CTE

Centrum för tillämpad etik
Linköpings Universitet

Ethical Fading and Biased Assessments of Fairness

- RAMÓN PONCE TESTINO -
Master's Thesis in Applied Ethics
Ethiek Instituut
Universiteit Utrecht
Presented June 2007

Supervisor: Prof. Gijs van Donselaar, Universiteit Utrecht

ETHICAL FADING AND BIASED ASSESSMENTS OF FAIRNESS

RAMÓN PONCE TESTINO

ABSTRACT

In this thesis I present and discuss the phenomenon of ethical fading, and its association with biased assessment of a fair action. Ethical fading is an intuitive, self-deceptive, unconscious mechanism by which even morally competent agents are lead to disregard the ethical consequences of a particular choice. In engaging in this psychological mechanism, I argue, agents are also presupposing a biased assessment of entitlement. This biased assessment of fairness is intentionally dubious, and to be found in decision frames and reinforced by contexts. In the final part of the work I present an applied ethics case to show how ethical fading may be a quite prevalent pattern of behavior.

ACKNOWLEDGEMENT

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1. WHY ETHICAL FADING AND FAIRNESS?

Let's consider this case.[1] During twenty years Paul Feldman, the head of a public research group in Washington, had brought bagels for his employees at the office. He used to leave a cash basket with the suggested bagel's price and his co-workers paid for those they had taken. It was an almost successful honor-system commerce scheme: he usually recouped 95% of his costs. When in 1984, Feldman lost his job he decided to sell bagels. Now, he delivered bagels to the snack rooms of 140 companies in Washington…and the payment scheme was the same. Feldman kept a rigorous business record of bagel's costs and payments done by the employees. By doing that, his business became an interesting experiment, for he could tell how honest his costumers were; and here is where the interesting part begins:

> In the beginning, Feldman left behind an open basket for the cash, but too often the money vanished. Then he tried a coffee can with a money slot in its plastic lid, which also proved too tempting. In the end, he resorted to making small plywood boxes with a slot cut into the top. The wooden box has worked well. Each year he drops off about seven thousand boxes and loses, on average, just one to theft. This is an intriguing statistic: the same people who routinely steal more than 10 percent of his bagels almost never stoop to stealing his money box –a tribute to the nuanced social calculus of theft. From Feldman's perspective, an office worker who eats a bagel without paying is committing a crime; the office worker probably doesn't think so. This distinction has less to do with the admittedly small amount of money involved (Feldman's bagels cost one dollar each, cream cheese included) than with the context of the 'crime'. The same office worker who fails to pay for his bagel might also help himself with a long slurp of soda while filling his glass in a self-serve restaurant, but he is very unlikely to leave the restaurant without paying. (Levitt & Dubner, 2006: 44)

Why people indulge with this type of behavior in particular contexts and what is the psychology behind it? Why some people can short the bagel man but are not eager to rob

---

[1] I am rephrasing the original story from Leavitt & Dubner (2006: 41-7).

his money box? And, more importantly, do they think that they are acting unethically or not? In recent years, interesting research on unethical decision making has showed that morally educated persons can be engaged in "ethical fading". Ethical fading is an intuitive, psychological mechanism by which a person takes a decision without noticing the ethical consequences that her choice implies. Driven by different factors –like pre-commitments, unconscious prejudices, ambiguous contexts, lack of knowledge, and self-representations of oneself as a morally competent agent– the person self-deceives herself in intuitively judging that her decision does not require an ethical orientation.

I believe that the hypothesis of ethical fading is especially important for two reasons. First, it shows that fading ethical considerations in decision making is more common than what we might have thought. In that sense, it opens a space for inquiring how morally competent persons can be rather blind, insensitive or indulgent with certain dubious behaviors under some particular circumstances. Second, it distinguishes some features which enable the engagement in unethical judgments: biased self-representations of moral competence, satisficing types of decision making, personal pre-commitments, ambiguous contexts that shade moral responsibility, unconscious decision frames, etc. As the proponents of ethical fading claim, these features are common to human psychology and dynamics of social interaction. So, in inquiring about ethical fading we are inquiring about how we cope with flaws in ethical reasoning.

Notwithstanding, I think there is also another element implied in ethical fading. An element that is not so clear by itself. People normally engaged in ethical fading seem to be reluctant to acknowledge their misdeeds; a minor amount of people do, but most faders don't. I believe that in this reluctance is an important element for ethical reflection. Biased and blinded by pre-commitments and particular situations, these people seem to feel entitled to act as they did. My guess is that in committing ethical fading these agents are also assessing a biased and intuitive conception of fairness.

1.2. THEME AND PURPOSE OF THE STUDY

The theme of my thesis is the relation between ethical fading and biased judgments of fairness. What I want to show is that in taking unethical decisions ethical faders are not only lead by a self-deceptive psychological drive or by being self-absorbed in a situation: they are also opaquely assessing a notion of fairness that is self-serving. If I am correct in this, it means that some unethical actions which are not motivated by reckless intentions imply biased considerations of entitlement.

1.3. OVERVIEW OF THE STRUCTURE

This work is structured in three chapters. First, in Chapter 2, I present a simple account of moral behavior. The reason of presupposing an idea of ethical behavior is necessary if the main theme to be discussed in this work is unethical decisions. I adopt a utilitarian point of view on moral norms and explain that I will face ethical problems as basic conflicts of interest. Later in that chapter I remark how moral perception is important for understanding how moral agents take or not ethical orientations in particular situations.

Chapter 3 is the main section and is devoted to present and discuss ethical fading. I expose in what consist the psychological mechanism that leads to the fading phenomenon and also present other features of context and behavior that enable its occurrence. The second part is an evaluative subsection. There I draw some practical and normative conclusions of ethical fading. The part on the "normative problem" is where I present the core discussion of this work.

In Chapter 4, I present a case that wants to be the practical exposition of how ethical fading can override normative guidelines. The case presented is a discussion on the notion of futility in biomedical ethics. Finally, in the conclusion I briefly summarized mt general argument.

CHAPTER 2

AN APPROACH ON ETHICS

This chapter would have been called "Rule-consequentialism, conflicts of interest, and moral perception". But that title was too long. After all, the actual one is congruent with what I want to say about ethics. A venerable account says that ethics is about how to live, especially about how to live a good life.[2] That account is not a mislead one. Ethics considers how one should behave; and in inquiring about that question it looks for elements of judgment that help to respond to that issue. However, I will defend a utilitarian approach to ethics, which I think is more appropriate for discussing ethical problems. So, here, personal ruminations about the good life are left aside for virtue ethicists, communitarians, and others.

I will rather present the view on ethics from which I will approach the problem of ethical fading that comes in subsequent chapters. That view is rule-consequentialism. In defining ethics I aim at a simple, functional understanding of why we (should) care about it. I do that in the first part of the chapter, in which I also defend that it is possible to analyze ethical problems as basic conflicts of interest between parties. However, though it is possible to define ethical conflicts in an objective, impersonal way I also claim that subjective understandings of the situation by the parties involved in it trouble the picture. Thus in the second part I will present moral perception as an additional element in the relation of agents to an ethical situation. In sum, my aim in this chapter is to present an objective and a subjective element in our understanding of ethics.

2.1. RULE-CONSEQUENTIALISM

Why ask how to behave? I have that question to be a justified one for we cannot avoid the fact of developing norms and moral beliefs that shape our view of human life. Whether we feel morally motivated to respect them or externally constraint to not trespass them, it

---

[2] See Aristotle's *Nicomachean Ethics*.

surely makes life easier and better. We may disagree on how to define or apply moral terms, but, not regarding exceptional cases, to develop a moral sense is a natural fact. Our attitudes towards what is good and what is bad, what is better or worse, what we like and don't, what we respect and what we hate, and so on towards greyer areas of leisured behavior, are all inescapable to life and do indeed have practical consequences. As Blackburn says, "(…) there is no getting behind ethics. It comes unbidden. It comes with living." (1998:2)

Here I will talk of ethics as concerned with justified moral grounds for action. Because ethics is a matter of common concern justifications for action should therefore be universally accepted as valid reasons. Moral responsibility involves objective standards, maybe not to be discovered as facts but that become obvious and answerable in practical reasoning (Singer, 1993; Blackburn, 1998). As Duff (2005:441-2) claims, responsibility is a relational and a practice grounded concept; the first because one is responsible for something to someone, and the second, for responsible actions are held within particular practices; it is because of these characteristics that responsibility involves always some kind of answerability that is presupposed.

