which can be couched in terms of their autonomy and so on). Some legal persons will be pragmatically affected by this policy while others such as certain heterosexual couples or single individuals will not be directly affected. Obviously, this does not mean that heterosexual couples and single individuals should not be legal persons because a certain pragmatic or policy-oriented consideration does not apply to them. When looking at particular examples of whether a certain treatment or policy uses a given individual as a mere means, one must consider the interests (and related capacities) of that individual. I have argued that the capacity for core self-awareness entails an interest in not being used as a mere means. When considering a particular core self-aware being such as a mouse, she will not have an interest in a policy about same sex marriage. She will, however, have an interest in not being injected with cancer cells. The interests of other core-self-aware-beings-as-legal-persons will differ. However, underlying all cases is a basic interest in not being used as a mere means. In this chapter, I have not merely presented arguments to the effect that the property mentality will not easily shift. Rather, as argued above, the person-property distinction is the only one that is capable of protecting the basic interests of those who have them. In other words, legal personhood is a necessary or foundational condition for having one's basic interests protected. Again, the notion that inanimate objects were legal property, humans of European ancestry were legal persons and human slaves were "quasi-persons" or "things-plus" is inherently unstable and was impossible to implement. This is not the result of, for example, the fact that human slaves of African ancestry may have benefited

less from a policy of only teaching European history than their legal owners would have. Rather, it is due to the fact that legal "quasi-persons" or "things-plus" are inevitably treated as a mere means and the prohibition of this constitutes the minimum condition that is held in common by all legal persons despite their many differences. Therefore, beings who morally ought not to be used as mere means morally ought to be legal persons. Otherwise, their moral status cannot be enforced, which – as argued in chapter one – would be immoral.

SECTION II: Moral Personhood

After an exhaustive survey of the philosophical literature that purports that non-human animals are not moral persons, Sztybel observes that this literature invariably stipulates various criteria for moral personhood or full moral standing, stipulates that these criteria add to the richness of human life and fails to provide any substantive arguments for accepting them. In other words, the philosophical discourse on the non-moral-personhood of other animals in the 28 years since this subject has been an issue, without exception, takes it as given that other animals are not moral persons and then proceeds from this undefended assumption.[235] As shown in chapter three, section three, sub-section B, Kant (and Aristotle) do not fare much better. Recall, however, that Kant defines moral persons as those beings whose nature indicates that they are ends in themselves who should never be used merely as a means.[236] As shown in chapter three,

[235] Sztybel, *Op. cit.*, pp. 7-18.
[236] Kant, *Op. cit.*, pp. 202-203.

Kant's claim that rational nature indicates this conclusion is highly questionable. This, however, is no reason to reject Kant's definition of moral personhood; moral persons are beings whose nature indicates that they should not be used merely as a means. Since, in chapters three through six, I have argued that all beings who possess core self-awareness should not be used merely as a means, it follows from Kant's definition that all beings who possess core self-awareness are moral persons.

Whether or not one accepts Kant's definition of moral personhood is ultimately extraneous. So long as the moral claim that all beings who possess core self-awareness should not be used merely as a means is abided by – and this is either done through the mechanism of legal personhood or a feminist ethic of care that becomes so accepted and widespread in the future that it provides the same or more protection than legal rights and personhood currently do – then it does not matter whether we label all beings who possess core self-awareness as moral persons. In this case, every being who possesses core self-awareness will be given exactly the same basic respect that is currently only given to humans who are presently labelled as moral persons.

One good reason to accept Kant's definition of moral personhood and to reject those found in the philosophical literature that go undefended is that Kant's definition accords very well with how the word "person" has been used throughout its history. Originally, moral and legal persons were defined as human rich white male landowners.[237] Only they had the legal rights that were afforded by legal personhood, which was in turn justified by a purported moral personhood status that only applied to

them. Slowly, the meaning of moral and legal personhood changed to include the poor, the landless, children, non-whites and women.[238] Underlying all of these views is the essential idea that a person is someone who is a member of the moral community. The Supreme Court of Canada stated that beings who have substantive legal rights are legal persons[239] and that ascribing legal personhood to a being is a moral endeavour[240] involving philosophical debate.[241] This suggests that, not only do Kant and the history of the word "person" suggest that persons are members of the moral community, but the Supreme Court of Canada does as well. As shown in chapter one – at the least – if a being is a member of the moral community, she or he must not be used as a mere means. Therefore, beings who possess core self-awareness who, as such, should not be used merely as a means are moral persons. The root linguistic meaning of the term "person" originally referred to the mask of an actor. Eventually, the term came to refer to the actor's persona and finally to the actor herself.[242] The nature of core self-awareness is such that it exemplifies the essential element of 'I' within a persona. Core self-aware beings are persons.

---

[237] Francione, *Introduction to Animal Rights*, *Op. cit.*, p. 168.
[238] *Ibid.*
[239] Canada, The Supreme Court of Canada, Ruling, *Tremblay v. Daigle*, 1989, pp. 548, 551.
[240] *Ibid.*, p. 553.
[241] *Ibid.*, p. 552.
[242] "Person, from Middle English, from Old French *persone,* from Latin *persona* actor's mask, character in a play, person, probably from Etruscan *phersu* mask, from Greek *prosOpa,* plural of *prosOpon*

BIBLIOGRAPHY

Aristotle. *Nicomachean Ethics*. Indianapolis: Bobbs-Merrill, 1962.

Arras, John D., and Steinbock, Bonnie, eds. *Ethical Issues in Modern Medicine*. 4th ed. Mountainview, California: Mayfield, 1995.

Bentham, Jeremy. *The Principles of Morals and Legislation*. Amherst, New York: Prometheus, 1988.

Bermúdez, José Luis. *The Paradox of Self-Consciousness*. Cambridge, Massachusetts: The MIT Press, 1998.

Boetzkes, Elisabeth, and Waluchow, Wilfrid J., eds. *Readings in Health Care Ethics*. Peterborough, Ontario: Broadview Press, 2000.

Cavalieri, Paola, and Singer, Peter, eds. *The Great Ape Project: Equality Beyond Humanity*. New York: St. Martin's Griffin, 1993.

Francione, Gary L. *Animals, Property and the Law*. Philadelphia: Temple University Press, 1995.

Francione, Gary L. *Rain Without Thunder: The Ideology of The Animal Rights Movement*. Philadelphia: Temple University Press, 1996.

Francione, Gary L. *Introduction to Animal Rights: Your Child or the Dog?* Philadelphia: Temple University Press, 2000.

Johnson, Oliver A., ed. *Ethics: Selections from Classical and Contemporary Writers*. Fort Worth: Harcourt Brace, 1994.

Singer, Peter. *Animal Liberation: A New Ethics For Our Treatment of Animals*. New York: Avon Books, 1975.

Sztybel, David. *Empathy and Rationality in Ethics*. Toronto: University of Toronto Press, 2000.

---

face, mask." (Merriam-Webster Dictionary, http://www.m-w.com/home.htm)