2004

Irrationality and human reasoning

Michael Aristidou
Louisiana State University and Agricultural and Mechanical College

IRRATIONALITY AND HUMAN REASONING

A Thesis

Submitted to the Graduate Faculty
of the Louisiana State University
in partial fulfillment of the
requirements of the degree
of Master of Arts

in

The Department of Philosophy

by

Michael Aristidou
BSc, Aristotle University of Thessaloniki, 1999
MSc, Louisiana State University, 2001
May, 2004

To my parents Theodotos and Kristia,
to my sister Eliza, and to my brother
Konstantinos for all their love and
support through all these years.

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Jon Cogburn for his great patience and valuable instruction on my thesis. I would also like to thank, Dr. Mary Sirridge for her advising and help through my graduate years in Philosophy, the members of my thesis committee Dr. Husain Sarkar and Dr. Ian Crystal for their cooperation, and Taneka Welch for her constant willingness to discuss issues related to my thesis, and her technical support.

TABLE OF CONTENTS

ABSTRACT

In his account of intentional interpretation, Donald Davidson assumes that people are mostly rational. Several psychological experiments though, reveal that human beings deviate drastically from the normative standards of rationality. Therefore, some psychologists arrive to the conclusion that humans are mostly irrational. In this thesis, I raise some objections to both points of view. On the one hand, ascribing rationality to humans in an a priori manner seems a suspicious position to adopt, considering the empirical data that show otherwise. On the other hand, the validity of the experiments and what exactly they test can also be put in question, since the position that humans are in general irrational is also unacceptable intuitively. In this thesis, I suggest that the discrepancy is due to the notion of rationality we adopt, which I bring into question. I do not find convincing reasons that humans should be thought a priori as rational and I do not also see why humans should be called irrational just because they fail certain tests. Many of the alleged "irrationalities" in the tests can be explained if we adopt different styles of reasoning than the "traditional" ones. Hence, humans can count as rational in another way. But, is this what Davidson thinks of rational, or does he think of rationality in the traditional sense? I think the type of rationality that Davidson endorses relies on Classic Logical conditions, which makes it inflexible. A type of rationality that relies on Fuzzy Logical conditions, as I claim, is more appropriate to describe human rationality.

> "The reasonable man adapts himself to the world; the unreasonable one persists in trying to adapt the world to himself. Therefore all progress depends on the unreasonable man".
> George Bernard Shaw (1856-1950), *Man and Superman*, 1903.

CHAPTER 1: INTRODUCTION

In their accounts of intentional interpretation, both Donald Davidson and Daniel Dennett assume that people are mostly rational. For Davidson, high degrees of rationality are needed because rationality relates with large numbers of true beliefs, which are necessary for interpretation. For Dennet, rationality also requires a large number of true beliefs, which are necessary for our survival and success as species. The former gives a conceptual argument for this assumption and the latter gives an evolutionary argument. Both have been criticized by other authors (e.g. Stich 1990 and 1985 respectively). Here, in Chapter 1, we will focus more on Davidson's theories and their implications.

In Chapter 2, I briefly offer Davidson's views on rationality and criticize his demand of ideal rationality. In Chapter 3, I present what can be thought of as empirical falsification of Davidson's theories; it is the claim that people deviate drastically from rational thinking since they robustly fail certain psychological experiments. I will criticize that view as well, and after introducing some of Stich's views in Chapter 4, I will present an alternative, in Chapter 5, that sits somewhere in the middle of the controversy.

As I will show, my view is close to Stich's epistemic pragmatism in the sense that it consists of a concrete realization of his theory. Fuzzy rationality will be an example of a cognitive system that is "better" than others in the line of Stich, yet not the "best." In this manner, it opposes Davidson's theory and is able to explain many of the so-called “irrationalities” that appear in the experiments, without being committed to the view that humans are largely irrational.

CHAPTER 2: DEFENCE AND CRITIQUE OF RATIONALITY

A. Principle of Charity

Davidson's famous Principle of Charity (PC)[1] roughly states that **one should interpret an agent's utterances in such a way that most of her assertions turn out to be true and most of her inferences turn out to be rational** (Davidson 1984, pp.183-198; Davidson 1982a, pp.302-303). In other words, when we are interpreting someone and we find that her reasoning deviates too much from what Davidson holds as "rational" (we will see what this comes to), then it is more probable according to Davidson that we are interpreting the subject incorrectly, and less probable that the subject is irrational (Davidson 1982a, p.303). But let us see in more detail what brought Davidson to that conclusion?

Inference that is continuously (or frequently) irrational, according to Davidson, is conceptually impossible. That is because inference - a process that generates beliefs - must have high levels of rationality and truth. But, why is that? And how could one

----

1. H. Jackman says, that "the Principle of Charity was actually first formulated by N.L. Wilson as the following semantic rule used to determine the referents of the names in a speaker's language" (Jackman 2003, p.145-146):

   > "We select as designatum [of a name] that individual which will make the largest possible number of [the speaker's] statements true" (Wilson 1959, p.532)

   As Jackman says, "this may initially seem like a mere description of how we go about guessing what the independently determined referents of a speaker's words are" (Jackman 2003, p.145). But as Jackman claims, "Wilson clearly intends Charity to be part of a more general account of what determines the referents of the speaker's words. As Wilson puts it, the Principle is part of an answer to the question "how do words hook up to things?" (Jackman 2003, p.145). Although Wilson's Principle of Charity has the objective of maximizing the number of true beliefs of an agent (like Davidson's PC), nevertheless, it has been distorted, especially by Quine's reformulation and later Davidson's, and "rather than being part", as Jackman says, "of a philosophical account of what determines the semantic values of our terms, Charity becomes more of a "common sense heuristic" "maxim" which guides the interpreter generally" (Jackman 2003, pp.151-152).

link rationality with beliefs? The keyword here I think is **interpretation,** and Davidson's main idea, which revolves around interpretation (Davidson 1982a, pp.302-303; Davidson 1982b, p.327, Davidson 1984, pp.195-198, p.170) goes as follows:

(1) For an agent to count as rational she must have beliefs.
(2) Having a belief requires having the concept of belief.
(3) An agent that has such concepts requires being a language user (i.e., an interpreter).
(4) For interpreting others, the assumption that an agent has most of her beliefs true is required.
(5) Interpretation and rationality are interrelated.

What (1) claims is that rationality is associated with propositional thoughts, and beliefs are fundamental to thoughts in general. According to Davidson:

> … belief is central to all kinds of thought. (Davidson 1984, p.156)

For (2) I would say that having the concept of belief derives from the fact that we can understand our beliefs as being true or false; that is, we understand that we might be right or wrong on a certain issue. But being able to evaluate our beliefs as true or false means that we must have the concept of belief. As Davidson says:

> Can a creature have a belief if it does not have the concept of belief? It seems to me it cannot, and for this reason. Someone cannot have a belief unless it understands the possibility of being mistaken, and this requires grasping the contrast between truth and error-true belief and false belief. (Davidson 1984, p.170)

Note that (2) implies:

(2a) Beliefs do not occur one by one but as groups of beliefs.

That is because one cannot have a single true belief about something because necessarily she must have many more true beliefs about that something. For example, we cannot really hold the belief "elephants are big" if we do not hold as true as well many other beliefs about elephants, for instance, "elephants are mammals," "they have big ears," etc.

Note also that if one does not hold many other true beliefs about elephants, then as we said, we cannot ascribe the belief "elephants are big" to her, precisely because we will not be able to ascribe to her the concept of an "elephant."

Statement (3) says, that since we are capable of interpreting and understanding each other's beliefs linguistically, then it follows from (2a) that most of our beliefs are true. Therefore, only language users have concepts of belief and hence beliefs. We read from Davidson that:

> …a creature must be a member of a speech community if it is to have the concept of belief. (Davidson 1984, p.170)

With respect to (4), as William Taschek claims:

> …we damage the intelligibility of our interpretations of the utterances of others if our method of interpretation has us *usually* and *inexplicably* disagreeing with them. (Taschek 1988, p.8)

Finally, statement (5) indicates that good interpretation and rationality are very close. We find, intuitively, an agent's utterance irrational if we fail to have an interpretation of that utterance. Note though, that to be able to ascribe irrationality (or rationality) it is necessary to be able to communicate (i.e., interpreting each other). As Davidson says:

> The conclusion of these considerations is that rationality is a social trait. Only communicators have it. (Davidson 1982b, p.327)

What one could conclude from statements (1)-(5) conjointly is that the meaning of a word cannot be fixed, if what the agent utters by that word does not accord with what she means by that word, and what she believes about that word. For example, suppose that an agent X believes that there is an elephant in front of him and she wants to communicate that. Suppose that X says "gavagai" whenever she sees an elephant. Then "gavagai" means elephant. Hence, fixing the beliefs and the utterances one can fix the meaning. The idea works backwards too. Fixing the meaning and the utterances (or meaning and beliefs) one can fix the beliefs (or the utterances respectively). It is clear now why Davidson needs the PC. Davidson uses the PC because he wants a theory of meaning. So, Davidson must use the PC to project rational beliefs on an agent X so that the interpreter can go from the utterances of X to the meaning of X's words. Now, the PC is plausible because radical error is impossible.

Finally, one also observes that what Davidson means by rationality is not clear at all. But Davidson's conclusion is that inference must be a rational process. It sounds as if, if inference is not rational, then there is no inference at all. So, if people reason, then they must be reasoning rationally. That is basically because people are capable of interpreting each other, which requires a large amount of shared beliefs held true.

A rough first guess of what makes an agent rational would be that an agent counts as rational if she engages herself in a process (inferences) of generating beliefs that maximize agreement with the beliefs that the interpreter holds, with high degree of confidence, as true. The agent has to come somehow to include most of the interpreter's almost "certain" truths to count as rational. But then, she must be rational, according to the PC, since she communicates with the interpreter. So in a sense, rationality depends on which beliefs the interpreter holds, with a high degree of confidence, as true. But, based on the fact that we all share a large number of beliefs, according to Davidson's theory, then we necessarily will come to share some of the ones that the interpreter holds as true with a high degree of confidence.

The argument would be perfectly fine if, to even be able to start interpreting each other, the set of almost "certain" beliefs for the interpreter was well defined. But which beliefs are those that the interpreter holds as true with a high confidence, and why should people assign the same degrees of confidence to certain beliefs? This debate is definitely controversial. Davidson assumes that certain beliefs, or propositions (e.g. the law of non – contradiction), should count as true with a high degree of confidence, which consequently defines the "type" of rationality he endorses. But, as we will see later (p. 18), this concern will have serious implications for his theory.

Let us now consider some obvious objections to the idea of PC. In the next section, I will focus more on the objection to the notion of rationality that Davidson uses in his PC. The controversy regarding rationality will be the main subject of the thesis. To premises (1)-(5) and to the whole idea of PC, one could raise the following suspicions:

(a) It is true, I think, that to have one true belief, one must have many more true beliefs on the same issue. But then, we have the following paradox to resolve. If for X's belief, say P, for that P to be true requires some other beliefs, say $Q_1, Q_2, \ldots, Q_n$ to be true, then we can never say P is true. Why? Because if each of those true Q's requires again with the same reasoning another set of true beliefs $R_1, R_2, \ldots, R_k$ (say, those that are to justify Q_2), and each of those R's requires its own set of true beliefs, etc, then, one reaches the conclusion that for X to justify P she will need to have infinitely many other true beliefs. But that is obviously impossible. The strange thing, though, is that Davidson himself seems to be aware of this problem. As he says:

> There are good reasons for not insisting on any particular list of beliefs that are needed if a creature is to wonder whether the gun is loaded. Nevertheless, it is necessary that there be endless interlocked beliefs. (Davidson 1984, p.157)

Clearly, Davidson has a holistic idea for beliefs in the sense that the relevant beliefs are too many and similar. But although one could have an intuition of similarity among beliefs, it is not clear at all with his choice of the word “endless,” whether he literally means infinitely many or just too many. But, how many? And even if it is a choice of finitely many among infinite, what is the qualitative choice of some beliefs among others? Taschek also talks about “relevantly similar, open-ended, interlocking collection of beliefs…”(Taschek 1988, p.11) and in an attempt to clarify the issue above, he states that:

> Lots of these beliefs are trivial and would hardly bear mentioning, and yet if most of them were not held, if say instead we supposed their contraries to be held, then the plausibility of counting the original thought a thought about guns would quickly evaporate. (Taschek 1988, p.11)

How do we classify some beliefs as trivial? And even if we can successfully classify them, the dilemma would still remain. The point is not whether those trivial beliefs are mentioned but whether they are being held, since, as Taschek says, they are necessary if we are going to grant to an agent beliefs about guns after all.