For practical reasons, I will take for granted that people already acknowledge the system of moral norms that is prevalent and reasonably justified in a particular social environment. Yet, I have to make clear that I defend that the more appropriate approach to the discussion and evaluation of ethical problem is rule-consequentialism.[3] Rule-consequentialism is a rational explanation of the utility of some preferences. It claims that the moral code to follow is the code that maximizes expected social utility. 'Expected social utility' means that individuals find utility in respecting a common preference that, they expect, others will also respect. These preferences have a social value which, it is presupposed, all rational agents would be interested in promoting: 'keep your promises', 'respect property rights', 'respect some social-role dependant obligations', 'respect individual physical and psychic integrity', and so on. However, the important thing in the prescription of the code is that it should work in two interconnected fronts: because

---

[3] My whole explanation of rule-consequentialism is based on the exposition of Harsanyi (1985).

rational people strategically expect that others will follow the code, they feel morally motivated to follow it too, and viceversa.[4]

In this line of thought rule-consequentialism also gives a rational explanation and justification of individual rights and obligations. The main reasons for acting morally in this environment is, first, that it is better to act respecting interpersonal, reasonable moral standards of action that imply consequences unto others, in a way that enforces stability and predictability; and, second, that these standards may well be not only a matter of interpersonal calculation but interiorized as desired motivations or "commitment devices" that become interests in themselves for humans (Frank, 1990:73&ff).[5]

According to rule-consequentialism the only reason to act morally is because in doing so persons are serving the common good (Harsanyi, 1985:49): that is, to preserve a justified expectation –one which also they rely on— about the rules that mediate personal and social interaction. And yet, moral rules have the form of hypothetical imperatives. "If you want to serve the common good, then always do action A under conditions $C_1, \ldots, C_n$." (Harsanyi, 1985:49) The conditionality of rules is due to the acknowledgement that concern for common good cannot always require a great amount of personal sacrifice in the agent.

A final thought on immorality. Immoral behavior is not irrational per se. We require moral behavior in our interaction with others, for we need

> "(…) some minimum of decency in dealing with them. Yet, this 'minimum level of decency' we must display as a matter of our own self-interest is typically very much below the standard of behavior we are morally required to maintain. It may be irrational, even in terms of one's own self-interest, to deviate too far and too obviously from society's moral code. But carefully chosen limited deviations from morality that do not provoke too many

---

[4] Rule-consequentialism is very strict in the observance of moral standards but in no sense prescribe the imposition of a specific moral code to others or to personal choice (Harsanyi 1985:46).
[5] Philosophers and social thinkers of different backgrounds agree on this; see, for instance, Scanlon (1982:116; 2003:159), Singer (1993:10, 323-4), or Harsanyi (1985).

unfavorable reactions from other people may be very rational from a selfish point of view." (Harsanyi, 1985:52)

If this is so, as it is, it means that there exists a continuum between rational choice and morality. But how should we evaluate those "limited deviations"? Where is the cut-off point that tells us that those limited deviations have already passed the threshold where its inchoate relation with expected social utility turns serious?

2.2. CONFLICTS OF INTEREST

For the purposes of this paper I will hold that ethics becomes obvious when we see practical conflicts of interest among parties. In conflicts of interest it is clearer how the interest of one party uneasily confronts the interest of the other. This is so for "(…) in the decision, improving the payoff to one party requires bearing some cost by another." (Frohlich & Oppenheimer, 2000:86)

This model is fruitful for discussing ethical issues for in weighing the particular interests involved we will need to incorporate reasons that go beyond the self-interest of each party. In that way, this model does not presuppose any maximal account towards moral norms but rather induce us necessarily to endorse some kind of moral stance for further evaluating or try to solve the problem. As I said above, I will consider a rule-consequentialist approach to moral rules. The reason for this is simple: if a situation involves a conflict in which necessarily the best interest of one part is overridden by the best interest of the other, the ethical element can be objectively defined regardless of the possible biased approach of each part. Thus we may be able to reason objectively about the moral wrongness in the situation.

And yet, in considering ethical responses it is relevant to consider how the agent subjectively relates to a situation. Frohlich & Oppenheimer consider that it is important to see this in order to evaluate whether there exists concern for others or "other regarding behavior" in agent's action. They argue (2000:86) that evaluating an act as ethical or not

requires the consideration of a subjective element: namely to inquire if the agent is cognitively "limiting self-interest or simply acting instrumentally (even if it is in a way that appears to limit those interests) for some other gain."

Here is the main point that I will argue: the premise of fairness that is required for engaging in ethical behavior isn't an assumption without trouble; and that precisely this assumption is subjectively presumed and constructed at the same time. I believe this is one of the less researched topics in how people define ethical situations, and it should be incorporated in the way we reason normatively about ethics if we want to understand people's reactions to their environment. The relevance of the approach promoted by Frohlich & Oppenheimer in evaluating how people relate to ethics converges with the approach that sees the agent as one with an active role in assuming dispositions to the (ethical) situation.

2.3. MORAL PERCEPTION

After having presented the background approach on ethics that supports our idea of moral rules, and having also said that we will see ethical problems as conflicts of interest, it becomes important to consider how agents evaluate if a situation requires a moral response. Thus I have to say something about the active role of the (moral) agent.

The previous subsection advanced this in some sense. I believe Blackburn's (1998) approach of how ethical agents are a kind of *input/output* models may help us in this task. Blackburn views the human agent involved in an ethical situation as similar to a devise that is capable of receiving input takes and delivering other output ones, a double-way function that maintains internal, dynamic relations. The input takes in this case are the representations of a context, situation, character, etc., in a way that their features are cognitively characterized for the agent; the output takes are the due practical responses to the world that the system delivers in the form of attitudes or constraints for action (1998:4-5). Blackburn's idea of depicting ethical sensitivity like a two-takes system

model resembles intentional explanations of how humans and other animals relate to the world.

In viewing a system (in this case the subject) as endowed with rational capacity,[6] the ability to perceive and process information, and the capability to act purposively towards a goal, we can say that ethical responses should also be viewed as intentional states, that is as attitudes that are related to the way the system cognitively adjusts information processes and behavior (Dennett, 1987; 1971) --as we will see in Chapter 3 this account of ethical responses does not mean that all delivered attitudes are conscious responses, for some of them may be unconscious and automatic. The relation between input and output takes is not per se a causal one; the former only serves as motivation or intentional pressure to the latter (Blackburn, 1998:7). The dynamic relation between both should explain why some outputs in the device are evaluated by the agent as ethical responses, while other outputs are held to be not required of ethical orientation.[7] But why we judge that some of our paths of action (outputs) don't need an ethical orientation?

The easy answer to this is to look back to our utilitarian approach. According to rule-consequentialism, a moral agent may disregard ethical consideration in his path of action if it does not go against a moral rule. In this line of thought, she may follow her self-interest and act without worries for ethical considerations. In this case it is not possible to find an ethical conflict of interest. And yet, sometimes things are not that easy. Sometimes the required ethical component is not that visible. We must then enter in the terrain of moral perception, which is particularly important when we want to inquire why some agents are not subjectively aware of ethical features precisely when a situation requires an ethical response --this is a problem that I will touch in depth when discussing ethical fading in Chapter 3. What I want to show briefly is how an agent may act unethically but failing to perceive the moral features involved in the situation. Take this case:

---

[6] That is, a minimal pre-theoretic and logical disposition towards things (even without language).

[7] In accepting Blackburn's model I am not opening the door to one-sided, relative evaluations of what an ethical conflict is. For the purposes of this paper, we –or an impersonal observer– define when an ethical situation is the case and when it is not.

Non awareness of ethical salient features. In a full bus with all the seats occupied, John occupies one seat. Near to him, a pregnant woman carries huge bags of food and is flanked by two children. John is aware of the woman, sees her bags, see some kids there, etc., but cannot extract from what he sees the ethical salient feature: she is pregnant and in a very uncomfortable situation. He may be self-absorbed in the moment. He sees what happens but his attention is insensitive to present salient ethical features (example taken from Blum, 1991). It is possible to not see a moral feature while not having the intention to disregard ethical behavior.

To go further, let me summarize what I have been presenting until now. The general premise from which I departed was this one: that the notion of fairness is subjectively assessed by the agent. This is important for in presupposing some particular idea of fairness people define situations and entitlements, and therefore react ethically or not to the environment. So, it is important to see how agents approach situations, for there is an active cognitive role in how they identify (even unconsciously) moral features of contexts. A moral response is required when the context implies the presence of a moral rule; and yet, a moral rule sometimes is not "visible" to the agent. When this happens, there are two things to evaluate:

(a) It is possible that the agent doesn't see a moral feature, while at the same time it shouldn't be presupposed that behind this inability there is an intention to disregard a justified moral rule. So, the agent doesn't see the ethical consequences of her action; her behavior turns to be unethical, for as a consequence of her choice other's interests are undeservedly affected; but a reckless intention for acting unethically is not to be presupposed.

(b) It is possible that the agent doesn't see the moral feature, while at the same time the agent holds prevalent beliefs, motives and drives that diminish the regard to a justified moral rule. So, does this mean that the agent in unconsciously engaging in unethical behavior is acting recklessly or with some important degree of intentionality?

In both cases the agent is not limiting self-interest; but in (a) there is no intention at all, while in (b) that affirmation is not so clear. Ethical fading implies entering in the murky realm of (b). I will claim in the next chapter that some ethical faders grasp a notion of fairness, but a biased one: while acting they are in a psychological state of self-deception and, also, they are (strangely and in a not obvious way) biased in their assessment of fairness. In what follows I will try to say that ethical fading is not only motivated by a psychological mechanism of self-deception, but that it might be grounded in a normative problem with the assessment of fairness.

CHAPTER 3

ETHICAL FADING

In the previous chapter I explained how I approach ethics in a utilitarian way. I claimed there that the objective nature of ethical conflicts may help us to see a moral disagreement. And yet, I also pointed out how moral perception is a pervasive subjective feature in our understanding of conflicts of interest.

Now, the present chapter is dedicated to a more subjective phenomenon: ethical fading. In the first part I will show what the concept of ethical fading is, and how it is based in a self-deceptive, psychological mechanism that leads agents to fade away the moral color of a decision. In explaining the concept I will also expose the features of context and behavior that enable its occurrence, like for instance the role decision frames play in it.

