Taking Business Ethics Seriously

Taking Business Ethics Seriously:

Beyond "How Much Would You Sell Your Mother for?"

By

Santiago Ojeda Couchoud

Taking Business Ethics Seriously:
Beyond "How Much Would You Sell Your Mother for?"

By Santiago Ojeda Couchoud

This book first published 2019

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Copyright © 2019 by Santiago Ojeda Couchoud

TABLE OF CONTENTS

FOREWORD

If you are reading this, there is a good chance that you are a postgrad student in Economics or Business Administration, the population towards which the text is addressed. If that's the case, I have both bad news and good news for you (although be patient, the latter will take some time to be spelled out!): the bad news is that most authors, with the acquiescence of the editorial industry and the higher reaches of academia, think poorly of you. *Very* poorly. They think you have a woefully short attention span; that, being raised on a diet of poor quality TV and social media, your reasoning abilities are at best underdeveloped (and at worst irrecoverably stunted); that having to compete with the siren songs of the Internet, uncommon words scare you and long sentences bore you, so the only way to engage your attention is to present you with a text barely above that of *Winnie the Pooh.* Not that I have anything against the writing style of good A. A. Milne, which I find admirable for its purpose of entertaining very young children… I just don't find it the proper tool to acquaint university students with complex, highly abstract topics.

The result of such a sorry state of affairs is a majority of Business Ethics texts that this author finds inadequate to the task of familiarizing their purported readers with the difficulties of the subject, or of providing them with ideas that may serve to guide their ethical decision-making in their professional lives. And the reason for such inadequacy is that there is no way on God's Green Earth to discuss such matters while limiting oneself to a level of discourse proper to be understood by a ten-year-old (or at best by a not-too-bright fifteen-year-old). I'm convinced that ethical thinking is a *skill*: an ability that can be improved by conscientiously *training* it, not that different from other skills like playing chess or weightlifting. As any practitioner of those other skills will tell you, once you are past the rank novice stage, to keep on improving you need to train in ways that challenge your current level. If you just keep playing with inexperienced six-year-olds in chess, or never lift anything but the pink rubber dumbbells (weighing a whooping total of two pounds) in your general-purpose gym you will never reach your true potential, and it doesn't matter if such potential is becoming a grand master, an Olympic champion, or something more modest (but still rewarding).

But we shouldn't judge the surely well-intentioned authors of most Business Ethics texts too harshly, even if their works are the intellectual equivalent of a half-pound pink rubber dumbbell: neatly written, heavy on "practical" applications of the scant theoretical insights and full of citations of recent articles in peer-reviewed journals pertaining to the field, but woefully short on theory, especially of full-blown engagement with the classical texts of plain old Ethics (no adjectives needed). Such texts are a harbinger of our modern times, so it is not surprising if their authors fear not being taken seriously by their readers if they talk too much of Aristotle. So uncool! An old, mushy guy who lived more than two millennia ago! What may he have to tell us, inhabitants of a present that rushes at increasing speed towards a promising future full of miracles and wonders? Nothing at all, that's what! Even worse if they talk about Kant. The "categorical imperative"? you must be kidding! Who even talks like that nowadays? Same for John Stuart Mill. Just look at his portrait! It screams lack of coolness, hipness and fun! Definitely better, in the environment of Business Administration and Economics schools, to just present lots of cases and discuss lots of business situations (and, again, quote as many recent works as possible, regardless of merit) so their readers get the impression that they "get it" and theirs is as respectable and serious a discipline as any other, like accounting or marketing.