(b) Davidson claims that inference is a belief generating process. Agreeing that one needs more than one true belief on an issue in order for us to grant someone a true belief on that issue, why should rationality be a true-belief generating process? Davidson seems to have this true-belief generating process as an underlying assumption for rationality. But, an interpreter that has the potential to distinguish clearly what is true from what is true for him in the totality of her beliefs is a little rare to find. One could conclude, then, that interpretation (and hence rationality) is subjective, but a subjective rationality is not what Davidson wants. Davidson tries to “fit” objectivity into the picture, as an effort to avoid subjectivity, by claiming that:

> To have the concept of belief is therefore to have

> the concept of objective truth. If I believe there is a coin in my pocket, I may be right or wrong; I am right only if there is a coin in my pocket. (Davidson 1982b, p.326)

But the problem is that what is needed, for people to arrive at the subjective objective-contrast, is linguistic communication. Communication in turn, depends on "the concept of a shared world, an intersubjective world. But the concept of an intersubjective world is the concept of an objective world, a world about each communicant can have beliefs" (Davidson 1982b, p.327). It seems to me that Davidson is taking objectivity to be what we agree to be the case based on what surrounds us, which then makes Davidson's objectivity seem like mere intersubjectivity. As he claims:

> …what gives each [of the two creatures communicating] the concept of the way things are objectively is the base line formed between the creatures by language. The fact that they share a concept of truth alone makes sense of the claim that they have beliefs…(Davidson 1982b, p.327)

Even worse, the addition of the new parameter "objectivity" makes his argument circular. That is because for two people to have the concept of how things are they must be able to communicate (i.e. must interpret each other). But to communicate, they must share many common true beliefs (consequently, beliefs), which presupposes that they should have the concept of belief. But one has the concept of belief, only if one has the concept of objective truth [2] . Hence, we are driven to the circular conclusion that people have the concept of objective truth, if and only if they have the concept of objective truth. In the next Chapter, I will present examples by psychologists where humans allegedly reasoned irrationally, so according to the PC it is the psychologists that are most probably

----

2. According to the second quote in p. 9

incorrect and not the subjects. But before we examine who is right and who is wrong (Chapter 5), let us examine more closely what Davidson means by "rational" and, consequently, what is "irrational."

B. Davidson's Rationality

As mentioned before, it is not clear what Davidson means by rationality. Davidson uses the term rationality to define irrationality, although he never explicitly defines the former. He says:

> Irrationality is a mental process or state-a rational process or state-gone wrong. (Davidson 1982a, p.289)

Irrationality is present for an agent when the agent somehow generates beliefs, attitudes, and actions that do not cohere, or are not consistent, with the pattern of beliefs, attitudes, and actions of the same agent. Examples are: A person X acting contrary to her own "best" judgment, X holding a belief discredited by evidence etc. Irrationality is *not* present if the agent took in full consideration her desires, ambitions and acted based on her knowledge and values. For example, cases like X climbing Mount Everest with no oxygen, X believing in astrology or X trying to square the circle, do not count as irrational for Davidson (Davidson 1982a, p.290).

I am afraid that the two sets of examples just mentioned share common ground, that is, their boundaries are not very clear, since for example anyone who we take to act against her own interest might not know that actually she is acting against her interest (alike the one that might not know that the circle cannot be squared). One would expect Davidson to say, that if no one showed to X (or convinced X) the truth of certain propositions, then X cannot be accounted irrational. That is, if X does not have in her possession the true belief that the circle cannot be squared, then, she does not count as irrational if she is trying to square it. But this forces Davidson to a relativism (that is, relative to what X came to be aware of), which as we will see right away, is incompatible with a kind of ideal rationality account which he endorses. Roughly, Davidson thinks that everyone should hold certain beliefs, propositions, and logical properties, with high degree of confidence as true. The criticism on Davidson focuses on his denial to drop the idea of ideal rationality, which I think is unachievable. A certain relative position on the other hand I think it can be more feasible.

Davidson gives the example of the man in the park, who stumbles on a branch that lies in his path and then picks it up and throws it aside so other people will not get hurt (Davidson 1982a, p.292). But then, in his way back home (in the bus), he thinks that where the branch is now (aside) poses somehow more threat to people than before and he decides to get off the bus, although he wanted to go home, and goes back to the park and restores the branch to its initial position. Then Davidson claims, that the man in the park was irrational for returning, because, assuming he had the principle that one **ought to do what one holds to be the best**, he went contrary to that principle (Davidson 1982a, p.297).

Davidson does not explain though what one should conclude about agents who do not necessarily adopt the above principle. When are they irrational and when are they not? And based on the second set of examples I mentioned in the beginning (i.e., it is irrational when X acts contrary to his best judgment), one is justified in assuming that Davidson takes the principle above as a condition for rationality. But then, several problems arise since "ought to do what one holds to be the best," sounds more like "ought to tell the truth." But as something might be a lie but give pleasure similarly something might not be best but give pleasure. Davidson is making the assumption, I would guess, that what is best is also what gives the most pleasure, but this is not the case in general. For a utilitarian who has it: best = most pleasure, there is no problem at all in explaining the man's action (restoring the branch) in the park as rational. X wanted to go home, but it gave him more pleasure to restore the branch in its initial place. Hence his act was rational, and is not as if on his first act (throwing the branch aside) X reflected and said that he will throw the branch away based on his principle, but rather because he stumbled on the branch and that did not "please" him, so he threw the branch away.

If Davidson does not assume a utilitarian point of view, then at least he identifies what is "best" with what is "most rational." That begs the question though. Hence, if it is not what is most rational to X (that is relativism), and not what is most rational to the interpreter (that is subjectivism), cases which I think Davidson is trying to avoid, then what is left is an **ideal** ("objective") standard of rationality that counts as what is "best" for Davidson. Now, the problem with that is that Davidson thinks that an agent, "miraculously" somehow, knows what is best or when his best is the most rational. Furthermore, Davidson also takes the principle to be somehow a priori "good" by nature (which is not necessary), and the agents adopting the principle to have some special access to the "good." But take for example, a person X who adopts the principle "ought to do what one holds to be the best" and she comes to hold that: best = torturing a child. Suppose then that X for some reason comes and overrides her principle. Does this make her irrational? Davidson has to say "yes," to be consistent with the principle he suggests, but this is absurd. But even if he says "yes," on the grounds that her "best" was irrational in the first place, then he begs the question as we mentioned above because he assumes what is best is what is most rational. If Davidson says "no," then Davidson has to explain in which cases is acceptable to override the principle. Anyhow, this latter position will force Davidson to present other 2nd order criteria to decide when is acceptable for the principle to be violated and when is it not, something which leads to an infinite regress.

Davidson himself raises the question of how scientific a science of the mental can be (Davidson 1982, p.301). Narrowing it down to the question of rationality, and consequently irrationality, I claim that it should be extremely hard to formulate a scientific theory that describes it. Davidson says the same thing, when he claims that:

> …starting out from scratch to construct a theory that would unify and explain what we observe - a theory of the man's thoughts and emotions and language – we should be overwhelmed by the difficulty. (Davidson 1982a, p.302)

Let us see what a few reasons might be. Recall what theory means. It derives from the Greek "theoria," which means contemplation, which is to say that one has to see things from "above" or "outside" if you like[3]. But to see things from outside and do science, one has to disengage from the object of study if one wants to have objective results. But on the other hand, according to Davidson, we have to suppose that a person X is more or less like us in order to understand that person (Davidson 1982a, p.302). That is, we mirror a lot of what we hold as rational to the subject's mind and by this we affect in some degree the way X responds or is forced to reason, etc. For a scientist to study objectively rationality, she should find a way to disengage herself from her rationality or her scientific community's rationality, which is implausible. What is just stated, at least,

----

3. I got the idea to argue in this fashion from C. Taylor's paper "Rationality" (1982), although I disagree with his claims that theoretical understanding relates (in the sense that implies) with rationality (p. 90).

admits a subjectivism on the interpretation of others, which as one might suspect leads to many and different kinds of rationality theories. I do not know if that is what Davidson wants, though, since his approach to rationality through interpretation, demands high degrees of consistency. Hence, any approach that allows inconsistency is not acceptable to Davidson. That is clear when he says that:

> …inconsistency breeds unintelligibility. (Davidson 1982a, p.303)

The other alternative is to disengage completely from the subject matter, that is study rationality without "looking" let us say into people's heads which is absurd, because then we completely neglect all the empirical data (that is, of how people reason ordinarily) and end up with an intellectualist account of what is rational and what is not. Since one is talking of rationality of the people then how rational people are, one could claim that it should be a matter of empirical investigation and not some kind of a priori legislation. But as we will see in the next Chapter, empirical results regarding rationality have their problems too.

An intellectualist account for rationality might be a serious problem according to Stich, because many times philosophy tried to issue a priori legislation to science and philosophy was wrong. Examples might include, Kant's assumptions that space is Euclidean and the laws of physics are Newtonian (Stich 1990, p. 11), Kant's absolute views on Logic treating it as the complete science, or Quine's claims that "fair translation preserves logical laws" (Cherniak 1981, p.174). As Kant states:

> There are but few sciences that can come into a permanent state, which admits no further alteration. To these belong Logic and Metaphysics. Aristotle has omitted no essential point of the understanding….Indeed we do not require any new discoveries in Logic, since it contains merely the form of thought. (Haack 1996, p.27)

A more detailed account of Stich's views and critique on Davidson is included in Chapter 4. But the point just made above, relates with the ideal type of rationality conditions that Davidson assumes. Like Kant, who thought that the laws of Logic describe how we think, Davidson in certain occasions thinks that the "laws" of rationality should be Logical.

Davidson commits his rationality account on purely logical conditions, for example, consistency. As he says:

> If we are intelligently to attribute attitudes and beliefs, or usefully to describe motions as behavior, then we are committed to finding in the pattern of behavior, belief, and desire a large degree of rationality and consistency. (Davidson 1980, p.237)

As Cherniak claims though, Davidson does not just require large degrees of consistency but actually ideal consistency conditions (Cherniak 1981, p.174). That is because, as Davidson states:

> I do not think we can clearly say what should convince us that a man at a given time (or without any change of mind) preferred a to b, b to c, and c to a. (Cherniak 1981, p174)

That is clearly a requirement from Davidson that certain logical properties (e.g., transitivity) should be included in the account of rationality. This is equivalent to say that Davidson moves on to indicate which beliefs (or propositions) according to his view should be held as true with a high degree of confidence. Davidson identifies some beliefs (or propositions) that supposedly, we should all share. Clearly, statements like $\sim(A \wedge \sim A)$ (the law of non contradiction) count always as true for Davidson. The next chapters will be devoted to undermine the need to include in human rationality strict logical conditions like, for example, modus ponens etc. That will serve as an argument to undermine Davidson's requirement for ideal types of rationality. My goal is not to undermine all deductive ability of people in general. The claim will just be that people's deductive abilities do not always reduce to logical deductive rules. Furthermore, I will not claim that people should not try to eliminate inconsistencies in general, but to seek for more flexible rational conditions that would allow inconsistencies on certain occasions, since people naturally do include them in their reasoning.

Close to the above position is Cherniak's Minimal Rationality idea, which is formulated by imposing some minimal conditions on rationality, inference, and consistency. As Cherniak says, "the rationality conditions below are only necessary conditions for having beliefs and desires" (Cherniak 1981, p.166). All other minimal conditions on rationality derive from the **minimal rationality condition** (Cherniak 1981, p.166), which states that:

(1) If X has beliefs, then he would attempt some, but not necessarily all, of those actions which would maximize her goals.

The above implies that X should have some deductive abilities. But how much? Cherniak claims that there is also a minimum requirement on deduction, too. The **minimal inference condition (on deductive ability)** (Cherniak 1981, p.167) states that:

(2) If X has beliefs, then he would make some, but not necessarily all, of the sound inferences from those beliefs that maximize her goals.