The second part will present an evaluation of its importance. Jointly with some practical consequences, I will say that the pervasiveness of ethical fading is not only due to an intuitive, psychological propensity but to a more troubling notion of fairness in moral perception. I support this conclusion with relevant research on "bounded ethicality", which is precisely what is presented in the third and last part.

3.1. EXPLAINING ETHICAL FADING

Ann Tenbrunsel and David Messick (2004) introduced "ethical fading" as an explanation model for unethical behavior in business organizations. They were interested in explaining how unethical behavior was so pervasive on a daily basis, a pattern of conduct that was so extensively found in corporate and occupational life that the hypothesis of isolated cases of bad, reckless actors seemed insufficient. In fact, one of the astonishing facts of these happenings was that many of the agents, after having committed bad deeds, claimed to be innocent and rejected the idea of an unethical action being done (Darley,

2005; Banaji *et al.*, 2003; Friedrichs, 2004; Green, 2004).[8] I think their model of explanation may be easily extended to individual agents outside the context of business decisions and corporations.

A conduct that trespasses ethics is not always a likely transparent issue. The ethical implications of a particular choice can be faded away in decision making in a way that the agent can act while apparently being unaware of the existence of these ethical features. When that's the picture, there may be *ethical fading*. Ethical fading is a psychological mechanism that fades out of view the moral implications of a decision, making possible for the agent to choose without perceiving the ethical elements present in her choice. As Tenbrunsel & Messick (2004:224) explain,

> We introduce the term 'ethical fading' to define the process by which the moral colors of an ethical decision fade into bleached hues that are void of moral implications. If we are correct, educating individuals on moral principles is only useful when individuals perceive that the decision has ethical coloration, but useless when ethical fading has occurred or may occur.

But how can this happen? Two elements are fundamental here: its intuitive, unconscious nature, and the possibility of self-deception.

3.1.1. Ethical fading as based in an intuitive judgment. Let's begin with what an intuitive response is. One of the common assumptions about ethical judgments is that they imply reflective deliberation. Reflective deliberation is the rational process by which an agent evaluates a possible action, or motive for action, against some accepted terms of moral justification; that is, she regulates her possible motives for action in order to adapt them to some shared standards of ethical acceptance, or to some categories of 'good' and 'bad' that she may be able to justify to herself and others. In that sense, ethical deliberation may be seen as a trade-off between self-interest and moral principles

---

[8] For an interesting and remarkably well-documented account of the very thin layer that separates legitimate practice from a more dubious incursion in white-collar offenses, see Green's book (2006). Also for white collar crime, Weisburd *et al.* (2001) present investigative data that show that occupational offenders don't have special charactereological differences from normal moral agents.

(Tenbrunsel & Messick, 2004:225). Certainly, ethics is not the extreme opposite of self-interest, but it is understood that for the former to have room there should be a reining in of the latter (Frohlich & Oppenheimer, 2000:86; Frank, 1990). So, even though both can usually converge in a happy way, ethical judgments have many times been held as involving a rational process of adequacy between individual motives and moral guidelines. The problem with this picture is that many of our ethical decisions are not necessarily reasoned but intuitive. Intuitive choices to act involve a sort of judgment but their nature is rather unconscious, uncontrolled, and automatic.

For instance, imagine a traditional Pietistic and deontologically-driven philosopher who has been told that one of his friends is going to be a father but has decided not to form a home with the mother of his future child. In his conservative beliefs about what it means to raise a family, the philosopher does not consider relevant information about the particular relationship between the man and the woman, and therefore does not take the time to give further thoughts to the issue: he just feels an instant feeling of moral disapproval against that friend. This is what an intuitive moral judgment is. It is a psychological mechanism that avoids the possibility of a deliberative, weighing-like reasoning. It is as automatic and affectively driven as an aesthetic judgment can be. Greene & Haidt (2002:517) support this view in the following way:

> We see an action or hear a story and we have an instant feeling of approval or disapproval. These feelings are best thought as affect-laden intuitions, as they appear suddenly and effortlessly in consciousness, with an affective valence (good or bad), but without any feeling of having gone through steps of searching, weighing evidence, or inferring a conclusion. (…) People certainly do engage in moral reasoning, but, as suggested by studies of informal reasoning, these processes are typically one-sided efforts in support of pre-ordained conclusions.

The previous example of the Pietistic moralist may lead readers to believe that all the intuitive judgments I am referring to are morally laden. It is not the case. With the same automaticity with which moral judgments may be done, other judgments that intuitively disregard moral content may also be taken. And that's the case of ethical fading. In

getting involved in this intuitive, unconscious process the person can act illegitimately against the best interest of others or either trump required moral norms without noticing that he is acting wrongfully. In this way ethical fading gets rid of moral constraints and is directly biased by self-interest.

This mechanism is so natural and common that being morally educated is not a warranty for being free of wrongfully mixing, say, professional responsibilities and personal interests. As Moore & Loewenstein (2004) explain, self-interest and consideration for others[9] are two models of thought that influence behavior through quite different cognitive processes. Self-interest is automatic, involuntary, unconsciously dominated, effortless, and even inscrutable in the same way inner processes of vision work; while, to the contrary, other-regarding behavior operates through controlled, more analytically effortful, voluntary, and accessible to introspection processes (Moore & Loewenstein, 2004:190). Cognition and decision making are precisely possible because, normally, both automatic and controlled processes act together reinforcing each other. And yet, sometimes their workings enter into conflict; and when this happens ethical judgments — which operate via controlled processes— are prone to be dominated by intuitive and emotional reactions: therefore in those cases personal interest prevail (Moore & Loewenstein, 2004:190,193; Greene & Haidt, 2002:517). In this model other-regarding behaviors are psychological second guests, and in fact operate only as overriders of a more automatic and prevalent process, namely intuitive judgments biased by unconscious, self-interested motives (Moore & Loewenstein, 2004:193).

3.1.2. Ethical fading as self-deception. Ethical fading is also based in self-deception. The interesting issue about the automatic, intuitive process explained before, is that the self-interested bias of the agent is not by itself "bad": the decision is not seen to be based in a subaltern motive that trumps ethical considerations, because the agent is unaware of the sole existence of that self-interested bias. For her, the intuitive judgment seems to

---

[9] I will use 'consideration for others' and 'other-regarding behavior' as interchangeable terms. I took the latter term from Frohlich & Oppenheimer (2000).

carry no apparent contradiction with some ethical standard.[10] Thus it is not obvious for the agent that in intuitively taking a choice she is "bleaching" the moral color of her decision. Tenbrunsel & Messick (2004) argue that this behavior involves a self-deceiving element.

But how could it be? If an agent is not sensitive to the context in which she is making an intuitive judgment, and therefore is unaware of some criteria that otherwise should concern her, how could she be deceiving herself? Tenbrunsel & Messick (2004:225-6) imply that ethical fading is a kind of self-deception for the person is not aware of how she is forming her opinions and judgments; unawareness that implies that she is obscuring to herself, or putting out of focus, the kind of unreasoned judgment that she is doing, in a way that she avoids unconsciously an evaluative monitoring of what she is entailing in engaging in an automatic choice.[11]

Self-deception should be understood here as the production of beliefs or evidence that one is biased to believe (Mele, 2001:11).[12] For the purposes of this paper I will refer to self-deception as a postulate for explaining the possible underlying behavior that is at stake when a person is likely to carry dubious and contradictory states of belief.

Let's consider some cases of self-deception. I might be an alcoholic but I am not so sure if that label suits me deservedly: I go to bed every night drunk or almost drunk, I've been trying to reduce my drinking for quite a time and yet I drink everyday, and my alcohol habits have begun to interfere with my regular obligations; however, I don't quite realize my chronic behavior, for every time I feel the need to drink my emotional motivation for doing it is so strongly permeated with feelings of self-pity, inability to cope with the moment, depression, etc. For outsiders it may be quite easy to say that I am lying to myself and hiding an alcoholic pattern of conduct, but to me that's not clear. I may or

---

[10] Here decision frames play an important part, but I will talk of them below.

[11] Certainly, there is a debate regarding the extent to which a person could be really unaware of the intuitive forces that drive her to unethical choices, but this is a criticism that I won't consider here.

[12] This is a pervasive pattern of behavior and we can trace in our own daily life common examples of this. Moore & Loewenstein (2004:193) use Tom Gilovich's account of how information-processing may be biased by implicit attitudes: "(…) when people want to believe something, they adopt the implicit 'Can I believe it?' whereas when they do not want to believe it their attitude is 'Must I believe it?'"

may not be able to see the pattern of behavior, but the fundamental issue of my condition is that I am not able, as a unitary self, to do a general judgment on it, for I face a conflict of opposing motivations towards which I am not able to set a hierarchy. It is a problem of inter-temporal bargaining, a tricky conflict of interest within a mindset, and I am not in control of it (see Elster, 1993). Let's see another example. Zeno is a talented student but is unable to organize himself and he is also a high rate procrastinator. He realizes that the way to accomplish his daily reading and writing is to go to work everyday to the Law Library, where no distractions are available. However, every morning, though he is aware that he has made the promise to change his work scenario, he rather perceives that maybe he was wrong and that –besides the usual appearance of flatmates, and his tendency to check the internet and tame his writer's anguish with recurrent visits to the kitchen— working at home is the same as working at the library. Later he will realize that he was (again) fooling himself, but back then he seemed to "honestly" believe that --and in that way unconsciously adapting to a want that biased his judgment. So, he was self-deceiving.