The underlying problem of all such texts is to have been written in an era when the core ethical beliefs that, being widely shared, could implicitly guide the actions of most social actors, are dissolving or have been abandoned. Thus, no "serious" writer aspiring to be seen as an undisputed bearer of the "objective truth" would dare to appeal to such a core. Unfortunately, lacking a core of ethical beliefs to rely on, ethical texts end up talking much about business and little about ethics, cursorily presenting the most popular exemplars of each traditional taxa without committing to any or seriously considering they could be in any sense "true". However, if you take truth out of Ethics, if you judiciously try to avoid committing to any one ethical tradition for fear of being accused of subjectivity, or partiality, or lack of scientific rigor, you may as well abandon ethical teaching altogether, as you find yourself limited to a bloodless discussion of dry concepts that are no more relevant to the students' lives than a classification of earthworms would be (a fascinating matter for those so inclined, to be sure, but a bit out of place in a business-oriented curriculum). So instead of analyzing what the great masters of yore had to say about how to live *as if it really mattered to us today* they rush to discuss how whatever belief ever labeled as "ethical" may apply to everyday business situations, assuming the students have little patience for

the musings of ancient Greeks but are eager to hear about the potential shenanigans that can be directed from the boardroom of a powerful corporation.

The sad effect of such a sorry state of teaching is plain for all to see: as ideas are not presented properly, substituted by "tools" that in theory are "ready to be applied" and in practice are "sure to be ignored" or rather "sure to be bent to justify and rationalize whatever behavior seems more advantageous to self, others be damned", students do not learn much of value. But alas! *Values* are precisely what is at stake, and require a darned abstract reasoning to be identified, communicated and eventually shared. In the meantime, society maintains its current course of mild degradation towards the universal reign of nominally utilitarian values where the strong and clever take what they want, and the weak and less brilliant suffer what they must. Not an original predicament, as one of the greatest exponents of ethical thinking (Henry Sidgwick, of whom we will have much more to say in section IV) already said in his *Methods of Ethics*:

> *The truth seems to be that most of the practical principles that have been seriously put forward are more or less satisfactory to the common sense of mankind, so long as they have the field to themselves. They all find a response in our nature: their fundamental assumptions are all such as we are disposed to accept, and such as we find to govern to a certain extent our habitual conduct. When I am asked, "Do you not consider it ultimately reasonable to seek pleasure and avoid pain for yourself?" "Have you not a moral sense?" "Do you not intuitively pronounce some actions to be right and others wrong?" "Do you not acknowledge the general happiness to be a paramount end?" I answer 'yes' to all these questions. My difficulty begins when I have to choose between the different principles or inferences drawn from them. We admit the necessity, when they conflict, of making this choice, and that it is irrational to let sometimes one principle prevail and sometimes another; but the necessity is a painful one. We cannot but hope that all methods may ultimately coincide: and at any rate, before making our election we may reasonably wish to have the completest possible knowledge of each.* (p. 14)

But I mentioned in the opening paragraph both good and bad news, didn't I? Given the universal bleakness of what I've talked about so far, the reader is justified in wondering (impatiently, as it behooves his or her state of bubbling youth) when the vaunted good news is coming.

I have proposed mastering the game of chess, or acquiring a certain proficiency in lifting weights, as valuable examples of skill development, of which ethical thinking is but another instance. The good news is that

being a novice in regard to any skill is a wondrous, extraordinarily rewarding state: comparatively little effort can yield enormous improvements, and in the case of Ethics, a lifelong relationship with one of the highest productions of the human spirit can be soundly grounded in the short duration of a semester if the right concepts are grasped and the right habits start to be cultivated. The key lesson to absorb is that to apply ethical theories, to soundly take good ethical decisions, you need to deeply understand (and contemplate and consider and mull over and turn in your head and discard and come back to them with a renewed understanding) the underlying ideas. But to reach that level of familiarity there simply is no alternative to studying how they came about, what the life circumstances of those that first formulated them were and what kind of questions they were trying to answer.

To think ethically in a business situation, you don't need stakeholder theory or to know what sustainability is or what kind of contract you would have signed in an ideal communication stance, as those are pretty intuitive concepts that can be as much applied to a sound piece of ethical reasoning as to a muddled one. What you do need to know is what a good life consists of, for you and for your fellow citizens; what life you are trying to lead; what legitimate claims other people have on your time and other resources and what consequences of your actions have moral salience.