Without condition (2), as Cherniak claims, X would never be able to satisfy condition (1). That is, without (2) X would never be able to take actions that would be maximizing his goals given X's beliefs. As an example, Cherniak invite us to imagine an agent who has in her belief set the belief that "if it is rains the dam will break," and let us say that it is raining. Then the agent, according to Cherniak, "would never conclude that the dam would break, even if this would be obviously useful" (Cherniak1981, p.167). But although I would agree that an agent must be able to make at least some sound inferences if she going to be able to undertake some actions that would result to achieve her goals, the example still will not do for what Cherniak wants to claim. I cannot see why the agent would not be able to conclude that the dam will break. The fuzzy word here is “raining.” Do we know how much is it raining now? It could be raining and the agent did not infer that the dam would break, because it was raining not very heavily. That does not mean that the agent could not in principle infer that the dam would break. Conversely, say that the dam will break if it rains (suppose if it rains heavily). And say now it rained a few raindrops. Is it rational for the agent to conclude that the dam will break? Let us even assume that she concludes that, i.e. that she satisfied condition (2). Should the agent move on to satisfy condition (1) too? In other words, say that because of these few drops the agent infers that the dam will break, and then she takes some necessary actions too, say she takes her family and moves to another state. Is that rational? The problem is not the minimal inference condition requirements, but the inference “type”, or style if you like, that we are trying to include to the minimal set. Strict logical deductive abilities (of the form: if p the q, and say p, hence q) will not always do, especially in everyday inferences. We need to relax the type of inference (e.g., if p then q, and say p, hence q under some circumstances), if those deductive abilities will have any natural value at all. But ignoring the example, Cherniak positions are quite realistic. He realizes the fact that the minimal inference condition is “vague” and admits that vagueness has advantages (Cherniak 1981, pp.175-176). Accuracy is not always necessary. He leaves open the question though, of whether inferences like modus ponens should be included in the minimal conditions. As he says:

> We shall treat as an open question here whether there are particular inferences-the most “obvious” ones, like modus ponens-which any creature that qualifies as having beliefs must be able to perform. (Cherniak 1981, p.177)

But in conclusion, Cherniak at least realizes the fact that modus ponens, for instance, is not a kind of inference that in the strict logical sense can be always applied. It depends on the circumstances. Only in some cases (crucial for an agent in achieving his goals) if p implies q, and say p, then the agent must imply q. Cases where first of all, the inferences required to be made are feasible to the agent (e.g., ones that require real life time). Also, inferences that are “positively useful for [the agent] at a given time” (Cherniak 1981, p.179), and easy for him to perform. Many inferences are inferences that might be sound but not reasonable (not useful) for an agent to make at a given time [4] . As Cherniak says:

> Not making the vast majority of sound and feasible inferences is not irrational, it is rational. (Cherniak 1981, p.180)

Finally, let us just mention that, according to Cherniak, the deductive ability requirement for minimal inference and rationality must contain a **minimal consistency condition** (Cherniak 1981, p.172) which states that:

(3) If X has beliefs, then if some inconsistencies arose, then the agent would eliminate them.

One cannot demand from an agent to eliminate all inconsistencies or adopt all inconsistencies. The former case is unrealistic since people do have inconsistent beliefs and violate consistent conditions all the time. The second is of no value since we lose every predictive ability for people’s behavior.

----

4. We will see later that in certain tasks that have a more realistic scenario, people perform better.

Before passing to the next Chapter, which will present the other side of the coin (i.e. that people frequently reason irrationally), I will say as a final remark that I would not have a problem agreeing with Davidson's PC as an idea if the conditions on rationality had been a little more relaxed. The type of conditions (e.g. consistency) on rationality that he demands will not do. For Davidson's theory to be more realistic, Davidson has to moderate the conditions on rationality that he requires. The type, or style, of logic matters in my view. Insisting on a particular type, for example the Classical Logic type, will not work. Not only because, as we will see in the next Chapter, there are many experiments that can falsify Davidson's ideas, but also because adopting Classical Logic to account for rationality, Davidson is forced into the following dilemma: Either inconsistency means absolute irrationality (because the Logic is Classical) which then would imply that none of us are believers by Davidson's theory, or inconsistency does not mean absolute irrationality which then would imply that Davidson needs a non-classical Logic (e.g. a Paraconsistent Logic). If we stop insisting on strict type of Logics, then Davidson's theory might be more plausible. We have to look for something other than the ideal type of rationality with strict logical laws, if we are going to come closer to how humans reason.

CHAPTER 3: DEFENCE AND CRITIQUE OF IRRATIONALITY

A. Everyday Reasoning and Experimental Results

To get a better grip of how humans reason, psychologists have engaged their subjects in a series of experiments, the results of which demonstrate a high degree of deviation in "rationality" from what some would take it to be the normative standards of rationality. Some of these experiments are: (a) The Selection Task (Wason P.C. and Johnson-Laird P., 1972), (b) The Conjunction Fallacy (Tversky A. and Kahnemann D., 1983), (c) Pseudo-diagnosticity (Doherty M.E., Mynatt C.R. and Tweney R.D., 1977) and (d) Belief Perseverance (Nisbett R.C. and Ross L., 1980). I will be interested more in experiments (a) and (b). Experiment (a), falls into what I call <u>Logical Approach</u> of arguing that humans are occasionally irrational, where experiment (b) falls into what I call <u>Probabilistic/</u> <u>Statistical Approach</u>. I will argue that although experiments (a) and (b) are quite useful in giving an insight into human cognitive capacities, **none** is sufficient to establish irrationality on behalf of humans. Nevertheless, they are good enough to falsify Davidson's belief that humans are a priori rational.

Humans might seem "irrational" with respect to Logic or Probability systems of inference (and that is the best these experiments can achieve), but we have no reason to believe that those two systems are the only ones that can give an account of human reasoning. There can be alternative systems that come closer to how humans reason in their everyday lives and explain many of the allegedly irrationalities that resulted from the experiments. The system in mind, which I introduce in Chapter 5 in a relative detail, is Fuzzy Logic and the claim, in the lines of Stich, is that it is a "pragmatically superior alternative" to the ones mentioned before. In other words, it is a "better" system of inference since once adopted; it is more general, more explanatory and leaves room for improving human cognitive skills.

B. Logical Approach and Problems

First Order Logic (FOL) was characterized as the calculus of certainty (Oaksford and Chater 1998, p.13), in the sense that the concepts that the calculus refers to are well defined, that no additional information can alter its inferences, and that the results we obtain can be asserted as 100% true or false. Therefore, because of the uncertainty that everyday human reasoning is embedded in, FOL is not a good candidate to account for human rationality. There are many examples-experiments to illustrate this (e.g. Wason's Cards, that we will explain later), and one does not have to go far to realize the gap between FOL and everyday reasoning. FOL is incompatible with human reasoning in its very foundations. The conditional "if…then…" is perhaps the cornerstone of FOL. It is also of most significance in everyday reasoning. But the very same conditional seems to lose its properties (Oaksford and Chater 1998, p.9) when carried from FOL to everyday reasoning. Thus, either we are talking about two different conditionals, say, logical conditional and natural conditional, or it is the same conditional but is not good enough to account for everyday logic.

Consider the following example:

If x is a *square*, then x has *four* sides
If x is a *square* and is *red*, then x has *four* sides.

One can observe, that adding to the premise (i.e., making the premise stronger) does not seem to alter the conclusion. In FOL, that is always fine. But if one tries to carry this in everyday reasoning, then the conclusion can be totally falsified or undecided. Take for instance:

If x is a *pigeon*, then x *flies*
If x is a *pigeon* and is *one second old*, then x *flies*.

or,

If x is a *pigeon*, then x *flies*
If x is a *pigeon* and is *three weeks old*, then x *flies*.

Obviously, the first conditional is generally true, in both examples, in the mind of everyday humans, but the second is false in the first case and undecided in the second case. Is the second conditional true in the first example, if one changes it, for example, to "pigeon is two seconds old"? Probably yes. But how about ten seconds old? Or a week, or three weeks old? How about three weeks plus or minus a day old? Is the conditional still true? There is a gray zone surrounding the word pigeon. But even if one knows for sure that a pigeon will fly in exactly three weeks, then, one could still make the second conditional false by changing it, let us say, to "pigeon is three weeks old and its wings are broken." That is, one can add one more premise and can falsify the conditional. So, if humans were supposed to use the conditional as used in FOL, then, they should have gone through, probably, an infinite number of relevant premises added to the conditional, in order to infer something. But, that is impossible. And even if the list of premises was somehow finite, then experience shows that this is not the way that humans reason. Humans do not require all relevant information to infer. Hence a fundamental property of the conditional does not always transfer[5] in everyday logic. The problem, I think, lies in

----

5. Mike Oaksford's and Nick Chater's similar example for non-transferability:

If x is a bird, then x flies
If x is a bird and is an ostrich, then x flies. (Oaksford and Chater1998, p.9)

although it contains the point I made about the vagueness on the concept of "bird," still it will not do, since the first conditional is false. x can be a penguin too.

the fact that our everyday concepts are vague and not well defined. It is not clear to us where the thin line of the word pigeon and not-pigeon is drawn. Is a pigeon with no wings a pigeon? How about with no wings and no legs? Or with no wings, no legs, and no head? And how essential is the property of flying in the definition of a pigeon? Could a car count as a car, if for example it could not run? On the other hand, a square is very well defined. It is the shape with all of its four sides equal and its four angles ninety degrees (supposing Euclidean Geometry). We cannot add premises of the form, "square with one of its angles ninety-one degrees", because we do not talk about a square any more.

Another non-transferable property of the conditional is the one that says that if the antecedent is false, then the conditional is always true. For example, the following is perfectly true in FOL:

If my fish has *200 IQ*, then the earth *revolves* around the sun.

As mentioned, this conditional is true in FOL but it does not make any sense at all in natural logic. It is not coherent.

All the above is just a small indication that human reasoning, at least with conditionals, is *uncertain*. That is, the results obtained are not 100% true or false. Oaksford and Chater explain this point very well when they say that:

> Logically this is an inference by modus ponens: *if the car is gone* (p)*, then someone has driven it away* (q) and *the car is gone* (p), therefore *someone has driven it away* (q). However, a crucial aspect of this inference is that it is uncertain: it is possible that no one drove the car away even though it is gone, someone may have towed it away, a helicopter may have removed it, […], and so on. Any of these additional pieces of information would *defeat* this inference […]. The problem with modeling such inferences logically is that classical logic is *monotonic*, i.e. no additional information, like that just outlined, can defeat a logical inference. Although this is a desirable property in mathematical reasoning it is almost antithetical to the kind of everyday

> reasoning that we use to explain our own and others' behavior. (Oaksford and Chater1998, p.4)

Human reasoning is a function of different factors and it varies under circumstances.

A famous example, that illustrates how much people deviate from formal mathematical-logical reasoning, is the well-known *Wason's Card* selection task (Johnson-Laird and Wason 1972). In this experiment, subjects have to pick those cards out of four that makes a certain rule true or false. The four cards on the side that the subjects can see have a "4," "7," "A," and "D" written on them. The rule is "If a card has a vowel on the one side then it has an even number on the other." On the question of which cards' other side one has to see in order to falsify the rule, around 90% of the subjects responded incorrectly "A" and "4" where the right answer is "A" and "7." This controversial experiment revealed some flaw in the deductive capacities of humans, but it surprised also researchers much more when a more concrete version of it (Robertson 1999, p. 69-70), dealing with everyday matters dramatically increased the correct responses that people gave. For example, subjects did far better when they were asked to play the role of a policeman that is trying to find whether the law of under aged drinking has been violated. The subject has to go in a bar, where at one of the tables four people are sitting. The first person is 15 years old, the second is 45 years old, the third is drinking lemonade and the fourth person is drinking beer. The rule is "If a person is under 18 years old, then she is not allowed to drink alcohol." Note that this task is nothing but the Wason's selection task stated above, where "lemonade" corresponds to "4," "beer" to "7," "A" to "15" and "45" to "D." On the question of what one has to do in order to check the rule, most of the subjects found the right answer which is check the 15-year old, and check the beer to see by who is being drunk. The reason why subjects found the concrete version of Wason's Task easier is, according to Roberston, that "certain contexts made the selection task much easier because the task fitted in with people's previously constructed schemas about those contexts"(Roberston 1999, p.71). People are more familiar with permission/obligation contexts, therefore it was easier for them find the correct answer, namely "15" and "beer," since the reasoning that generates those answers is closer to the permission/obligation schema.

The discussion above, at least shows that just because people failed the formal Wason task that does not make them irrational. Also, just because they passed the concrete version that does not make them rational either. The discussion above, also, reveals at least a discrepancy in the view that some philosophers had that reasoning is governed by the laws of logic (e.g. Boole's naming of his treatise on logic as the "Laws of Thought," Kant's view that logic contains the form of thought (Haack 1996, p.27) or even Davidson's requirements for some logical properties on rationality). This suggests that FOL is insufficient to capture everyday reasoning because the latter is imbedded in uncertainty. Informal reasoning seems to be more natural to humans. One might think that Probability Theory presents a good alternative to FOL, since the former deals with uncertainty, but, as one can see in the next section, Probability Theory falls short too.