3.1.3. Self-deceptive mechanisms and decision frames. According to Tenbrunsel and Messick (2004:226-31) some mechanisms enable self-deception.[13] These enablers are (i) language euphemism, (ii) the slippery slope of decision-making, (iii) errors in perceptual causation, and (iv) constraints induced by self representations. I will briefly explain them and then will explain what role decision frames play in all this.

(i) A *language euphemism* is a common mechanism by which the agent labels positively a decision or choice that otherwise would also receive a more negative characterization. The relevant thing is that the mechanism works on the assumption that the label doesn't totally cover an opposed meaning but leads the agent to believe that because a favorable label seems suited for the occasion then some legitimate entitlement for that meaning might be the case. For instance, harmful and not justified consequences of a military strategy might be labeled as "collateral damage" or "externalities", in a way that the

---

[13] The core explanation of each of the four enablers of self-deceptive unethical behavior is based on Tenbrunsel & Messick's exposition. I will only refer to them when making direct quotes.

negative result seems not to be denied but tamed by a justified military achievement that is already implied in the positive label (see for instance on labeling and framing states of mind, Lakoff, 1987 & 1988; Lakoff & Johnson, 1980; Lakoff & Turner, 1999). Cognitive science research supports this view: one of the early achievements in thought development is to be able to re-represent knowledge that the child has already described for himself, in a way that later he or she would be able to maintain overlapping but different descriptions of one same experience (Clark & Karmiloff-Smith, 1993:495).

(ii) The *slippery slope of decision-making* is a sort of psychological mechanism that tames the processes of ethical evaluation in decision making by the force of routinization. Two kinds of mechanisms lowdown our ethical reactions here: one that is characterized by a psychic numbing towards ethical features that is caused by repetitive and routinely practice, and the other that is called the 'induction' mechanism. The numbing effect is caused by a constant exposure to the same kind of experience, developing in the agent a repertory of patterned responses or solutions that may alienate new, recalcitrant occurrences that require more detailed evaluations. The second one, the induction mechanism, is complementary to the numbing effect: "what we were doing in the past is OK and our current practice is almost identical, then it too must be OK." (Tenbrunsel & Messick, 2004:228).[14]

---

[14] It's interesting to see the similarity of this pattern of behavior with the underlying choice constraints that lead to 'satisficing' paths of decision. See, for instance, this example (case reformulated on the basis of an example proposed by Dennett, 1986).

> The Utrecht Philosophy Department has set an open competition for a twelve year fellowship for a graduate student. 250,000 candidates apply and it's obvious that a complete and deserved evaluation of all the material is impossible by the time available before the deadline (hiring additional assistants would severely diminish the fund of the award). Nonetheless, equal consideration and the obligation to select the best candidate are required. What to do?
> Finally, it is decided to "choose a small number of *easily checked* and *not entirely unsymptomatic* criteria of excellence —such as grade point average, number of philosophy courses completed, weight of the dossier (eliminating the too light and the too heavy)— and use this to make a first cut. Conduct a lottery with the remaining candidates, cutting the pool down randomly to some manageably small number of finalists —say 50 or 100— whose dossiers will be carefully screened by a committee, which will then vote on the winner." (Dennett, 1986; 8)
> The decision process is done in a very anomalous way: it is a mixture of the brainstorming of some rough alternatives, while also considering the pressure of time, first letting go the possibility to control the process, while finally retaking some last accountable control in decision-making while certainly knowing that all the procedure seems more like a lottery but in which no better or more fit (under the odd circumstances) management was available for coping with the situation.

(iii) *Errors in perceptual causation* means that perceptions about the moral responsibility involved in particular cases erroneously lead some agents to disregard a due ethical concern. There are three reasons for these mistakes in perception. First, there is a proclivity to focus on individual responsibility and not on how the environment or a more broad arrangement of things works: the systematic arrangement in an environment (say, for example, a corporation) may be the key for understanding the prevalence of ethical failures in their employees (see Darley, 2005, for an empirical explanation of corruption chains). Second, the reversed image is also true: a physician could be responsible of a bad operation and therefore be the subject of legitimate blame, but he might deflect blame by self-biased perceiving that others put unrealistic expectations on medical profession and the health system. Finally, people seem to hold a diminished perception of moral responsibility in acts of omission than in acts of commission (see Smith, 2003).

(iv) The *constraints induced by representations of the self* refers to the common subjective perspective that we unwillingly take when judging how and why things happen around us and how we conduce ourselves towards others. While judging the motives and actions of others we tend to believe that we can objectively put ourselves in the shoes of others and understand persons or situations in an unbiased way; but normally this is not so and we cannot get rid of our subjective perspectives. As social psychology research shows, persons tend usually to rely on the assumption that they are moral, competent, and deserving (Chugh *et al.*, 2005; Banaji *et al.*, 2003; Bazerman *et al.*, 2004)

The occurrence of these mechanisms may lead to self-deception and therefore to unethical behavior. But, as I remarked before, the judgment process that leads to self-deception is not a conscious mechanism. So, how do these elements have a weight on the way a decision finally turns out if the agent is not especially aware of them? Here is where the importance of decision frames appears. A decision frame is the subjective notion that the agent has on the alternatives for action, possible outcomes, and associated

What we have here is a particular context for decision-making that poses a problem for normative evaluation, and by which people act accepting ethical legitimate gaps in their behavior.

consequences that are linked to a particular path of choice. Decisions can be framed in various ways, dependant on the cognitive heuristics that underlie the decision maker that faces them. A heuristic or frame is a cognitive devise that helps the agent to optimize her lack of knowledge or uncertainty. As Igou (1999) argues,

> A framing effect is a change of preferences between options as a function of the variation of frames, for instance through variation of the formulation of the problem. For example, a problem can be presented as a gain (200 of 600 threatened people will be saved) or as a loss (400 of 600 threatened people will die), in the first case people tend to adopt a gain frame, generally leading to risk-aversion, and in the latter people tend to adopt a loss frame, generally leading to risk-seeking behavior.

Self-deception and decision frames are behavioral keys that should direct our view to the environment in which the agent is taking decisions (Tenbrunsel & Messick, 2004: 225).

3.2. PRACTICAL AND THEORETICAL CONSEQUENCES OF ETHICAL FADING

What should be our evaluation of the phenomenon of ethical fading? I think a due response to this deserves multifaceted orientations. I believe ethical fading is not a pattern of unethical behavior among others: its importance is particularly special for practical as for theoretical reasons. In this section I will first focus on two practical conclusions. Later, in the final part, I will propose what I believe is the controversial normative problem that lies behind it and overlaps with its psychological nature.

On the first part of my practical conclusions I follow Tenbrunsel & Messick's (2004) picture of the problem, while at the same time I present complementary views that support their account on the pervasiveness of ethical fading. As a next step I present a theoretical normative conclusion. I am aware that the normative problem that I associate with the pervasiveness of ethical fading is a controversial one; to avoid dilation I say in advance that my thesis here (finally, the thesis of this paper!) is that some of these unethical behaviors —which are finally based in intuitive, self-deceiving framings, as I

explained before— imply also a normative assessment (that is, a particular understanding) of the notion of fairness.

What I will defend is that self-deceived agents prejudge fairness in a very biased way. As I will try to support, fairness is a context-dependant notion. What many of the self-deceived agents do is to rely on a biased but contextually justified notion of fairness. It is like some contexts lead agents to self-deceive themselves, and because of that they feel entitle to act in that way. However this is contradictory with the whole description of ethical fading, for it would require a conscious rationalization and intentional unethical behavior. So, what is the explanation? I presume that this assessment of fairness in contexts is patterned in decision frames! If I am correct in defending this claim that means that ethical conflicts are not only difficult to resolve but also that a normative, theoretical problem might be involved and needs further discussion.

3.2.1. Practical issues. The first practical problem that should be mentioned is that ethical fading is hard to correct if the efforts are put only in normative education or ethical training (Tenbrunsel & Messick, 2004:233). As these authors explain, this kind of training comes too late or its accomplishment is no guaranty for avoiding ethical fading to happen. The blinding to ethical considerations comes mainly from the sort of decision framings that the agents intuitively endorse in situations. One of the key elements in the unconscious endorsement of biased frames is self-deception, and Messick & Tenbrunsel emphatically remark that this behavioral pattern "is hard to correct" (*Op.cit.*:234). However, one good advice is to check for the surrounding environment in agent's decisions, for the presence in it of self-deception enablers like those exposed above is determinant for predicting proneness in unethical behavior (Ibid.; Banaji *et al.*, 2003). In fact, how agents relate to their environment, and how the environment seems determinant in the behavioral variables that the agents will follow, is an important new direction in which ethical research has entered in trying to see if the presumed permanence of character traits --which virtue ethicists wholeheartedly presuppose-- is an accurate premise or not. New approaches on empirical ethics and moral psychology seem to prove that it is not the case: the presumed virtue ethics' assumption that moral education may

be directed to the formation of character is misleading, for the permanence of character traits is a hypothesis contradicted by the variability of the agent's reaction to particular contexts (Doris, 1998 & 2002; Harman 1998; Railton, 1995:94).