You need, in summary, more ethics and less business, and to get there you need to know a bit of the history of the prevalent idea of good, a bit of anthropology to understand what makes men tick, a bit of sociology to grasp how incentives are shaped socially and how they are transmitted and embodied in institutions. This book does not pretend to be a treatise on History, Anthropology or Sociology, nor to present a complete picture of the History of Ideas focusing on practical philosophy (as it would require too much space and tax inordinately the student's attention, if having to be absorbed in a single semester), but it will draw freely from all of those fields to present a rollicking example of what sound ethical thinking looks like, and hopefully to pique the students' interest to search for more detailed information (available in the selected bibliography provided at the end of each section).

What this book does pretend to be is a plea for what is almost universally deemed to be a lost cause: that there *really* is something good and noble and inspiring about human life, as opposed to, say, quarks and wavelengths and algorithms (which can also be good and noble and

inspiring in their own right, but in a different sense). Something no less real, but infinitely more valuable. Something that requires us to take free will seriously, and humans as the kind of beings that can exercise it. Something that marks certain states of the world as inherently *better*, more *valuable* than others. What is even more joyous and exciting, such distinct goodness and value and nobility can be learned and acted upon, and people that invest in learning and acting upon them lead lives that are a thousandfold more meaningful, more fulfilling, more worthy than those that do not.

How, then, do I propose to *start* the teaching (because learning about it can only be a process that lasts for the whole duration of each individual life) of such wondrous goodness? I will begin by describing in a very intuitive sense what the subject matter of ethics is, and will advance a more rigorous definition in Section 1. After such a foundation has been laid down, in the next three sections I will take my readers for a ride through the ages to consider the most salient answers that have been offered to the vexing question concerning how we should live. Rather than present the main ethical schools (classical or virtue ethics, utilitarianism and deontology) in an abstract manner, I will in each case describe first the kind of society in which such schools originally took root, so the attentive reader can draw some parallelism between the time when those theories were formulated and our own. I will discuss the questions the greatest thinkers of the age were trying to answer, how they formulated their response, and in some cases the unexpected paths in which their followers took their initial insights. After the main "positive" answers on how to live have been presented and discussed I will devote a brief section to how such answers came to be rejected after the advent of modernity, although the arguments most modern critics appeal to were already honed at the dawn of Classical culture. Finally, to close the book, I will introduce a final section on the application of ethical theories to the business world, understood as a means to achieve universally valid criteria to foster more just economic organizations, with special attention paid to the most salient conflicts that typically arise between the groups within or around those organizations.

A word has to be said at this point about the (for some people's taste awful) style with which this book has been written: This foreword should have provided ample evidence (four pages long already! Who has patience for such verbosity these days?) that this is not your average analytically inspired textbook, but a sophisticated, circuitous, arduous, and hopefully exciting romp through some of the most disputatious and disputed issues

in human history. Reading it cannot but be a tough, challenging exercise that, as a side benefit, will make you better at understanding and discussing complex ideas. Such betterment is a necessary step in the path of becoming a more accomplished moral agent, so consider the potential difficulties and pitfalls of the torturous sentences I've filled the following pages with my very conscientious effort to contribute to your required training, that the current audiovisual deluge you are typically exposed to does more to hinder than to foster.

Also, in my experience the best way to train students to reason ethically is not with (shorter or longer, doesn't really matter) expositions of ethical theories, or presenting them with a bunch of examples of ethical dilemmas "in the real world" without values and without a shared understanding of what is right or wrong (so any possible outcome can be considered "good"… or not), but showing them how to build a narrative of how ethical reasoning itself evolved by answering to the changing circumstances of the social world in which it developed. Because the ability to reason morally is essentially an ability to tell ourselves (and later, others) why we acted as we did. An ability, that is, to provide plausible justifications for our conduct. As much as we would like ethics to be about the *future*, to help us decide in advance what we should do but still haven't done, as frequently as not it is really an exercise about making sense of the *past*, when the dice are already cast and the deed is actually done (and, at its best, of using the lessons we have learnt from that interpreted past to guide our future behavior).