C. Probabilistic/Statistical Approach and Problems

Probability Theory (PT) was called the calculus of uncertainty (Oaksford and Chater 1998, p.13), in the sense that when we cannot answer for sure that P will be true, we can at least say that will be true for example 80% of the times. Because of that, many philosophers thought that PT would be the appropriate modeling of human reasoning. PT, as a model of rational thought in uncertain situations, was used from its very beginning. First, it was thought as a tool in decision making especially in gambling by B. Pascal and D. Bernoulli (Ekenberg, p.4-6). J. Bernoulli, in his *Art of Conjecture*, interpreted probability as "rational degree of belief" (Oaksford 1998, p.14). But D. Bernoulli was one of the first that also came to realize the discrepancy between the objective method and a subjective conception of probability (Ekenberg, p.4-6). In the 20th century, the distinction between "objective probability" and "subjective probability" became finally clear. Subjective probability means that the outputs reflect the decision maker's actual beliefs and not the objective outputs. Probability now can very well be taken as degrees of belief. For this to happen though, PT should not be seen as reasoning about flipping coins or rolling dice but seen from a statistical point of view. Statistical inferences use observed data to infer a certain situation. For example, as Oaksford and Chater explain:

> Given the observation of 50 heads in 200 throws what is the likely bias of the coin? (Oaksford and Chater 1998, p.15)

Hence, viewing PT as statistical inference makes it a form of inductive inference, which is fundamental to human reasoning. This way of statistical reasoning is ultimately expressed by Bayes' Theorem, which states:

$$P(H_j/D) = \frac{P(D/H_j)P(H_j)}{\Sigma\, P(D/H_i)P(H_i)}$$

In words, the probability of a hypothesis H_j given the data D depends on the probability of the data D given each possible hypothesis Hi and the prior probability of each H_i. The probability $P(H_j)$ can be even interpreted as an initial degree of belief of the hypothesis H_j. According to Oaksford and Chater:

> Bayesian approach relates probability and statistics most directly to the problems of belief updating, and hence has the most natural relation to cognitive processing. (Oaksford and Chater 1998, p.16)

In other words, Oaksford and Chater adopt a Bayesian approach to account for human reasoning. It seems that this approach can explain many of the "irrationalities" generated by adopting a FOL account of human reasoning. For instance, this probabilistic approach can give a natural interpretation of conditionals in terms of conditional probability. E.g., the conditional we have seen earlier, "If x is a bird, then x flies," can be seen as, the conditional probability of "x flies given that x is a bird" is high. That is, P(Flies/Bird) is high, say 0.85 (Oaksford and Chater 1998, p.17). Also statements like "if x is a bird and is an ostrich, then x flies" can be viewed as a conditional probability but with different outcomes. Hence, P(Flies/Bird$\cap$Ostrich) is low, say 0.01. That is, both statements now are completely compatible within a probabilistic approach. They just take different values.

In their book, Oaksford and Chater also eliminate the Wason's Cards paradox by giving a "rational analysis of performance" (Oaksford and Chater 1998, Ch.10-13), according to which human performance far from being incorrect, in fact, displays an optimal adaptation to the environment. As they claim, the selection task is viewed by people as an inductive rather than a deductive task, in the sense that people must check the validity of a rule from specific instances (Oaksford and Chater 1998, p.27-28). As they argue:

> In particular, subjects face a problem of *optimal data selection*: They must decide which of four cards (*p, not-p, q, not-q*) are likely to provide the most useful data to inductively assess a conditional rule, *if p then q*. The standard "logical" solution is to select just the *p* and the *not-q* cards […]. This solution presupposes a "falsificationist" approach to inductive reasoning, which dictates that people should only collect data in order to

disconfirm, not to confirm, hypotheses. In contrast, […] rational analysis uses a *Bayesian*, rather than a falsificationist, approach to inductive confirmation […], and specifically to optimal selection […]. According to this approach, people assess whether to select a card by the *expected information gain* […] from turning that card. (Oaksford and Chater 1998, p.28)

Therefore, people were more than "rational" in choosing "4" over "7" since "4" has a higher expected information gain than "7" (where "A" corresponds to p, "D" to ~p, "4" to q, and "7" to ~q respectively), based on the calculations of Oaksford and Chater (Oaksford and Chater 1998, pp 177-185). Without getting into the mathematical details, I would just mention the basic formulas that enabled Oaksford and Chater to capture these intuitions probabilistically:

1. $I(M_i) = \sum P(M_i)\log_2(1/P(M_i))$ (Shannon-Wiener information)

where i = 1, 2 and M_i models 1 or 2 (e.g., M_1 = the model where "vowel" and "even - number" are depended, and M_2 = the model that "vowel" and "even - number" are independed). Also, $P(M_i)$ = the probability with which M_i is believed true. The Shannon-Wiener information is used to quantify uncertainty.

2. $P(M_i/D) = \dfrac{P(D/M_i)P(M_i)}{\sum P(D/M_j)P(M_j)}$ (Bayes' Theorem)

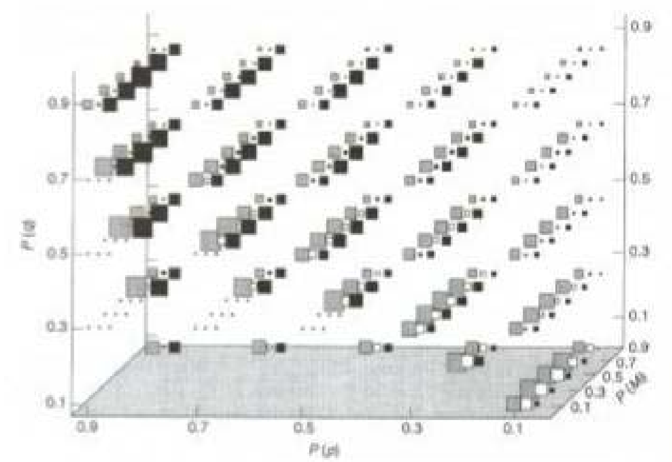
where $P(M_i/D)$ means the probability with which M_i is believed true given data D.

3. $I(M_i/D) = \sum P(M_i/D)\log_2(1/P(M_i/D))$ (information given data D)

4. $I_g = I(M_i) - I(M_i/D)$ (information gain)

5. $E_g = I(M_i) - \sum P(D_j)I(M_i/D_j)$ (expected information gain)

Oaksford and Chater summarize their results in the following graph (Oaksford and Chater 1998, p.184) which shows the pattern of expected informativeness for cards p, q, and ~q (the information gain from ~p is zero):



The gray, white, and black squares represent $EI_g(p)$, $EI_g(q)$, and $EI_g(\sim q)$ respectively, and the area of each square is proportional to the EI_g for the corresponding card. One can observe now that for P(p) and P(q) small (that is a consequence of Oaksford and Chater's "rarity assumption" in which people should treat p and q as rare), one gets that $EI_g(q) > EI_g(\sim q)$ which justifies people's choice of card "4" over "7."

At the end of their introduction, Oaksford and Chater conclude:

> So, rather than representing a blatant example of human irrationality, performance on this task can be viewed as an example of human rationality. Crucially our account reconciles the paradox between the apparent irrationality of human performance on the selection task and the manifest success of human reasoning in the everyday, uncertain world. (Oaksford and Chater 1998, p.20)

The points made above are very important since the points suggest that by changing the type of reasoning expected, one could explain the so called irrationality the subjects displayed in the experiment. Davidson's theory might be right after all, since the appeared irrationality was due to the fact that the experimenters thought that their subjects were reasoning in a certain way but the subjects reasoned in another way, i.e. the experimenters did not interpret their subjects correctly. Nevertheless, I am not sure whether Oaksford and Chater's theory could be the right account for human rationality. Although their account "rationalizes" human reasoning in the Wason Task, there are plenty of counter-examples to their probability-based model as I will show next.

It might be the case that calling the performance on Wason's Cards "irrational," sounds like a false term, since a probabilistic approach can answer it. So, Davidson's proposition that we have to ascribe rationality to humans might still be on the table. But, the probabilistic approach being effective on selection tasks does not necessarily imply that the same probabilistic techniques might work in other reasoning tasks, too. Proponents of the probabilistic approach think that Bayesian laws are at the heart of the human minds' reasoning (carried away by the success of Wason's Task), but, as A. Tversky and D. Kahneman (1982, p.361) will reveal, it is exactly the opposite. People tend to violate Bayes Law in many circumstances, as it will be shown. "Irrationalities" are also everywhere, even by adopting a probabilistic approach to explain reasoning.

For example, people seem to conform to the conjunction fallacy in probability judgments. The most famous example perhaps, is the one of Linda (Tversky and Kahnemann 1982) that states:

> Linda is 31 years old, single, outspoken and very bright. She majored in Philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. (Shier 2000, p. 69)

The subjects were asked to state, which is the most probable from the following five statements:

(1) Linda is a teacher in an elementary school.
(2) Linda is a bank teller.
(3) Linda works in a bookstore and takes yoga classes.
(4) Linda is active in the feminist movement.
(5) Linda is a bank teller and is active in the feminist movement.

The majority of the subjects chose (5) as more probable that (2), which shows that they violate a fundamental law of probability that says that the probability of a conjunction must be less or equal than the probability of its conjuncts. Therefore, that is an "irrationality" case, according to the experimenters. But is it? Or is it an "irrationality" case within the probabilistic approach? If yes, then one could say that the above "irrationalities" are irrationalities with respect to the approach one chooses to model human reasoning. It is not irrationality in the "traditional" sense. They seem to be "irrationalities," because of the strict rationality conditions the experimenters imposed on rationality.

Oaksford and Chater, commending Tversky and Kahnemann's results, argue that:

> If this viewpoint is correct, then the whole idea of rational models of cognition is misguided: cognition simply is not rational" (Oaksford and Chater 1998, p18).

I could not disagree more. If human reasoning does not conform in the attempts to be formalized with mathematical theories such as PT, with Russell/Frege logical assumptions (i.e. $\sim(A \wedge \sim A)$, Law of Non-Contradiction (LNC)), that does not make human reasoning irrational. It makes it "irrational" relative to the formal systems, but obviously that is something different. It depends what we mean by irrational and how we choose to formally model rationality. Do we model rationality based on a Russell/Frege approach, where LNC is always true? There is no reason why we should not accept A and ~A, simultaneously, as basis of our reasoning since evidently it seems we think in such a way. We do seem to have contradictory beliefs and ways of thinking. Therefore, any approach of modeling human reasoning in the traditional (Russell/Frege) mathematical approach is unpromising. We have to change the logical axioms in our axiomatization.

There is no reason why rationality should be tied in the LNC. We can still speak of rationality, even by accommodating $A \wedge \sim A$. It is by doing our formal systems more "fuzzy" that we can come closer in describing the fuzzy reasoning, and behavior, of human beings.

To sum up, I think that all the cases of "irrationality" revealed in the above discussion are not real irrationalities. They are irrationalities with respect [6] to the systems of Logic (FOL, PT) that the subjects were forced to respond. Both systems revolve around a crucial axiom, namely the LNC, which the everyday system of logic of humans does not appear to have it as central. Therefore, it does not make sense to judge humans as irrational, simply by not responding accordingly to something that is secondary (or alien) and restrictive to their way of thinking. In Chapter 5, one can see how a Fuzzy Logic approach can come closer to how humans reason. Before that though, let us examine first in the next Chapter some views by Stich, in order to see how Stich's theory anticipates in a sense the idea of a Fuzzy Logic system, as an alternative to FOL and PT.

----

6. The expression "with respect to the system" does not mean to imply a general relativism on what is irrational. It simply says that if we choose to model rationality on FOL then responses based on, for example, Paraconsistent Logics would look irrational. It is like religions. They all try to formalize what is a moral act, so necessarily a person of another religion looks immoral, at least in theory. But that neither precludes the possibility of existence of a morality beyond the morality of the current religions, nor does it say that morality is capricious or subjective, and nor that the common morality of all religions is the "right" morality.

CHAPTER 4: ALTERNATIVE THEORIES – STICH'S THEORY

Steven Stich argues (Stich1990, p.15, 17) that if Davidson's position is true then the following problems arise:

a) Our ability to empirically explore irrationality is undermined; hence we are losing a good insight into human cognition, since the interpreters are the ones most probably mistaken.
b) Any concern about bad people doing bad theories, which in consequence affect our lives, is also undermined since in principle humans cannot be irrational.
c) The view that cognitive processes of ordinary people can be improved is also undermined, since the claim that "departures from normative standards of reasoning are impossible" sets the bounds.
d) The claim that bad reasoning is conceptually (or biologically) impossible leads to a normative theory of rationality of no practical importance, since the theory turns its back on the empirical results.