It is also important to realize that the conflicts of interest ethical fading raise are not of a homogenous type. Some conflicts of interests are easily seen, while others remain invisible or obliquely obvious to the common eye. As we saw in Chapter 2, conflicts of interest are usually good for detecting ethical issues involved. Some of these conflicts are easy to detect, like for instance when a physician accepts economical benefits from an institution and in return refers her patients to clinical trials (Banaji *et al.*, 2003:7; Chugh *et al.*, 2005:20); or when a lawyer obstructs justice by destroying material evidence that incriminates her client (Green, 2006). But seemingly there are also other cases in which the dubious ethical feature seems not so clear to detect, turning itself almost invisible. Take, for example, a case imagined by Blum (1991:704-6). Julio is a leg disabled man who works in a department whose head-office is Theresa. Because of his disability he experiences recurrent pain and part of his work in the division turns to be more difficult to achieve. For that reason he approaches Theresa to legitimately ask for a plan in the office that could help accommodate the company to his particular condition. By legal obligation Theresa accepts the plan but she also begins to release him from more work than he deserves, and in further events Julio begins to feel that Theresa is treating him in a condescend way, like she were not really understanding his condition. In fact, Theresa is unconsciously biased against him for she sees his pain as a kind of weakness or lack of character. She is not really aware of that but the consequence of her reactions to him is that Julio feels offended. There is a conflict of expected interest here, but the way it is put makes its latent characteristics quite inchoate --until more intentional action makes it totally manifest.

The second practical problem is a consequence of the existence of actual but invisible conflicts of interest. That is, that the resolutions of this kind of conflict are difficult to resolve. The reason is that one of the conclusions we can extract from the patterned and biased behavior that leads to ethical fading is that the same agents may be permeated

against criticism; and therefore won't easily accept that they were incurring in bad behavior. Normally, we may expect that the agent will be easily persuaded by reasons and a detailed explanatory "unpacking" of the way she framed her unethical decisions (as, for example, Banaji et al. 2003, recommend), but this needs previous agreement on the existence of an unmotivated or clear-cut moral offense. So, just a platform for discussion here looks like the search for a basic "overlapping consensus" between parties, which is something very hard to achieve in practice (see Bohman, 1995).

3.2.2. A normative problem. There is a complementary normative issue in self-deception, and I believe it is a fundamental problem relative to the assessment of moral norms. Here I reach the core point in my thesis. What I want to defend in this section is the following claim. I believe that in judging or taking decisions ethical faders are not committing a bad slip on the basis of a pure psychological propensity: I rather think that in their self-deceptive state of mind they are also assessing an account of fairness, but under a biased interpretation that is not even consciously reasoned by the agent. Thus it is not only that they are self-absorbed by a situation or by an intuitive drive, and that in that way they act without knowing or without choosing. What I say is that they choose, but they choose following a decision frame that incorporates a very opaque, oblique assessment of fairness that is self-serving in itself.

If this is true, if the notion of fairness is involved in decision frames while committing an ethical fading, then it means that the psychological mechanism also mediates a biased relation (but a relation after all) with a normative stance. To rephrase it: decision frames and self-deception are intuitive, unconscious, automatic drives, but they also imply an oblique normative assessment of fairness that enables the agent to act in self-serving ways. This is not a small issue. In fact, I believe that ethical fading is not a slight slipping but a fundamental key in understanding the problem of prescribing ethical behavior when moral norms and self-interest enter in conflict in an environment (or arrangement of

things) that doesn't reinforce the necessity of acting morally.[15] Again, if this is so, it would rather mean that we should consider in more deep focus how complex is the way we attain moral stances toward others.

Let me now go back and explain how we relate to fairness. Bicchieri (1999), based on experiments on Ultimatum and Dictator games, claims that it seems presumable that our norms of fairness and reciprocity become prevalent not by natural fair-mindedness but by the particular circumstances in which individual interaction and decision making take place. There is no unitary notion of fairness, but many. The assessment of fairness is context-dependent, as is the application of norms due to that previous assessment. As Bicchieri says, "There is continuity between real life and experiments with respect to how 'rights' and 'entitlements', considerations of merit, desert or sheer luck shape our perception of what is fair." (1999:233). Thus in the way people relate to situations there may be a clue on how they relate to fairness; in the same line of thought, there may be also a clue on how to act accordingly to a perception of a fair context that induces or mediates ethical considerations to others (Frohlich & Oppenheimer, 2000:85). Consideration for others or other-regarding behavior, to be possible, requires a previous assessment of a fair context. In acting morally all ethical agents rely on fair play. But what is fair play?

As we saw in Chapter 2 our basic moral stance is rule consequentialism. This utilitarian approach states that we act morally on the grounds of our rational necessity to commonly agree and accept some rules of interaction for all of us. These rules are based on social preferences; that is, because we have evolved to arrange what our common preferences are —namely, the ones that we want to rely on in order to develop our own lives on a stable and predictable basis in our interaction with others— we agree to respect some rules. These rules may be as simple as 'don't kill others', 'don't cheat', 'don't take what doesn't belong to you', 'don't attempt against the physical integrity of others', and so on; but because they touch us so close, because they enable the further freedom to choose and

---

[15] This may be due to cultural, institutional or egoistical drives. However, I won't pay attention here to what of those is the best explanation for understanding how the environment can also bound individual persons' judgments.

to develop ourselves as persons, we deeply internalize their mandate and therefore develop also some higher level stance to feel negatively against those that don't play fair.[16] Nevertheless, it is important to have clear in mind that the only reason to act morally (respecting individual rights and obligations) is to act with a previous commitment towards the promotion of expected social utility (Harsanyi, 1985:54-5). Leaving aside the constraining force of sanctions, if one acts without a pre-commitment to the public good one may be driven by self-interest without feeling the necessity to act morally. Moral rules are therefore based in a rational coordination strategy.

In acting morally --that is, in weighing and avoiding the trespassing of a moral threshold—one is playing fair in trying to accomplish personal preferences, for in aiming at that one is not deflecting from a common expectancy to follow some guidelines in interacting with others. That agent is presuming and acting in accordance to fairness. And yet, Frohlich & Oppenheimer propose (2000:86) to see that an ethical question may be the case when the best interests or preferred alternatives of two parties don't coincide. The interesting element in this minimalist definition of ethical conflict is that it opens the door to the possibility of viewing that the conflicting individual preferences of the parties involved don't belong within the common realm of morally accepted rules or preferences. Normally, I would say, it doesn't matter that the other person has different, particular preferences than those that I have: the existence of an ethical conflict has to do with the exceptional fact that in gaining some payoffs for following my particular preference, the other part will have to cover some undeserved costs. If I disregard this possibility I am incurring in playing unfairly against her.

But what happens if in weighing our own individual preference or interest against a moral rule that we should not trespass (like playing fair), we also judge that the context puts us in an unfair situation towards a possible decision? This is a legitimate question. One of the reasons to disregard respect for fairness in decision making is to know in advance if the arrangement of things in the context puts us in a disadvantaged position. As with

---

[16] On the development of norms and meta-norms as a psychological consequence of social evolution, see Axelrod (1986).

games, cheating and rule-breaking are allowed when motivated by a counteraction that intends to negate unfair advantage (Green, 2006:67). Thus in cases where others are not playing fair it is neither prudent nor morally obligatory to act fairly. And why is this reflection important? It is important because it is possible that when a person weighs a very personal preference against the background of a possible offense (to others or to public trust) but within a context of weak impermissibility, the agent may legitimately feel, intuitively and automatically, that she deserves the possibility to act according to her preference; and that the counter scenario in which she should constrain her desire is rather unfair: for what she would like to do is something that others would do while being in her same situation. And therefore having a moral predisposition turns to be irrational and the imaginary force that constrains her to act ethically becomes an unfair, blind adviser. So, in taking decisions here the agent is framing fairness in a particular way: by making the assumption of fairness she returns to a metaethic realm in which the question on acting self-interestedly or acting morally is again assessed.[17]

This sounds like mumbo-jumbo, I know. However, as I have previously argued, I believe that the agent obscures to herself these rational motives, or its intentionality, but beyond her self-deceptive state that predisposition is to be found in decision frames. What I say is that, in my view, ethical fading may not only be "fading" but also may imply a ruling out of normative reasons on the requisite or not to act more or less ethically; a ruling out that is present in decision frames. This awkward, ambivalent state of mind is what makes ethical fading self-deceptive in itself. So in fading the ethical elements of a decision, agents are also unconsciously rearranging the situation to make a judgment about the possibility to hold it as required or not of other-regarding considerations. In being bounded by a situation, by intuitive forces in decision making, and by previous pre-

---

[17] As an additional remark, I think this explanation of the psychological but also of the normative assessment of morality in these dubious kinds of judgments --like cheating on others, buying stolen goods, public urination, creative accounting, log rolling, etc (some may say that I am putting in the same bag patterns of behavior that should be separated, and maybe they are right; but I believe that all them rely on the same kinds of features that I have tried to expose)-- involved all patterns of behavior that maintain the same oblique relation to moral standards that Fletcher (1966) defended on the basis of how the love of God may indulge us in our involvement in particular situations that tame our weak nature.

commitments to personal preferences,[18] the person may feel therefore legitimately entitled to "bounded ethicality" not to "full ethicality".