Finally, some words about the title of this work are called for, as being addressed to students in the field of Economics and Business Administration, the antagonistic nature of its subtitle may be understood as an unnecessary snub. Not so, as I hope to convince the attentive reader that the evolution of the moral discourse in the last three hundred years has been oriented to first imagining, then configuring, and finally consolidating and cementing a shared space (indistinctly called "the economy", "the market" or "business") whose main feature is precisely to be exempt of the traditional ethical considerations. A space in which the rules on how to behave towards others were superseded by an overarching rule of "utility maximization" (but of a kind of utility that no classical utilitarian, the supposed masters of the concept, would recognize). A space populated by strange creatures, understood primarily not as "neighbors" or "fellow citizens", or "countrymen", but as "economic agents" whose main feature would be precisely, to have no agency at all, and as emotionally detached from ourselves as possible. I don't approve of such an initiative,

and I don't condone the supposed legitimacy of the market (or of any similar environment where business is conducted) to discard the precepts and commandments that humans have painstakingly developed in the previous two millennia, and replace them with what tends to be presented as impersonal, valueless rationality. No rationality can be valueless (rationality being a value in itself, and one that cannot be justified only with appeals to its own content) and no rule that tells us how to behave towards other people can be impersonal.

Which is to say, the ethics that should apply to business relationships is in no way, shape or form different from the ethics that should apply in any other area of our life: family relationships, religion, politics or friendships. That's why it makes sense, even for "business administration" or "management science" practitioners, to be suspicious of something (somewhat pompously) called "business ethics", and to be strongly, unequivocally, passionately and wholeheartedly committed to "ethics", with no labels attached and no unnecessary qualifications added.

SECTION 1 –

WHAT ETHICS IS ABOUT

Its Scope and Limitations

There is one sense in which ethics has to be considered one of the most difficult matters on Earth. After all, we as a species have been discussing it endlessly since we made our appearance on the African Savannah about two hundred thousand years ago, and documenting such discussions by writing them down in a recognizably modern form for more than two and a half millennia, without being able to reach any definite conclusion. Very brilliant minds have devoted significant chunks of their lives to analyzing "it", understanding "it" and, eventually, reporting back on "it", trying to tell the rest of us how we should behave, only to see their best efforts shot down and mocked by the next generation of similarly brilliant successors.

But there is another sense, which I would like to point the attention of my readers towards, in which ethics is pretty darn simple. Consider how the following personalities, or "complexes of traits", which we will call simply Complex A and Complex B, without further qualifications (yet) describe themselves:

Complex A:

- I follow a (very lenient) set of rules myself, but demand a different (much more stringent) set of rules for everybody else
- If I could freely distribute some scarce good (that is, without fear of retribution) I would keep most of it for myself, and would leave very little, if anything at all, to everybody else
- If I had to choose between scratching my little finger and letting the whole Universe implode or saving the Universe at the price of enduring a minor itch, I would choose the first option
- I'm perfectly fine with enjoying myself amidst great suffering of others

- I wouldn't sacrifice any time or material possession, however small, to ameliorate the life of others, even if the improvement for them were enormous

Complex B:

- I try to follow the same rules I expect everybody else to follow
- When it comes to distributing scarce goods, I understand I have no more claim to them than the rest, and am content with a portion similar to that given to everybody else
- I am willing to forgo considerable pleasures (and even endure substantial pains) if such forgoing (and such endurance) significantly alleviates the wretchedness of other people's lives (or significantly increases their wellbeing)
- I feel empathy and am necessarily distressed by the plight of any sentient being, so I could never enjoy or be at ease in the vicinity of any suffering
- I'm happy to donate time or material possessions of mine to see other people's lives improve, even a little

You don't need a PhD in philosophy to understand that Complex A is evil (or any other similarly negative term you may choose: unethical, bad, despicable, morally wrong, deserving of condemnation –eternal or otherwise, and what have you) and Complex B is good (again, also known as virtuous, ethical, admirable, morally right, deserving of praise, etc.)