Stich claims that, at best, Davidson's argument shows that irrational people cannot engage in "real inference" but only in "inference-like" processes, which generate "belief like" (that is, not intentionally characterizable) mental states (Stich 1990, p.12). But as Stich claims, as there is no way one could distinguish between "real beliefs" and "belief like" mental states, there is no clear distinction between "real inference" and "inference like" processes. Stich agrees with Davidson that content and good reasoning have a close connection, but he does not agree with Davidson's position that treats the two processes mentioned above, of "real inference" and "inference like" inference, as two distinct clear-cut cases. Therefore, Davidson's claim that content depends on inference is true, if and only if, one assumes that inference is "real inference", something that cannot be assumed because of the unclear distinction between "real inference" and "inference like" processes.

In other words, Davidson is right to say that content and "good" reasoning are linked, but wrong in the assumption that for "good" reasoning there is only one type of rational reasoning. The latter assumption is something that Stich opposes, since he claims that it is *not* the case that there are no alternative systems of reasoning (different) that are *all* rational. This is something that Stich calls **normative cognitive pluralism,** which is the claim that there are different "good" cognitive processes.

But how can one decides then that a system P is a "good" (i.e. a rational) cognitive system? Stich mentions N. Goodman's attempts to describe a procedure (or a test) that a system of inferential rules should go through (or pass), in order to count as rational. Goodman argued that:

> …is via a process of mutual adjustment in which judgments about particular inferences and judgments about inferential rules are brought into accord with one another. (Stich 1990, p.10)

But as Stich observes, it is very difficult to discern the relation between rationality and "right" test. Does not one assume a priori here that the test itself is rational? Or did that test pass a different test previously? Is there not a regress threatening? And even if a system passes the test, why should it count as rational? Does not one beg the question here? For these reasons therefore, Goodman's idea is implausible. But even if such a test existed, as Stich says, it must be an analysis (or explanation) of our ordinary concept of rationality. That is, it has to explain also irrationalities. Hence one needs a more realistic account of what makes a system rational. But a realistic definition of what counts as a rational inference process is implausible, because of the reasons I mentioned in Chapter 2(B) (e.g. reasons of what kind of logical properties should be preserved, reasons of subjectivity etc). Therefore, one is left with a comparative account of a system P being "more" rational than a system Q. By that, I imply the existence of a system P that contains Q, and can explain some irrationalities appeared within Q (e.g. Fuzzy Logic contains Binary Logic (see pp.40-41). By "contain," I mean that Fuzzy Logic does at least what Binary Logic does). This is similar to what Stich proposes; when he adopts a position that he calls **epistemic pragmatism** (EP), as an answer to Davidson's and Dennett's thesis. What Stich means by that is, in his own words, the following:

> One cognitive system P is preferable than a system Q if in using it we are more likely to achieve those things we value most. (Stich 1990, p.24)

Note that EP is a comparative account and it does not answer whether a system P is good or bad. It just says whether P could be better than Q.

Three objections that come quite naturally, as Stich indicates (Stich1990, p.25, 26), are the following:

(a) EP *leads to relativism*: It is clear by its definition that EP is relative to the values of an agent. But if a system P is better for an agent, and a system Q is better for a different agent, then which one is right? Or is it the case that anything goes?
(b) EP *leads to skepticism*: that is because EP separates completely rationality from truth. If different systems of reasoning are preferred by different people, then the systems generate different beliefs, on the basis of similar sensory input. How can one ever decide then what is true?
(c) EP *leads to circularity*: How does one know that her cognitive system is pragmatically preferable, without using the very system whose superiority she is trying to establish?

As far as (a), it is not the case that anything goes at all. As mentioned already in the previous section, one has no reason to always respond in "black or white" manner. If system P is right that does not mean system Q is necessarily wrong. It can be the case that both are right, and simply one might be used in the wrong context. Or it can be the case that P is "more" right than Q etc. Therefore, the dilemma that objection (a) is forcing on us is actually a "pseudo-dilemma." It can even be the case, as I show in the next section, that a system P could contain a system Q. Hence, even if one uses Q, it is feasible for her to adopt P as well (i.e., extend Q) and improve her "rationality."

Regarding (b), Stich responds by saying that even if one's chosen system of reasoning generally leads to false beliefs, then we should really first show that one has some reason to "want" true beliefs, something that has not been established yet (Stich 1990, p.26). I would add to that, that even if assumed that there is something valuable for one having true beliefs, then one can still answer the question that objection (b) is posing. That is by claiming that it can be very well the case that not two, but even one (and the same) system can generate different (or opposite) beliefs on the basis of similar sensory inputs where *both* of the beliefs might be true (see Appendix).

Finally, about objection (c) I would just say that whether a cognitive system is pragmatically preferable, or superior to some other system, is ultimately tested in praxis. If by trying P initially on a few occasions we see that we are able to achieve everything that Q allows us to achieve, plus more, then we will end up using P. And if by using P, we see that some irrationalities produced by Q can be now explained, plus we do not lose anything valuable we obtained from Q by using P, then P will prevail. Reasoning systems are not something fixed. They evolve relative to our knowledge, environment, circumstances, applicability, etc and they are adopted based on their pragmatic results, explanatory power etc. And as Stich insists, we have to keep searching of what makes "good" reasoning good and what makes "bad" reasoning bad. That means, one cannot afford to "rest her soul" on one reasoning system because there is always another one better.

Now, if one finds such a comparative account that EP proposes very relative, then Stich is making an attempt to say what makes a reasoning system "good." He suggests that:

> P can qualify as "good" if is as good as any
> other "possible" alternative. (Stich 1990, p.27)

He explains that by "possible" alternative he does not mean logically possible, but feasibly possible alternative. But which are the feasibly possible alternatives? Are there any in the first place? Stich says, that such feasible alternatives are not easy to be identified abstractly. In praxis though, as he says, such feasible reasoning systems are the ones that people could use and improve their cognitive performance (Stich 1990, pp.27-28). In addition, those systems are not to be found without empirical exploration. It seems then, that the only real problem we are facing is the problem of *existence* of even at least one such a reasoning system, having Stich's properties. The Chapter that follows has the objective to give one example of such a system. Fuzzy reasoning systems, I claim, will serve perfectly Stich's theory as I will show.

To conclude this section, let us just give Stich's final answer (and make a small remark) on the main question of the discussion of whether one is justified in claiming that the subjects in the experiments could be called irrational. As Stich argues:

> It is not clear that the subjects in the experiments were reasoning badly. To claim that they were we have to show that there exists a pragmatically superior alternative. (Stich 1990, p.28)

I could agree with the first half of his statement. What is not so clear is the second half of his words. I do not think that if one establishes a superior alternative, then one is always justified to ascribe irrationality to people. It makes more sense if one takes his second half of the statement to mean that people would be then called irrational *relative* to the inferior system, and not plainly irrational. I will read Stich's words as such then, and try to materialize his theory by the example of a Fuzzy Reasoning System that is presented in the following Chapter.

CHAPTER 5: A NEW IDEA

A. Paraconsistent and Fuzzy Logic

Paraconsistent Logic (PL) is not so much the rejection of Classical Logic [7] , but more a generalization of it. PL challenges the principle of "anything follows from contradictory premises." To be more precise, a Logic is called **explosive** if and only if its consequence relation satisfies (the *ex falso quodlibet*):

$$\{A, \sim A\} \models B, \forall\ A, B \quad (1)$$

A Logic is called **paraconsistent** if and only if it is <u>not</u> explosive (Priest 1984, p.3). A crucial consequence of rejecting (1), though, is that some PLs accept statements like $A\wedge\sim A$ (dialethias) as true, i.e. they violate the LNC (that is $\sim(A\wedge\sim A)$). Hence, for those PLs to be non-trivial, it has to be shown that, indeed, there are examples of dialetheias. The most famous one is the *Liar's Paradox* given below which, note, is both a true and false statement. A popular version of it is:

"This sentence is not true." (2)

Note that statement A (i.e. (2) above) and $\sim A$ are true which makes $A\wedge\sim A$ also true in standard 4-valued Logic.

One observes, though, that accepting true contradictions (dialethism) implies paraconsistency (because $A\wedge\sim A$ true $\models A, \sim A$ true), but not vice versa. For example, in non-adjunctive approaches to paraconsistency like the one of Jaskowski's (Priest 1984, p.7), one fails to obtain entailments of the form: $\{A, \sim A\}\models A\wedge\sim A$. As Priest says:

> …non-adjunctive approaches to paraconsistency do not

----

7. Although some authors like L. Zadeh (Zadeh 2002, p.1) and B. Kosko (Kosko 1993, pp.23-24) tend to identify the what is now known as Classical Logic with Aristotelian Logic, the reality is that although Aristotle held that contradictions cannot be true, his Logic was not explosive (Priest 1984, p.5). Aristotle was actually one of its first critics and he proposed modifications on it (Haack 1996, pp.xxv, 40).

> take the idea of dialethias seriously. For we have $\{A\wedge\sim A\}\models B$, and the only thing that prevents $\{A, \sim A\}$ from blowing up, is the non-standard behavior of conjunction. For this reason non-adjunctive paraconsistent logics are unsuitable as the underlying logic of important inconsistent theories such as naïve set theory. (Priest 1984, p.8)

Now, not only PLs exists as formal systems, but some are also the foundation (at least the ones that adopt dialethism) of many theories that are inconsistent but non-trivial[8].

Some examples of such theories are:

(a) Cantor's Set Theory: This is the theory that its first order language has two relation symbols, namely "=" and "$\in$," and its concept of a "set" is captured by the following two axioms:

(1) $\exists a \forall x\, [\, x\in a \leftrightarrow P(x)\,]$, where P is a property. (axiom of comprehension)
(2) $\forall a \forall b [\, \forall x\, (x\in a \leftrightarrow x\in b) \rightarrow a = b\,]$ (axiom of extensionality)

This is an inconsistent theory, since Russell's Paradox implies (by axiom (1)) the existence of sets with the property $P(x) =\{x\in x \text{ and } x\notin x\}$. In this theory though, we have some surprising results such as the proof of the Axiom of Choice (Priest 1984, p.14). This axiom, as one knows, is independent of the axioms of other set theories, like Zermelo-Fraenkel for example.

(b) Newton's Calculus: This is the theory of infinitesimals in which very small quantities ε (called infinitesimals) are considered in some cases as zero, and in other cases as non-zero (Priest 1984, p.14).

(c) Quantum Mechanics: This is the theory introduced by M. Plank in 1900 which

----

8. Inconsistent means that the theory contains both sentences A and ~A. Trivial means, for all sentences B, B is in the theory. According to Priest, inconsistency plus non-triviality is equivalent to Paraconsistency (see Priest 1984, p.3 for a proof).

roughly states that photons can behave both as particles, and as waves. Of great significance in this theory is the so-called Dirac δ function, which is defined as follows:

$$\delta(x) = 0,\ x \neq 0 \quad \text{with} \int \delta(x)dx = 1$$

As one can observe though, the above property of the Dirac Function is contradictory since the area of a point is always zero.

(d) Bohr's Theory of Atom: This is an example of a theory that was finally proved to be false. According to the theory, an electron that orbits the nucleus does not radiate energy but also does, since Maxwell equations (also part of the theory) say it should (Priest 2000, p.1). Nevertheless, Bohr's Theory was preserved since it verified experiments, made predictions, etc. Somehow, the contradictions were there but did not affect the rest of the theory. The theory was viewed from a more holistic perspective.

There is a variety of PLs where $\{A, \sim A\} \models B, \forall A, B$ fails. One of them (other than Jaskowski's) is a Many-Valued Logic proposed by F.G. Asenjo (Priest 2000, p.4). A Many-Valued Logic is a logic with more than two truth-values. For example, a 4-valued logic is one that asserts 4 truth-values. For example: True, false, true & false, neither true nor false. In other words, whereas Classical Logic has as Truth-value set the set {0, 1} (meaning {False, True}), 4-valued Logic has as Truth-value set the set {0, 1, 2, 3} (meaning {true, false, true & false, neither true nor false}). If one now accepts as Truth-value set the interval [0, 1], i.e. all real numbers between 0 and 1, one produces what is usually called *Fuzz*y *Logic* (FL). One notes that the 2-valued Logic (i.e. Classical Logic) is a special case of the FL since the discrete set {0,1} is a subset of [0,1], where 4-valued Logic is not. That is because in (standard) 4-valued Logic the statement $A \wedge \sim A$ is true, where in FL is "half-true", i.e. it has truth-value 0.5 (see Lukasiewicz's definition of degree of truth in p.52). FL implies that truth-values are "degreefied." One does not talk of whether a statement S is true or false, but of the degree that S is true. Let us make the notion of degree of truth above more precise, by relating also the concept of membership in a set (see Zadeh 1965, pp.339-340 for more details).