What is bounded ethicality? Bounded ethicality "represents a subset of bounded rationality situations in which the self is central and therefore, motivation is most likely to play a prominent role." (Chugh *et al.*, 2005:8) Motivation in this case is an unconscious drive that maintains self-worth. That is, implicit and unconscious mental processes (like stereotyping, decision frames, intuitive heuristics, etc.) in interaction with specific social environments,

> "(…) favor a particular vision of the self in our judgments. Just as the heuristics and biases tradition took bounded rationality and specified a set of systematic, cognitive deviations from full rationality, we endeavor to take bounded ethicality and specific systematic, motivational deviations from full ethicality. (…) Ethical decisions are biased by a stubborn view of oneself as moral, competent, and deserving, and thus, not susceptible to conflicts of interest. To the self, a view of morality ensures that the decision-maker resists temptations for unfair gain; a view of competence ensures that the decision-maker qualifies for the role at hand; and a view of deservingness ensures that one's advantages arise from one's merits." (Chugh *et al.*, 2005:9-10).

I believe that the way Chugh *et al.* (2005) propose bounded ethicality behaviors as not only based on psychological mechanisms but also on context sensitivity evaluations of agents, supports partially the thesis I am presenting.

3.3. BOUNDED ETHICALITY MADE IN THE SHADE

So far I have tried to present what ethical fading is; how it happens and how it depends on the way we engage intuitively and automatically in daily life. Additionally I have

---

[18] Many of the researchers in ethical fading and biased judgments are authors that seem to rely on the idea that unethical behavior is mainly induced by decision framings that are common in environments like business occupations, economic bargaining or, in more general terms, competitive practices. But is it not possible that those framings are more patterned in our dealings with others, to the extent that the self-absortion in competitive environments is not necessary for our own dealing with other's interests is equally conflictive? Or, could it not be for example that the decision frames that mislead actors to engage in unethical behavior are frames that are patterned in various social ways?

presented further accounts on moral agency that support the idea that moral actors seem to be more flexible to particular contexts or inconsistent with theoretically driven accounts of ethics than we should at first glance expect. There are psychological processes that lead persons to conflicts of interests in which they are biased to believe that they are acting fairly and not serving unethically self-interest. The importance of this fact is not that ethics is trumped by self-interest, but that ethical considerations are also bounded by personal or collective pre-commitments and that these criteria (incorporated in decision frames) should also be considered in thinking morally about the world. As I remarked before, ethical fading is not an isolated issue of morally uneducated people.

The reader will have to wait until Chapter 4 to see a detailed discussion of a real case, exposition by which I will try to defend that applied ethics approaches may easily overlap with more prevalent ethical fading patterns of behavior. Applied ethics as a field of work seems to have not seriously considered the possibility and the framing constraints that lead to ethical fading. I think this is not only a gap that applied ethics should cover, but mainly an issue that requires better consideration. Endorsing a moral norm is not enough for respecting moral beliefs. I think this fact is very often left out in applied ethics' approaches. That problem has to do with the way we endorse and defend ethical norms, and therefore it also questions the way we apply a normative framework to try to enhance moral behavior. Ethical fading contradicts any reasoned approach to ethical decision-making, for it is based on an intuitive basis. So what should be a proper response to it considering that its triggering mechanism is difficult to counter? My position on this is that in promoting ethical behavior justifying some moral standards is not enough. What should be done is to address the problem of intuitive understanding of situations in order to achieve ethical development. And in doing that, thinking about how contexts are designed or controlled is not a bad approach.

As I tried to propose in the previous section, the account of psychological mechanisms and constraints that contribute to ethical fading is tangential to the hypothesis of "bounded ethicality". As we saw, bounded ethicality is not other than the application of the framework for bounded rational decisions to choices that involve ethical matters.

Chugh *et al.* (2005:3) explain that being unable to recognize conflicts of interest in decision-making should not be limited to explicit cases of dishonesty, because much of our ethical failures are driven by misguided perceptions, limited rationality, and bounded by the particular constraints of situations and social motivations (see also Bazerman & Banaji, 2004; Mazar *et al.*, 2007). In that sense, the psychological model of ethical fading is congruent with the broader approach of bounded ethicality: which is to check pre-commitments, biased framings and intuitive judgments. The importance of deepen research in this direction is fundamental for ethics;[19] indeed, after 2000 renewed and novel efforts have already been put in that direction. Nevertheless, there remains some reluctance on the soundness of this path.[20]

However, I have to be clear in affirming that because people may be prone to self-serving conducts by being intuitively bounded it doesn't mean that that hypothesis implies the lowering of normative moral accounts of behavior. As with Chugh *et al.* (2005), I do not challenge the general cognitive ability of agents to recognize conflicts of interest and therefore behave ethically. I only do contend that in some cases they may be biased and unable to see the overriding nature of others' interests. My claim is rather that the way this happens and the way this liberates self-serving attitudes is difficult to tame, for decision frames are prevalent and useful cognitive tools. But in no sense I imply by this that we should consider some patterns of unethical behavior as part of our accepted moral standards.

---

[19] Ethical philosophical theory, I believe, has not enough addressed self-serving interpretations of fairness. Psychologists and behavioral economists are working on that right now: take, for example, Chugh *et al.*, 2005; Wade-Benzoni *et al.*, 1996; Babock & Loewenstein, 1997. One interesting philosophical article from the nineties that insists in deepening moral theory research in the direction of compatibilist assumption of responsibility, and from which I took the name for this subsection, is Railton's (1995).

[20] New research on intuitive responses and moral engagement seems to some (see, for example, Singer 2005) to not pose a serious criticism to our traditional philosophical accounts of moral normativity, for in doing that it seems to them that a path in that direction will quickly incur in the famous "naturalistic fallacy".

CHAPTER 4

A CASE EXPOSED

In Chapter 3 I have tried to show a relation between ethical fading and a normative assessment of fairness. I believe that relation is plausible, though I haven't proven it. I have argued that the relation is to be found in decision frames. And yet, to prove it I would rather require a sharper theoretical model and a bunch of empirical data, something that competent readers would have expected to find here and have not. I am aware of this problem in my thesis. Its lack implies important flaws in the development of my argument.

That aside, I believe that my theoretical discussion has practical consequences. I will try to explain what I mean. So far my cloudy argument has been that ethical fading involves not only a psychological component but also a normative one that involves a framing on fairness. A good way to approach the problem of fairness in ethical fading is, I believe, to look from the standpoint of the hypothesis of bounded ethicality. Bounded ethicality claims that in the same way agents with previous commitments, prejudgments, stereotypes, lack of information, and a variety of feelings, are bounded to take subjective assessments of what a rational decision is; in the same way, I say, they are lead conscious or unconsciously to be bounded when it comes to decide if a situation requires other-regarding behavior or not. It is like (un)ethical decisions were driven, and blinded, by an effect of "path-dependence". If the hypothesis of bounded ethicality is right, it would mean that in some particular cases the agent is faced to a decision in which being fair with others would mean to get rid of the pre-commitments that define him not only as a person but as a moral self.[21] If this is true, it means that ethical fading would imply an asymmetric understanding of fairness between the agent and others (those affected by the consequences of the agent's action). Is in this line of thought that I maintained that ethical fading implies a normative problem in the interpretation of fairness. And I think it is problematic not only because the agent is biased by an automatic and unconscious

---

[21] On moral priority and identity, see Nucci (2000).

propensity in the way she took a decision, but also because I believe there is a terrain in ethical reasoning in which the relation between other regarding behavior and selfish motives is not so clear; and it is so because the elements of bounded ethicality are ambiguously present in decision frames.

However, this picture requires further commentary. If what I have said by now is true it implies that because agents have all personal preferences, prejudments, stereotypes, etc, it would mean that "full ethicality" or the possibility to deliberate on the basis of reasonableness or a moral code would be impossible; for it would mean that agents faced with decisions that defy the beliefs and desires behind their bounded self are unable to be ethical in the usual sense. Obviously this is a misguided conclusion. Therefore I believe that bounded ethicality is a hypothesis whose explanatory force is mainly dependant on particular contexts that reinforce by institutional weakness the possibility of dubious intentionality in decision frames. I am not able to explain in what consists the features of these deinstitutionalized contexts, but one good shot may be to look at the four enablers of self-deception that were exposed in Chapter 3. Contexts in which normative assessments are hard to be introduced in decision making, by poor institutionalism or problems of cooperative action, may be good scenarios to test the prevalence of bounded ethicality. In sum, biased interpretations of fairness (in decision frames) and contexts (or factors in the context) are complementary and bounded ethicality remains only a prudent hypothesis for ethical fading phenomena and for some other special cases in which, for example, ambiguity or complexity leave more room for self-serving decision frames (like, as I said, scenarios of poor institutionalism).

In what follows I will present a case that exposes well ethical fading, as it also exposes ambiguities that are a consequence of context and which promote the possibility of overriding a justified normative guideline. That case is a case in biomedical ethics: the justified use of the notion of futility in cardiopulmonary resuscitation (CPR) decision making. I will use Kite & Wilkinson's (2002) research on futility to expose an imaginary case and evaluate it.

In 4.1 I present the normative account that Kite & Wilkinson give, I think justifiably, for disallowing the concept of futility as a sound criterion for decision making in CPR. Based on them, I present an imaginary case in 4.2 of physicians committing ethical fading. Finally in 4.3 I discuss the case. First, I will show how a normative assessment of the soundness or not of the concept of futility may overlap with ethical fading behaviors. Second, I will show how the context of decision making for these behaviors may present relevant factor that turn more feasible the possibility of ethical fading. Third, I will show how the mix of ethical fading and contextual factors can lead overloaded physicians to defend their unethical behavior on the basis of a biased decision frames.