In classic set theory, a subset A of X indicates a function f_A, which is defined as follows:

$$f_A: X \to \{0,1\}$$
$$f_A(x) = 1 \text{ iff } x \in A \qquad (\text{ hence } f_A(x) = 0 \text{ iff } x \notin A, \text{ i.e. when } x \in {\sim}A)$$

Analogously, in Fuzzy Set theory, a fuzzy subset A of X indicates a function:

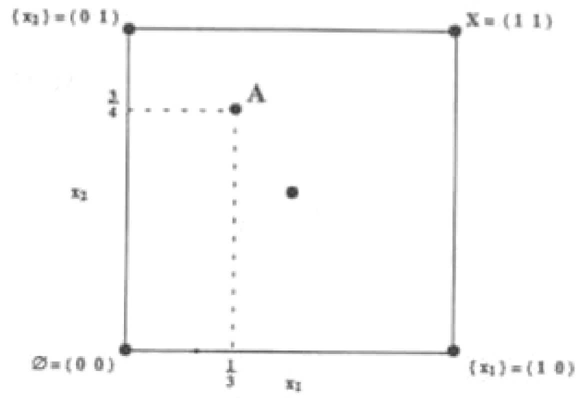
$$f_A{}^*: X \to [0,1]$$
$$f_A{}^*(x) = n \in [0,1]$$

The interpretation of the number "n" is as follows:

$f_A{}^*(x) = n$ iff (n-part of x) $\in$ A (hence, $f_A{}^*(x) = 0$ iff 0-part of x $\in$ A, i.e. when $x \notin A$ which is $x \in {\sim}A$ and $f_A{}^*(x) = 1$ iff 1-part of $x \in A$, i.e., when $x \in A$)

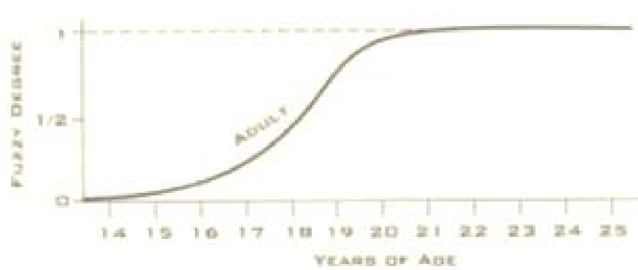
Therefore $f_A{}^*(x) = 0.5$ iff 0.5-part of x $\in$ A ie, $x \in A$ and $x \in {\sim}A$. Note that if A is not a fuzzy set, $f_A{}^*$ reduces to f_A. In other words, $f_A{}^*|_{A=\text{not fuzzy}} = f_A$.

There is also a geometric way to understand the notion of a fuzzy set, which is usually more intuitive. Rather as viewing a fuzzy set as a membership function, we view it as a point in an n-dimensional unit cube (Rubik's cube), denoted by $I^n = [0,1]$, where n is the number of elements in a set X. That n-dimensional cube is actually defined as the fuzzy power-set of all fuzzy subsets of X. So for example, if $X = \{x_1, x_2\}$, the classical power-set of X is $\{\varnothing, X, \{x_1\}, \{x_2\}\}$. But the fuzzy power-set of X is represented by the unit 2-cube (it is a square), and the fuzzy set A by the point A in the cube as given by the figure (Kosko 1990, p.217) below:

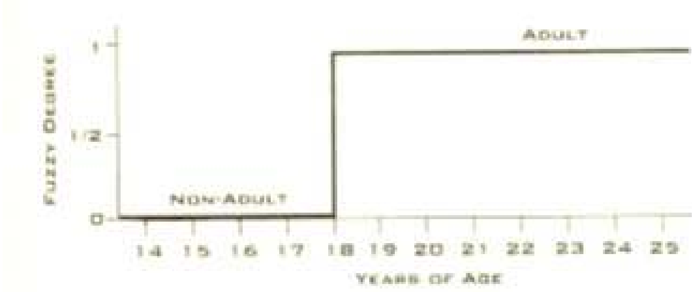


The numbers 1/3 and 3/4 are the degrees of membership in A, of x_1 and x_2 respectively. One also observes that at the corners of the cube, one has the classical power-set of X (the lattice B^n).

As an example of a fuzzy set from real life, consider the fuzzy set "Adult" (i.e. the set of all adult people). It is a fuzzy set because there is no way that we could draw a clear distinction between adult and non-adult. This is obvious if one sees the graph of $f_A{}^*: X \to [0,1]$ (Kosko 1993, p.35), where X = [0,100] and A = {adult people}:



Thinking A as <u>not</u> fuzzy results to the following graph for f (Kosko 1993, p.35), graph of which its interpretation seems absurd:



And I say absurd because in reality there is no way that a "rational" person would think of a person who is 18 and one-day years old, as always an adult. The only "reality" which an 18 and one day old would be thought always an adult, is the formal system of law, which humans violate every second. I am not claiming that every time people break a law, they were confronted with fuzzy situations. I am saying, though, that if a situation is fuzzy to people, there is higher likelihood that they will break a law that surrounds the situation. Are all the people that break the law that says you must be 18 to drive irrational? Obviously they are not. What is the problem then? The problem is that, the clear cut law of "18 or not 18" is not good enough to contain the fuzzy set "adulthood," hence unable to capture the fuzzy thinking about it. The formal law does not leave any room for cases like "17 but mature," or "19 but irresponsible." If a person x is 19 and irresponsible, but nevertheless drives, then who is more irrational; the lawmakers, or the people? Hence, reasoning purely with conditionals, say, "if x is 18 then x is an adult" is completely wrong. When the set we refer to is fuzzy, then it imposes too many constrains on us to consider - if we still want to preserve reasoning with conditionals.

B. Fuzzy Rationality

It is clear from the discussion so far that human reasoning violates the LNC quite often. The question now is, by doing so, whether human could be thought as irrational.

Could it be rational to accept contradictions? If there is enough evidence to accept a contradiction, I would say yes. Some people would argue that *contradictions are false and is absurd for rational people to believe them*. But, arguments like this are begging the question since contradictions are not always false. They are always false assuming the LNC. But, humans have no reason to accept LNC a priori since truth and falsity are not always disjoint for them. For example, it is not quite true to say that a 25-year old is an adult but is not quite true also to say that he is not. Now, why statements like the one italicized above are simply false, could be seen by the famous "Paradox of the Preface," which goes as follows:

> A person after a lot of research writes a book m which he claims that $A_1,\ldots,A_n$ are true. He has every rational reason to believe them but he also aware that no factual book has ever been written which did not contain some falsehoods. The inductive evidence for this is overwhelming. Hence quite rationally he believes $\sim A_1 v \ldots v \sim A_n$ too. Clearly his belief set is inconsistent. Yet he believes it and is paradigmatically rational. (Priest 1986, p.107).

Other proponents of Classical Logic would argue that since Fuzzy Logic accepts to some degree dialetheias, i.e. true contradictions, then a person could not be forced rationally to abandon a view held. For if a person accepts A, then when an argument for ~A is put up, they could just accept both A and ~A. But again, arguments like the one just stated are simply false since nowhere has been stated that if one accepts sometimes A and ~A, then one always does it. One should recall that both A and ~A are only accepted in the light of enough evidence. All the above now, dictates somehow how humans' fuzzy rationality is to be understood.

Firstly, it has to be clear that at the heart of human reasoning are perceptions and not formal syllogisms or computations. And by saying this, I mean that if, for example, a car

30m in front of an agent breaks, then the agent would also break, approximately at a point x (with respect to their velocity), in order to avoid collision. The agent will not actually compute the real distance and velocities and find where exactly to break; neither would the agent reason in a way such as, “Oh, breaking now I have higher probability not to hit the car in front.” Perceptions are imprecise. That is, one is not sure whether in a cloudy weather, whether it is going to rain or not, and thus asserts as much truth as falseness. In other words, perceiving fuzzyfies sets that human have to act upon, and hence, fuzzy reasoning is the type of reasoning humans probably employ. It is the kind of reasoning exactly appropriate since confronted with degreefied situations, e.g. “x is more of y and not quite z”, one degreefies her truth tables too. For example, “x is more of y” is something like 85% true.

Secondly, humans surrounded by uncertainty, imprecise perception, or even contradictory perception (see Appendix), automatically have to extend their field of reasoning to account also for contradictions. Consequently, they have to adjust their conditions of what is to be a rational belief. It has to be a more liberal account of rational belief that considers also contradictions, but not in a sense of everything is accepted. I mentioned earlier that P is rationally accepted as true (P could also be a contradiction), if there are good enough reasons to support it. One might ask of course, what qualifies as a good reason, but as G. Priest puts it there are many good reasons, some of which, for example, are: experimental support, high statistical probability, something deduced from something already rationally accepted and so on. As he further explains:

> I do not suggest that these are the only kinds of reasons…to support a theory/belief, but equally, I am skeptical of the attempts…to reduce them to a single “master reason.” (Priest 1986, p.108)

And then he goes on to claim that:

> …an inconsistency can be supported by each and every kind of reason enunciated above. (Priest 1986, p.108)

Therefore, contradictions as well, if there is enough evidence to support them, could very well be accepted as true. And contradictions can be supported by each of the reasons mentioned above.

To end this section, as a final remark, I would like to make clear a distinction between Probabilistic reasoning and Fuzzy reasoning. It was more than clear from our discussion that FOL reasoning, is a very special case of Fuzzy reasoning. Could one though separate Probabilistic reasoning and Fuzzy reasoning, although both, somehow degreefy the truth of a statement? The answer is yes. The two theories are not the same, but they relate in the sense that Fuzzy Theory contains Probability Theory (i.e. Fuzzy Theory can do at least the things Probability Theory does). Probability Theory, as a special case of Fuzzy theory, <u>cannot</u> answer at all to particular problems that Fuzzy Theory can. Kosko gives the following example:

> Say you park your car in a parking lot with 100 painted parking spaces. The probability approach assumes you park in one parking space and each space has some probability that you will park in it. All these parking-space probabilities add up to 100%. If the parking lot is full, there is zero probability that you will park in it. If there is only one empty parking space, say the thirty-fourth space, you will park there with 100% probability. If the parking lot is empty, and if we know nothing else about the parking lot, you have the same slim chance, 1%, of parking in any one of the parking spaces.
> The probability approach assumes parking in a space is a neat and bivalent affair. You park in the space or not, all or none, in or out. A walk through a real parking lot shows otherwise. Cars crowed into narrow spaces and at angles. One car hogs a space and a half and sets a precedent for the cars that follow. To apply the probability model we have to round off and say one car per space.
> Up close things are fuzzy. Borders are inexact and things coexist with nonthings. You may park your car 90% in the thirty-fourth space and 10% in

> the space to the right of it, the thirty-fifth space. Then the statement "I parked in the thirty-fourth parking space" is not all true and the statement "I did not park in the thirty-fourth space" is not all false. To a large degree you parked in the thirty-fourth space and to a lesser degree you did not. To some degree you parked in *all* the spaces. But most of those were zero degrees. This claim is fuzzy and yet more accurate. It better approximates the "fact" that you parked in the thirty-fourth parking space. (Kosko 1993, pp.12-13)

Consider also the following example ("possible-not probable" dilemma), given by Zadeh:

> Suppose that 99% of professors have a PhD degree, and that Robert is a professor. What is the probability that Robert has a PhD degree? PT's answer is: between 0 and 1. More generally, if A and B are events such that the intersection of A and B is a proper subset of B, and the Lebesque measure of the intersection is arbitrarily close to that of B, then all that can be said about the conditional probability of B given A is that it is between 0 and 1. (Zadeh 2002, p.3)

Without getting into mathematical details, I will just remark that proponents of Fuzzy Theory claim that their theory can provide answers to problems like the ones above (Kosko 1990, p.233) where Probability Theory cannot. That is because Probability Theory is tied to Classical bivalent Logic, which the later restricts the former's applicability. With respect to the limitations of Probability theory regarding the human mind, Zadeh also mentions the following:

> …PT is lacking in capability to operate on perception-based information. Such information has the form of propositions drawn from natural Language-propositions which describe one or more perceptions. For example, "Eva is young", "usually Robert returns from work at about 6 pm"[…]. The inability of PT to operate on perception-based information is a serious limitation because perceptions have a position of centrality in human cognition. Thus, humans have a remarkable capability to perform a wide variety of physical and mental tasks without any measurements and any computations. Everyday examples such of such tasks are parking a car, driving in city traffic, playing tennis and summarizing a story. (Zadeh 2002, p. 3)

Now, when one redefines Probability Theory over Fuzzy Logic foundations instead of Classical ones, only then Probability Theory is able to cope with problems of the kind mentioned above, and is capable of operating on perception-based information. The resulting **Fuzzy Probability Theory (FPT)** definitely contains ordinary Probability Theory. A few basic concepts of the former, as well as some of its applications, I will present next in the following section.

C. Are Humans so Irrational?

It was mentioned in the previous Chapter (Section C) that assuming a probabilistic way of thinking, the "irrationality" appeared in Wason's Cards was a relative one. That is, humans are perfectly rational if we ascribe to them a probabilistic way of thinking rather than an FOL reasoning type. But how about the "irrationality" appearing in Tversky's Linda example? Could that be explained in a way that it would guarantee rationality for humans? The answer is yes, and there is a reasonable justification of why people chose premise (5) as more probable over (2) (see p.30). One could justify humans if one ascribes to humans a fuzzy way of reasoning. I will dissolve the "paradox" of Linda in two ways. I will justify people's choice of premise (5) over (2) by considering, (a) a Fuzzy-Set Approach to the problem of Linda, and (b) Lukasiewicz's Infinitely-Valued Logic:

(a) <u>Fuzzy Set Approach to the Problem of Linda</u>: The key concepts involved here, are the ones of *Subsethood Measure* and *Fuzzy Probability*. In simple words, the subsethood measure, denoted by S(A,B), indicates the degree to which A is subset of B (Kosko 1990, p.226). By fuzzy probability I mean the probability of a fuzzy event, which expresses (deterministically) degrees of subsethood instead of randomness. To make things more precise, let us just recall that a subset of a set is contained in the set, and is contained to 100% degree. But the set also is contained in the subset in some degree. In other words, if

A is a proper subset of B, then S(A,B) = 1 because B contains A completely. Similarly, S(B,A) = 0.2, if let us say A contains, or "covers" B at 20%. This degree that the whole is contained in its part is nothing but the probability (or the fuzzy probability, if the sets involved are fuzzy) of the part. As Kosko explains:

> In general the probability of a set or event A equals the degree to which the part A contains the "sample space" X. The probability of A is how much the whole sticks in or fits in the part of A. (Kosko 1993, p.60)

Kosko explains more rigorously all the above as follows: The Subsethood Theorem states that $S(A,B) = M(A\cap B)/M(A)$, where M(A) is the cardinality[9] of the set A, and is defined by $M(A) = \sum m_A(x_i)$, i = 1,…n, and m_A is the membership function (Kosko 1990, p.220). The formula for S(A,B) is nothing but a more general formula of the formula of conditional probability P(B|A) which, as one recalls, is equal to $P(A\cap B)/P(A)$ (Kosko 1990, pp.232-233). That is because if B contains N trials and A contains the N_A successful trials (i.e. $A\subset B$, and also A, B are non-fuzzy), then S(B,A) = N_A/N which is the relative frequency of successes in trials (Kosko 1993, p.60). The important thing though is that, as Kosko says:

> The N elements of B constitute the *de facto* universe of discourse of the "experiment". […] The probability N_A/N has been reduced to degrees of subsethood, a purely fuzzy set-theoretical relationship. […] Where did "randomness" go? The relative frequency S(B,A) describes a fuzzy state of affairs, the degree to which B belongs to the power set of A[…]. Consider B = X = $\{x_1, x_2\}$ and A = $\{x_2\}$ in the unit square: the frequency S(X,A) corresponds […] to the ratio of the left cube edge and the long diagonal to X. (Kosko 1990, p.233)[10]

Based on the above, we want to re-compute now the probabilities on the Linda

----

9. Kosko also shows (Kosko 1990, p.221) that $M(A) = d_1(A, \varnothing)$, where d_1 is the distance defined by: $d_1(X, Y) = \sum |x_i - y_i|$, i = 1,…,n. This is clear, as one can see, from p.42, that M(A) = 1/3 + 3/4 = | 1/3 - 0 | + | 3/4 - 0 | = $d_1(A, \varnothing)$.

example, viewing Linda as the set L = $\{x_1, x_2, \ldots, x_7, x_8, ,x_9, \ldots, x_{12}\}$ (Linda as a set of properties) which the first seven come from the description of Linda as single, outspoken etc, and the remaining five come from the properties for Linda that the subjects had to chose. For example, x_1 = 31-years old, x_2 = single, x_5 = majored in Philosophy, x_6 = concerned with issues of discrimination etc. The crucial properties of "Linda being a bank teller" and "Linda being active in the feminist movement" are precisely x_9 and x_{12} respectively.

Let $A_1 = \{x_9\}$ and $A_2 = \{x_9, x_{12}\}$ [11] be subsets of L. Now, based on the relation "relevant things go with relevant things" [12], it is more likely that elements x_5 and x_6 will "fall" [13] in A_2 than A_1, on the basis that usually less Philosophy majors become bank tellers, and usually more people that are "deeply" concerned with issues of discrimination etc are active members of feminist movements. Now, in our case $M(A) = \sum m_A(x_i)$, where the sum is taken now for i = 1,…12, and m_A is the membership function (I will take A to be A_1 and A_2 respectively). One observes now that properties x_5 and x_6 do make the quantities $m_{A1}(x_5)$ and $m_{A1}(x_6)$ smaller than $m_{A2}(x_5)$ and $m_{A2}(x_6)$, since x_5 and x_6 are more likely to "fall" in A_2. Hence, we have $M(A_1) \leq M(A_2)$ which in turn implies that $S(L,A_1) \leq S(L,A_2)$. Since these degrees can also represent the probabilities of A_1 and A_2 (call them $P(A_1)$ and $P(A_2)$ respectively), then the people did well choosing "Linda being a bank" as less probable than " Linda being a bank teller and active in the feminist movement."

----

10. That is because, when $A \subset B$, $S(B, A) = M(A)/M(B) = d_1(A, \varnothing)/d_1(X, \varnothing) = d_1(\{x_2\}, \varnothing)/d_1(\{x_1, x_2\}, \varnothing)$ = left edge/diagonal.
11. Observe here that I have given the operator "and" a weaker interpretation. I interpret "and" not as conjunction but as a natural additive "and," more in the sense of "I have the red book and I have the blue book," that is "I have the red book "and" the blue book". Or, "I have a dollar "and" two quarters"
12. That's a very realistic classification, since humans tend to classify things that way.
13. To be more precise, it is more likely that they will have high membership degrees.

Hence, people are not as irrational as they appear to be. In simple words, subjects thought that finding a person (who had feminist feelings before) to be strictly a bank teller with 0-degree of active feminist feelings, is less probable, than having a bank teller with some degree of feminist ideology. Therefore, people did well choosing statement (5) over (2) in the Linda problem (see p.31).

So what went wrong in Tverski's results regarding Linda? First of all, "being a bank teller" etc should not have been viewed as a set (or event) but as a property. What should be viewed as a set is the set {being a bank teller} that consists of one element, namely, the property "being a bank teller." This is more well defined, since viewing "being a bank teller" and "being a bank teller and active in the feminist movement" as sets, one could question of what these sets consist of and what is the meaning of their intersection. Instead, the set {being a bank teller, active in the feminist movement} gives a more natural representation to the double property "being a bank teller and active in the feminist movement." Defining the sets in such a way, one has the framework now to use probabilities as subsethoods. Second, even if Tverski and Kahneman intended to use the properties of Linda as sets in a similar fashion, they completely ignored the degrees of membership in A_1 and A_2 of the "relevant" properties. Their subjects, nevertheless, did not consider all properties as equivalent as they considered certain properties having more relevance and weight to the presented state of affairs that they had to choose.

(b) Lukasiewicz's Infinitely-Valued Logic: We recall the Lukasiewiczian definition of degrees of Truth in a Model. Let M = < D, v > be a model, where D is a domain of objects and v an evaluation function from a set of language symbols to D. Then, the truth value M(S) of a proposition S is given by the following rules:

(1) $M(Rt_1 \ldots t_n) = v(R)(v(t_1),\ldots,v(t_n))$
(2) $M(\sim R) = 1 - M(R)$
(3) $M(R \to Q) = 1$ or $(1 - M(R)) + M(Q)$, if $M(Q) \geq M(R)$ or otherwise respectively.
(4) $M(R \wedge Q) = \min\{M(R), M(Q)\}$
(5) $M(R \vee Q) = \max\{M(R), M(Q)\}$
(6) $M(\exists x(Rx)) = \sup\{Mdx/(Rx)\}$, where d is an assigned value for x (note that if D is finite or if Mdx/(Rx) has a greatest value for some d, then sup = max).
(7) $M(\forall x(Rx)) = \inf\{Mdx/(Rx)\}$, where d is an assigned value for x (note that if D is finite or if Mdx/(Rx) has a least value for some d, then inf = min).

One notes that, the function v assigns "names" to objects in D but also v(R) assigns n-place predicates to n+1-place 2^n tuples, where the first n entries are the $v(t_1), \ldots, v(t_n)$ and the last entry, is the degree of truth. I.e., v(R) is a function from objects to values in [0, 1]. Consider an example from Classical Logic (i.e., the degrees of truth are just 0 and 1):

> Let a, b be names such that v(a) = Jon and v(b) = Michael. Let also H = happy, be the 1-place predicate. Then, v(H) = { <Jon, 1>, <Michael, 0>} meaning that Jon is happy where Michael is not. Now, to find the truth value M(S) of the proposition S = Ha, meaning "Jon is happy", we use (1) which exactly tells us how to evaluate M(S). That is, M(Ha) = v(H)(v(a)) = 1, as Jon is happy from the v(H) set above. So, "Jon is happy" is true.

Same things apply when we consider a 2-place predicate. For example (in Classical Logic):

> Let a, b, v(a) and v(b) as above. Consider the 2-place predicate T = taller. Then v(T) = { <Jon, Michael, 1>, <Jon, Jon, 0>, <Michael, Michael, 0>, <Michael, Jon, 0> }meaning that Jon is taller than Michael etc. Then, M(Tba) = v(T)(v(b), v(a)) = 0, as Michael is shorter than Jon. Hence, "Michael is taller than Jon" is false.

It is not necessary though, for one to restrict on the classical truth degrees 0 and 1. One might as well take the degree of truth in the first example to be in [0, 1]. Repeating example one now, one might find results like v(H) = { <Jon, 0.8>, <Michael, 0.3> }, and M(Ha) = 0.8 (or M(Tb) = 0.3) which translate as "Jon is quite happy," or "Michael is not so happy." In turn, degrees of truth in [0, 1] make proposition like "Jon is happy" above as 80% true, without having to decide between the two extremes of 100% true or 100%

false. Degrees of truth also make the above scenarios more realistic and natural since rarely a human could count as 100% happy, or 100% unhappy. The notion of "happiness" has some built-in vagueness in it, and a degreefied approach seems more appropriate to capture "happiness" in a more natural way, since it avoids also the need for specifying the line dividing happiness from unhappiness, or justifying why every happy person should count as 100% happy, etc. Finally, the nice thing about Lukasiewicz's semantics is the fact that if the only possible values for truth are 0 and 1, then these semantics are equivalent to the classical ones. Thus, perhaps they show why classical semantics works when we factor out vagueness.

Now, in relation with Linda problem, one could interpret "Linda is a bank teller and active in the feminist movement" as $B \wedge F$, and "Linda is a bank teller" as $B \wedge \sim F$. But then, these two statements could be represented semantically by the Lukasiewiczian formulation above, with corresponding truth values $M(B \wedge F) = \min\{M(B), M(F)\}$and $M(B \wedge \sim F) = \min\{M(B), M(\sim F)\}$. Based on the information given about Linda, one is justified in assuming high degrees of truth for F (i.e. M(F) will be quite big, where M(~F) will be very small). Again, based on what we know about Linda, one expects M(B) to be at least smaller than M(F). Hence, one obtains the following inequalities:

$$M(B \wedge F) = \min\{M(B), M(F)\} \geq \min\{M(B), 1- M(F)\}= \min\{M(B), M(\sim F)\}= M(B \wedge \sim F)$$

In other words, the subjects gave higher truth-value to "Linda is a bank teller and active in the feminist movement" than to "Linda is a bank teller", which justifies in a sense their ascription to the former, a higher probability than the latter. Once more, people did well choosing statement (5) over (2) in the Linda problem (see p.31).

In conclusion, I would say that it is my belief that Fuzzy Reasoning is a better way to go if we want to come closer to describing how humans reason. To say whether humans are rational or irrational has to be searched within their domains of thinking, and not to restrictive formal systems that humans themselves have constructed to study non-realistic structures. Fuzzy thinking is closer to nature and it could incorporate human's natural way of thinking too.

D. Objections to Fuzzy Rationality

In this last section, I consider a serious objection to Fuzzy Logic, raised by Susan Haack. I will deal just with this objection because it relates with Fuzzy Rationality, and not with all objections against Fuzzy Logic per se. That is because the suggestion of the thesis is that some of the "irrationalities" ascribed to people could be removed if we assign to people a Fuzzy Logic way of reasoning rather than a Classical one. Therefore, one could at least be justified to claim that natural thinking (which deals with uncertainties, vagueness etc) is closer to Fuzzy Logic than Classical Logic, and if one wants to come a little closer to how humans reason, then Fuzzy Logic could be a better alternative to Classical Logic.