4.1. THE CONCEPT OF FUTILITY IN CPR DECISION MAKING

In a paper on biomedical ethics Kite & Wilkinson (2002) stated that, contrary to what it was commonly assumed in the past, physicians and medical guidelines should be very careful in using the concept of futility for taking decisions such as 'do not attempt resuscitation' (DNAR). In general terms, futility or a futile action were normally understood as an intentional attempt that won't achieve its goals (Kite & Wilkinson, 2002:638); but under a more sound evaluation futility seems to imply some other elements.

According to Kite & Wilkinson, a futility judgment:

(i) "(…) can only be understood relative to particular goals. An intervention may be futile with respect to one aim, eg, cure of the disease, but effective in relation with respect to another, eg, palliation. Hence, whether an action is futile or not depends on its purpose." (2002:639)

(ii) "(…) has to be understood in terms of probability. Futile treatments are those which are almost certain to fail, and have a probability of success that is close to zero." (*Ibid.*)

(iii) Should not be confused with harm. "Sometimes we withhold treatment because it is harmful, because the disadvantages outweigh the benefits (…) Withholding harmful

treatment has a straightforward ethical justification (the principle of non-maleficence, or 'do not harm'); however, the reasons for withholding futile treatment are less clear." (*Ibid.*)

Going further, futility also has two different meanings: physiological futility and normative futility. Physiological futility means that a particular medical treatment won't achieve its "specified clinical objective" (*Ibid.*), that is, won't reach the medical goal for which the treatment is proposed. On the other side, normative futility means that a particular treatment won't benefit the patient. Distinguishing between both is important for in judging normative futility, that is, in asking about the patient's benefit we are making a judgment about the patient's interests (*Ibid.*).

So, futility in itself turns out to be an empty but seriously misleading concept. For example, considering (ii), Kite & Wilkinson (2002) argue that physiological futility is normally not accurately measured in terms of probability rates, though to some extent it is believed that the professionals are able to make reliable reasoning on these issues. If this is the case, then the criterion should be based on a probable rate and not on the concept of futility. In cases of CPR, for instance, even a slight probability is still a good reason to maintain a treatment, even if that risks a non long lasting condition as a consequence of resuscitation, which is something that should be discussed with the patient. And here we face an equally problematic issue related with normative futility. The authors argue that the only person to decide on its benefits and wellbeing is the patient or, in due case, a surrogate decision-maker. So, futility implying normative futility is not a reason by itself, for the presumed futile benefit of a treatment is something that has not to do with the treatment itself but with the patient's autonomous will. Also: to some important extent, the use of the concept of futility for DNRA may also be adduced as a consequence of some kind of weighing: the critical condition of a patient may be weighed against a cost-benefit analysis that, for example, consider a serious lack of medical resources, or a huge demand of medical attention. But, again, if that is the case it means that the reasons for decision making are to be found in the criteria to be weighed and not in the concept of futility. I believe Kite & Wilkinson are right in their assessment of why the notion of futility should not be used in decision making.

4.2. AN IMAGINARY CASE

My problem with this is that a particular physician could judge that behind the label "futility" there may be precisely the right reasons for decision making that Kite & Wilkinson defend. Probability assessment, shared decision making, communication skills, and also contextual constraints (like lack of medical resources, critical demands of attention, etc.) are all factors that require a lot of time, human resources, managerial control, sufficient knowledge and availability of information, and so on, to be able to take decisions according to the normative justification that the authors mentioned rightly support. In fact, he might be thinking that he is taking the right decision but he is not 100% sure.

So, let's imagine a case in which the cut-off line between following the concept of futility and not following it is blurry. Something that would look like this:

Let's take a doctor. Galliani is an average professional, nothing incredible but not a bad professional. He is ethically reliable, but not a moral razor. He is an ethically sensitive man and, as a physician, he deeply believes in the principles of biomedical ethics, Beauchamp & Childress' style.

Galliani works in a public hospital in Lima, the Canevaro Hospital. The hospital is not strictly poor, but has not enough rooms and human resources for the amount of potential patients that fill formularies everyday to receive hospitalization. The public fill formularies to get a vacant and wait for 6 to 14 months for a bed and due treatment. At the same time, and because a past financial crisis in public remuneration, the nurses of the hospital developed a decade ago a strong syndicate in order to be able to defend their right to a decent payment –something that they haven't yet achieve and for which they have to do regular strikes to pressure the Ministry of Economy to increment their salary a little more every 5 months. Paradoxically, nurses rule in the hospital: because the hospital administration changes every year and due to problems of corruption, the more "reliable" and experienced human resources in the institution are precisely them, whose syndicate constitutes a cluster within the hospital. They rule every floor of Canevaro. As a consequence of the low wages they receive, they rotate every 6 hours in order to be able to

go out of Canevaro and work as private nurses or do whatever job they can find in order to subsist. Something similar happens with doctors —most of them good professionals, and some of them excellent professionals— who work for some hours in Canevaro and a higher amount of hours in private clinics, which are the institutions from which they really gain their economical support. Galliani is one of these doctors.

Though Galliani is ethically reliable in general terms, he has to deal with the pre-requisites of nurses, the scarce availability of high technology apparatus and of other medical resources, as probability tests of most of the treatments they realize in the hospital. Even though this habituation to constraining conditions makes his profession more difficult, he hasn't changed his ethical standards: he still discusses with his patients medical decisions when the interest of them is at stake. Yet, when it comes to CPR decision making things slightly change, or so Kite & Wilkinson (2002) might think if confronted with Galliani's pattern of behavior.

Why? For several reasons; but first let's see the facts. Galliani sometimes opts for DNRA based on the concept of futility. He knows the reasons behind Kite & Wilkinson's defense, but he also faces a complex situation that leads him to believe that a strange overlapping of reasons seems feasible with the context in which he does take choices. First, he cannot make 100% tests of probability of physiological futility, so he relies on his clinical eye. Second, when the only possibility to reject DNRA would be to receive intensive treatment by the nurses, he cannot rely on that for the nurses are not qualified nor have good relations with the administrative hierarchy of Canevaro to commit to these kinds of treatment on a full basis –precisely when they postpone their duties in the institution to go to the streets in fight for their political right. So, Galliani doesn't have a reliable basis for trusting attention by nurses, and matched with his melioristic evaluation of the patients' probability of recovery, it seems more prudent to decide for DNRA. Third, the most important thing is that the same Galliani is not aware that he is overriding criteria of evaluation that should be more carefully assessed. To some extent, it might be said that he has developed a faster and to some extend reliable notion of futility that is consistent with the context in which he works, and that make it possible for him to keep doing a public service and also having time for getting the profits that he deserves as a private and committed professional outside Canevaro.

4.3. ANALYSIS OF THE CASE EXPOSED

Here I will very briefly analyze the case. I have chosen a case in which ethical fading is not motivated by full blown self-serving biases, but in which it is possible to find decision frames that are difficult to reject if the conditions are the ones that were exposed in the example of Galliani.

4.3.1. Ethical fading behavior. Galliani is indeed an ethical fader. He acts intuitively based on a decision frame that has incorporated the diminished normative standards that he should follow in order to keep working under the conditions of health care that actually Canevaro can provide to patients. He is in a sense self-deceiving himself but that sole characterization is not enough. He is self-deceiving for, in practice, he contradicts his belief in the normative necessity to rule out futility as a justified criterion for DNRA decision making. But that contradiction is inherent to having already accepted the conditions of medical treatment that Canevaro can afford to give to patients. It seems gratuitous to reconsider this if the exposition of the case above seemed to show it in a quite obvious way. However, what happens when a relative of a patient that was not resuscitated (when other tests proved that there were probabilities of recovery) claims for an ethical justification of the usual "futility" procedure which Galliani frames some of his DNRA decisions? It will turn obvious the morally orphaned ground that he can adduced as an answer.

4.3.2. Factors that induce ethical fading. There are several factors in the example that promote the prevalence of decision frames that override the autonomy of patients in taking part of DNRA decisions. First, the dependency on the complex medical context in which Galliani works makes him adapt intuitively his biomedical ethical principles to a suboptimal level that allows him to keep operating there. Second, Galliani is not responsible for the prevalence of the constrained and poor context in which he labors; he would like things to be different but a serious change in medical standards depends on political, financial, institutional and cultural problems that are so complex and interdependent that the sole questioning the issue seems superfluous and, maybe, even frivolous. Third, he is not the only doctor in the hospital that has to assume this kind of conditions, and other doctors (maybe with the same questioning nature and ethical

convictions of Galliani) may be reasoning like him in accepting that besides the poverty of treatment that Canevaro can offer, the day to day health assistance that patients require are an obvious and demanding priority.

4.3.3. A biased interpretation of fairness. When asked about how guilty is Galliani in rejecting CPR to patients that may have 5% of probability of recuperation, he takes the question personally as a general criticism on his competence as a moral self. Even if he (as a morally sensitive person) finds this attitude as callous or unjustified, in taking the DNRA decisions he is been taking he is already committed with a biased framing of what "fair" treatment means within Canevaro. He strongly believes that any other physician would have to accept that working in that kind of context would involve some kind of contradiction when that person would have to come to terms with the idea that presuming sound arguments for every step of decision would signify an illusion. So, what is "fair" for a doctor committed with public service will (at least in this difficult context) involve some kind of "realism" that disregard the possibility of full evaluation of the reasons behind choices.