But let us remind ourselves of the rationale behind the suggestion that Fuzzy Logic is a better candidate for human reasoning, so that we can see better on what the objection focuses. T. Williamson could not have given the rationale better in the following passage:

> Imagine a patch darkening continuously from white to black. At each moment during the process the patch is darker than it was at any earlier time. Darkness comes in degrees. The patch is darker to a greater degree than it was a second before, even if the difference is too small to be discriminable by the naked eye. Given that there are as many moments in the interval of time as there are real numbers between 0 and 1, there are at least as many degrees of darkness as there are real numbers between 0 and 1, an uncountable infinity of them. Such numbers can be used to measure degrees of darkness. Now at the beginning of the process,

> the sentence "The patch is dark" is perfectly false, for the patch is white. At the end, the sentence is perfectly true, for the patch is black. In the middle, *the sentence is true to just the degree to which the patch is dark*. Truth comes in degrees. For "The patch is dark" to be true just is for the patch to be dark; for "The patch is dark" to be true to a certain degree just is for the patch to be dark to that degree. Even if we cannot discriminate between all these degrees in practice, we have made the truth of our sentence depend on a property which does in fact come in such degrees. Thus there are at least as many degrees of truth as there are degrees of darkness, and so at least as many as there are real numbers between 0 and 1, an uncountable infinity of them. (Williamson 1994, p.113)

So for example, if a patch X belongs to the set of dark patches to a degree of 0.3, then the sentence "The patch is dark" would be true to a degree of 0.3. This is what Haack would call the "first stage of fuzzification" (Haack 1996, p.234), where object-language predicates determine fuzzy sets, in which objects are members to those sets to certain degrees that range within [0,1]. There is also a "second stage of fuzzification", according to Haack, in which metalanguage predicates such as "true", denote fuzzy subsets of the set of values of in [0,1] (Haack 1996, pp.233-234). For example, the very same way that one can associate to the linguistic variable *Height* the fuzzy values *very tall, tall, not very tall* etc, the same way one could extend this to the metalanguage level on linguistic variables such as *Truth*, associating values like *very true, true, not very true* etc to that variable. Against this second stage of fuzzification Haack raises the objection, which I now turn.

According to Haack, from the first stage of fuzzification one will have, for example, that if X belongs to the set of tall people to a degree of 0.3, then the statement "X is tall" would be true to a 0.3 degree. But then from the second stage of fuzzification, one would assign the linguistic value *not very true* to the statement since its degree of truth is quite low (Haack 1996, p.234). But the problem, as Haack claims, is that it is completely subjective, hence arbitrary, of how values in the first stage are associated with linguistic values (like *not very true*) in the second stage, and to what degree (Haack 1996, p.235). Haack also says that although:

> There are rules for calculating what [numerical] values belong to belong to what degree to *very true* or *not very true*, etc; but the upshot depends on an initial, subjective assignment to the primary term [in this case the term is *true*]. (Haack 1996, p.235)

In simple words, why should 0.3 in our example above correspond to *not very true* or *not quite true* or *false* etc, and how (based on what criteria) one chose the division of *true* into the finite discrete segments consisting of *very true, true, half true, not very true, false* etc.

One could respond as follows: It is true that there is no universal agreement about what linguistic terms the numerical truth values should correspond to, but nevertheless there is big difference between subjective and arbitrary. The correspondence cannot be arbitrary since it would totally ignore the content of the statement "X is tall," and its truth-value of 0.3. Assigning the term *quite true* to the statement, although it has 0.3-truth value, it would undermine either the 0.3-truth-value, or the content of the statement. Intuitively, people when asked, they assign a term to 0.3 that in their mind has at least as much as qualitative difference in "distance" from true, as 0.3 has quantitative difference in distance from 1. That is, they assign something close to *not very true*. Hence, not all linguistic terms are free for choice.

There might be variation in choice among close terms, for example, whether 0.3 should be corresponded to *not very true* or *not quite true*, and perhaps this is where subjectivity enters. But this is not arbitrary. Arbitrary would be, if one assigned to 0.3 above terms like *quite true*, or randomly any other term related to the primary term, ignoring the content of the statement from the first stage or its truth value. On the contrary, assigning terms close to *not quite true* for 0.3, being consistent with what intuitively people might have thought of assigning, makes it anything but arbitrary. Regarding the subjectivity mentioned, one need not to worry too much because it is the type of subjectivity that is bounded by the facts and has factored out arbitrariness. So, subjectively one might choose to use the term *not very true*, or *not quite true*, but this is not important since both make the same point into people's heads which is that 0.3, with respect to 1, should correspond to something that is at least weaker than *half true* with respect to *true*.

Theoretically, the correspondence between truth-values and linguistic terms might look arbitrary, but pragmatically it is not. One might suggest theoretical models (like the examples that follow) to show the extent to which the correspondence can be arbitrary. It is possible for example, that in a society of ultra skeptical people, the 0.3 degree of truth of the statement "X is tall" would correspond to *almost false* instead of *not very true*. Or, it is possible that in a country that genetically engineered all of its habitats to maintain the same height after their 5th year of age, to correspond to 0.3 above the term *false.* It is also possible based on the context a statement is made, as well as its content, to have different corresponding linguistic terms. For example, "God is good" with degree of truth 0.8 might correspond to *false* to an atheist, and *very true* to a theist, or even *false* to a theist if the statement is viewed in the context of the Bible that says, for example, that God is "all good." Finally, it is possible that in a Matrix scenario, any degree of truth no matter what the statement is could correspond to the term *false* for a person outside the Matrix, since all statements made within the Matrix are false. However, in real life with real people, experience shows that 0.3 degree of truth for a statement will result in choices of terms that converge around the term *not very true.*

REFERENCES

1. Cherniak C. (1981), “Minimal Rationality”, *Mind*, Vol. XC, 1981, pp.161-183.

2. Davidson D. (1980), “Psychology as Philosophy”, in *Essays on Actions and Events*, New York, Oxford University Press, 1980.

3. Davidson D. (1982a), “Paradoxes of Irrationality”, in *Philosophical Essays on Freud* edited by R. Wollheim and J. Hopkins, Cambridge, Cambridge University Press, 1982.

4. Davidson D. (1982b), “Rational Animals”, *Dialectica*, Vol. 36, No4, 1982, pp. 317-327.

5. Davidson D. (1984), “On the Very Idea of a Conceptual Scheme”, in *Inquiries into Truth and Interpretation*, New York, Oxford Press, 1984.

6. Davidson D. (1984), “Radical Interpretation”, in *Inquiries into Truth and Interpretation*, New York, Oxford Press, 1984.

7. Ekenberg L. (?), “A Survey of Decision Analysis”, on the Internet at: www.ite.mh.se/~love/beslut/litteratur/oversikt2.pdf/ .

8. Haack Susan (1996), *Deviant Logic, Fuzzy Logic: Beyond the Formalism*, Chicago, The University of Chicago Press, 1996.

9. Jackman H. (2003), “Charity, Self Interpretation and Belief”, *The Journal of Philosophical Research*, Vol. 28, 2003, pp. 145-170.

10. Johnson-Laird P and Wason P.C. (1972), *Psychology of Reasoning*, Cambridge, Harvard University Press, 1972.

11. Kosko B. (1990), “Fuzziness vs. Probability”, *International Journal of General Systems*, Vol.17, 1990, pp. 211-240.

12. Kosko B. (1993), *Fuzzy Thinking*, New York, Hyperion, 1993.

13. Oaksford M. and Chater N. (1998), *Rationality in an Uncertain World*, East Sussex, Psychology Press, 1998.

14. Priest G. (1984), “Introduction: Paraconsistent Logics”, *Studia Logica*, Vol. XLIII, 1984, pp. 3-16.

15. Priest G. (1985), “Contradiction, Belief and Rationality,” *Proceedings of The Aristotelian Society*, 1985/86, pp. 99-116.

16. Priest G. (1998), “Dialetheism”, *Stanford Encyclopedia of Philosophy*, on the Internet at: www.setis.library.usyd.edu.au/stanford/archives/fall2001/entries/dialetheism/, 1998.

17. Priest G. (1999), “Perceiving Contradictions”, *Australasian Journal of Philosophy*, Vol. 77, 1999, pp. 439-446.

18. Priest G. (2000), “Paraconsistent Logic”, *Stanford Encyclopedia of Philosophy*, on the Internet at: www.setis.library.usyd.edu.au/stanford/archives/fall2001/entries/logic-paraconsistency/, 2000.

19. Robertson S. I. (1999), *Types of Thinking*, London, Routledge, 1999.

20. Shier D. (2000), “Can Human Rationality be Defended A Priori?”, *Behavior and Philosophy*, Vol.28, 2000, pp. 67-81.

21. Stich S. (1985), “ Could a man be an Irrational Animal?”, *Synthese*, Vol.64, 1985, pp. 115-135.

22. Stich S. (1990), *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation*, Cambridge, MIT Press, 1990.

23. Taschek W. (1988) “Making Sense of Others: Donald Davidson on Interpretation”, Fellows Seminar-Institute for the Humanities, Ann Arbor, University of Michigan, 1988.

24. Taylor C. (1982), “Rationality”, in *Rationality and Relativism* edited by M. Hollis and S. Lukes, Cambridge, MIT Press, 1982.

25. Tversky A. and Kahneman D., *Judgment under Uncertainty: Heuristics and Biases*, Cambridge, Cambridge University Press, 1982.

26. Williamson T. (1994), *Vagueness*, London, Routledge and Kegan Paul, 1994.

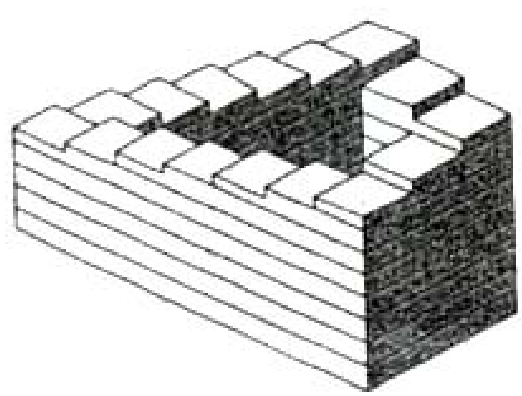
27. Wilson N. L. (1959), “Substance Without Substrata”, *Review of Metaphysics*, Vol.12, 1959, pp. 521-539.

28. Zadeh L. (1965), “Fuzzy Sets”, *Information and Control*, Vol.8, 1964/65 pp.338 339.

29. Zadeh L. (2002), “Probability Theory and Fuzzy Logic”, on the Internet at: www.Ieeesmc.org/dec2002/Probability%20Theory%20and%20Fuzzy%20Logic.Pdf, 2002.

APPENDIX: VISUAL CONTRADICTIONS

It seems that that there are many examples in bibliography of perceiving visual contradictions. Here we consider o visual contradiction that concerns impossible figures. The example was taken from (Priest 1999, p.441) and it refers to *Penrose's Figure* below:



If one begins from the closest corner so to say, to climb up the stairs he finds himself back to where he started from. But that's a contradiction since a continuous ascend from the lowest point should bring one to the highest point. Going back to where we started from implies that the lowest point is also the highest point. Hence we perceive A∧~A directly as true.

VITA

Michael Aristidou was born in Athens in 1975, by Greek-Cypriot parents, who now live in Cyprus. In high school, he was in the Division of Science, where he graduated 1st in that Division and 3rd from all Divisions, receiving two honor awards. He took the entry exams for admission to the Department of Mathematics of Aristotle University in Greece, where he entered 1st nationally. Before entering the University, he served for two years in the National Army as a Sub-Lieutenant a rank he received after six months of training, which included physical and written exams. He received his Bachelor in Mathematics in Greece (two of the four years he was on scholarship), and specialized in pure mathematics, mainly in algebra. Michael was awarded a teaching assistantship from Louisiana State University and passed the Comprehensive Exams at both master's and doctoral levels. He received his master's in mathematics from Louisiana State University in 2001.After passing the doctoral general exam in 2002, Michael started working with Dr. G. Olafsson on Lie Groups and their relations to Special Functions, in pursuit of a doctorate. In the meantime, he is working on his thesis on irrationality, under the supervision of Dr. J. Cogburn. The thesis is for the master's in philosophy, which he hopes to receive in 2004. He also expects to receive his doctorate in Mathematics in 2005. For eleven years, he has been tutoring mathematics at almost every level. He also writes poetry.