CHAPTER 5
CONCLUSION

My general attempt has been to show that some unethical behaviors are covered by a biased, subjective and unconscious assessment of fair action. In explaining how this is possible I have presented and analyzed the phenomenon of ethical fading. I have defended that ethical fading is not only a psychological, intuitive, self-deceptive mechanism by which agents choose to act while fading out of view the moral color of their decision. I have rather argued that in engaging in this kind of behavior a strange faded intentionality is also involved: the agent deceives herself (in not considering that she is acting dubiously) by letting herself be lead by a decision frame or heuristic that implies a self-serving interpretation of what "acting fair" means. In this way the person can engage herself in choices or decisions that release her from full responsibility.

I don't imply that in engaging in ethical fading the agent is necessarily blameworthy. The reason for this is that decision frames are prevalent, common and useful cognitive tools. If applied ethics wants to address the problem of how to improve ethical standards of behavior in particular contexts, special attention should be paid to how framing constraints work in every context or environment. Ethical training or moral education is not enough for improving these standards; they rather overlap easily with the prevalence of ethical fading.

REFERENCES

Axelrod, Robert. 1986. "An evolutionary approach to norms", *The American Political Science Review*, Vol. 80, No. 4 (Dec.), 1095-1111.

Babock, Linda, & George Loewenstein. 1997. "Explaining bargaining impasse: the role of self-serving biases", *Journal of Economic Perspectives*, Vol. 11, No. 1 (Winter), 109-26.

Banaji, Mahzarin R., Max H. Bazerman, & Dolly Chugh. 2003. "How (un)ethical are you?", *Harvard Business Review*, December.

Bazerman, Max H., & Mahzarin R. Banaji. 2004. "The social psychology of ordinary ethical failures", *Social Justice Research*, Vol. 17, No. 2, June.

Bicchieri, Cristina. 2006. "Philosophy: what is to be done?", *Topoi*, 25:21-23.

Bicchieri, Cristina. 1999. "Local fairness", *Philosophy and Phenomenological Research*, Vol. LIX, No. 1 (Mar.)

Blackburn, Simon. 1998. *Ruling Passions: A Theory of Practical Reasoning*. Oxford: Oxford University Press.

Blum, Lawrence. 1991. "Moral perception and particularity", *Ethics*, Vol. 101, No. 4 (Jul.), 701-25.

Bohman, James. 1995. "Public reason and cultural pluralism: political liberalism and the problem of moral conflict", *Political Theory*, Vol. 23, No. 23 (May), 253-79.

Chugh, Dolly, Max H. Bazerman, & Mahzarin R. Banaji. 2005. "Bounded ethicality as a psychological barrier to recognizing conflicts of interest", in Don Moore, Daylian Cain, George Loewenstein, & Max H. Bazerman (eds.), *Conflicts of Interest: Problems and Solutions from Law, Medicine and Organizational Settings*. London: Cambridge University Press.

Clark, Andy, & Annette Karmiloff-Smith. 1993. "The cognizer's innards: a psychological and philosophical perspective on the development of thought", *Mind & Language*, Vol. 8, No. 4 (Winter), 487-519.

Darley, John M. 2005. "The cognitive and social psychology of contagious organizational corruption", *Brooklyn Law Review* Vol. 70: 4, 1177.

Dennett, Daniel C. 1987. *The Intentional Stance*. Cambridge, Ma: A Bradford Book/The MIT Press.

Dennett, Daniel. 1986. "The moral first aid manual", delivered at the University of Michigan, the Tanner Lectures on Human Values (November 7-8).

Dennett, Daniel C. 1971. "Intentional systems", *Journal of Philosophy* 68: 87-106.

Doris, John M. 2002. *Lack of Character: Personality and Moral Behavior*. New York: Cambridge University Press.

Doris, John M. 1998. "Persons, situations, and virtue ethics", *Nous*, Vol. 32, Issue 4, 504-30.

Duff, R. A. 2005. "Who is responsible, for what, to whom", *Ohio State Journal of Criminal Law*, Vol. 2, 441.

Elster, Jon. 1993. "Egonomics: the economics of personal conflict", in *The Great Ideas Today,* 87-135.

Fletcher, Joseph F. 1966. *Situation Ethics: The New Morality*. Westminster John Knox Press.

Frank, Robert H. 1990. "A theory of moral sentiments", in Jane J. Mansbridge (ed.), *Beyond Self-Interest*. Chicago: Chicago University Press.

Friedrichs, David, 2004. "Enron *et al.*: paradigmatic white collar crime cases for the new century", *Critical Criminology* 12: 113-132.

Frohlich, Norman, & Joe Oppenheimer. 2000. "How people reason about ethics", in Arthur Lupia, Matthew D. McCubbins & Samuel L. Popkin (eds.), *Elements of Reason: Cognition, Choice, and the Bounds of Rationality*. New York: Cambridge University Press.

Green, Stuart P. 2006. *Lying, Cheating and Stealing: a Moral Theory of White Collar Crime.* Oxford: Oxford University Press.

Green, Stuart P. 2004. "Moral ambiguity in white collar criminal law", *Notre Dame Journal of Law, Ethics & Public Policy*, Vol. 18, 501.

Greene, Joshua, & Jonathan Haidt. 2002. "How (and where) does moral judgment work?", *Trends in Cognitive Sciences*, Vol. 6, No. 12 (Dec.).

Harman, Gilbert. 1998. "Moral philosophy meets social psychology: virtue ethics and the fundamental attribution error", *Proceedings of the Aristotelian Society* 1998-99, 99, pp. 315-331

Harsanyi, John C. 1985. "Does reason tells us what moral code to follow and, indeed, to follow any moral code at all?", *Ethics*, Vol. 96 (Oct.), 42-55.

Igou, Eric. 1999. "Frame, framing effect" (June 11); see http://www.sfb504.uni-mannheim.de/glossary/frame.htm

Kite, Suzanne, & Stephen Wilkinson. 2002. "Beyond futility: to what extent is the concept of futility useful in clinical decision-making about CPR?", *The Lancet Oncology*, Vol. 3 (Oct.).

Lakoff, George. 1988. "Cognitive semantics", in Umberto Eco, Marco Santambroglio & Patrizia Violi (eds.), *Meaning and Mental Representation*. Bloomington: Indiana University Press.

Lakoff, George. 1987. *Women, Fire, and Dangerous Things*. Chicago: Chicago University Press.

Lakoff, George, & Mark Johnson. 1999. *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.

Lakoff, George, & Mark Johnson. 1980. *Metaphors We Live By*. Chicago: Chicago University Press.

Levitt, Steven D., & Stephen J. Dubner. 2006. *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything*. London: Penguin Books.

Mazar, Nina, On Amir & Dan Ariely. 2007. "The dishonesty of honest people: a theory of self-concept maintenance", available at http://ssrn.com/abstract=979648

Mele, Alfred R. 2001. *Self-Deception Unmasked*. Princeton, NJ: Princeton University Press.

Moore, Don A., & George Loewenstein, 2004. "Self-interest, automaticity, and the psychology of conflict of interest", *Social Justice Research*, Vol. 17, No. 2 (June).

Nichols, Shaun, & Joshua Knobe, 2006. "Moral responsibility and determinism: the cognitive science of folk intuitions"; forthcoming in *Nous*.

Nichols, Shaun, & Ron Mallon. 2005. "Moral dilemmas and moral rules". Article in press, forthcoming in *Cognition*.

Nucci, Larry. 2000. "The promise and limitations of the moral self construct". Presidential Address presented at the 30th annual meeting of the Jean Piaget Society: Society for the Study of Knowledge and Development, Montreal, Canada (June 3).

Railton, Peter. 1995. "Made in the shade: moral compatibilism and the aims of moral theory", in Joselyne Couture & Kai Nielsen (eds.), *On the Relevance of Metaethics: New Essays on Metaethics*. Canadian Journal of Philosophy, Supplementary Vol. 21. Alberta: University of Calgary Press.

Scanlon, T. M. 2003. "Précis on *What we owe to each other*", *Philosophy and Phenomenological Research*, Vol. LXVI, No. 1, 159-61.

Scanlon, T. M. 1982. "Contractualism and utilitarianism", in Amartya Sen & Bernard Williams (eds.), *Utilitarianism and Beyond*. Cambridge: Cambridge University Press.

Singer, Peter. 2005. "Ethics and intuitions". *The Journal of Ethics*, No. 9, 331-52.

Singer, Peter. 1993. *Practical Ethics (2nd Ed.)*. Cambridge: Cambridge University Press.

Smith, Patricia. 2003. "Bad Samaritans, acts, and omissions," in R. G. Frey & Christopher Heath Wellman (eds.), *A Companion to Ethics*. Oxford: Blackwell Publishing.

Tenbrunsel, Ann E., & David M. Messick, 2004. "Ethical fading: the role of self-deception in unethical behavior", *Social Justice Research*, Vol. 17, No. 2 (June).

Wade-Benzoni, Kimberly A., Ann E. Tenbrunsel, & Max H. Bazerman. 1996. "Egocentric interpretations of fairness in asymmetric, environmental social dilemmas: explaining

harvesting behavior and the role of communication", *Organizational Behavior and Human Decision Processes*, Vol. 67, No. 62 (Aug.), 111-26.

Weisburd, David, Elin Waring, w. Ellen F. Chayet. 2001. *White-Collar Crime and Criminal Careers*. Cambridge: Cambridge University Press.