Unfortunately, few personalities and, more to the point, few behaviors conform neatly and indisputably to those extreme categories. There are, for sure, some historical exceptions where either extremely saintly or extremely knavish behaviors were exhibited, but even for those we could find advocates willing to justify them by appealing to special circumstances. The vast majority of people we cross our paths with, and the vast majority of actions we witness, take place in a continuum between what we have termed Complex A and Complex B, without it being possible to identify a clear-cut frontier, a prominently marked point, separating the behaviors belonging to one extreme category from the ones belonging to the other:

Externally perceived as:

- Deserving condemnation
- Shameful
- Spiteful
- Despicable
- Morally wrong
- ...

- Deserving praise
- Source of pride
- Worthy of recognition
- Admirable
- Morally right
- ...

Action 1   Action 2

A   **B**

Internally perceived as:

- Partiality towards myself
- Egoism
- Sloth
- Callousness
- Lack of empathy
- ...

- Equanimity
- Generosity
- Work for others
- Concern
- Sympathy
- ...
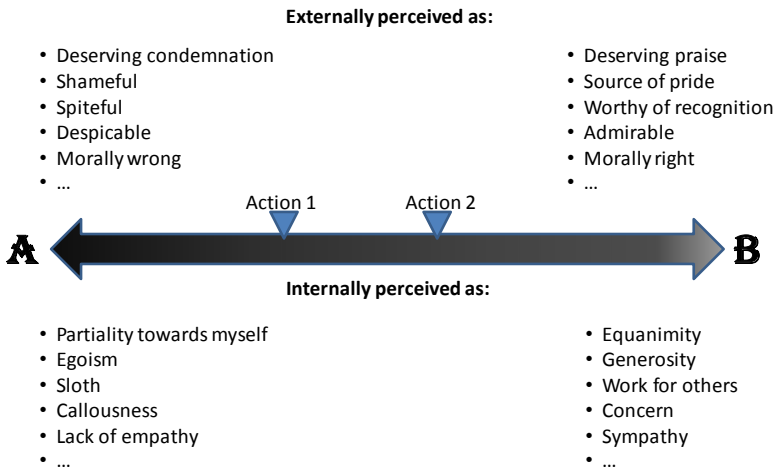
Figure 1.1 The Ethical Continuum

A couple of things have to be taken into consideration before we delve deeper into the subtleties of what makes certain behavior, or certain personality traits, be considered as standing closer to Complex A than to Complex B: 1) an innate bias that is part and parcel of human nature and 2) the symbolic nature of human language

1) Human nature (yes, there is undeniably such a thing, although it has fallen somewhat out of fashion in some circles) includes a certain bent towards egoism and confirmation bias, that makes us perceive our own actions as nobler, less self-interested and more deserving of praise than those of others in similar circumstances:

Figure 1.2 Ethical Valuation Bias

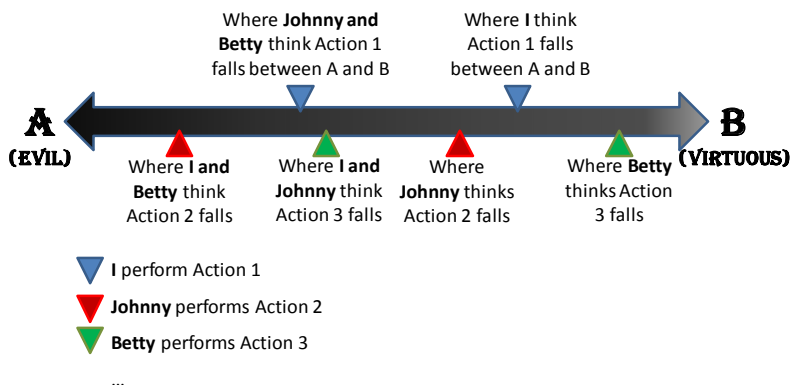The lesson we have to extract from such an expectable, and easily measurable difference in the perception of the moral worth of the actions of different moral agents we interact with is NOT that there isn't really something like "moral worth" to begin with, and that everything, being subject to different opinions, is "just relative", "just opinion", so any action in the end may be defended as being as praiseworthy as any other and requires necessarily to appeal to another source of legitimacy to be justified. Such a source may be the greater strength of those defending it, or its better alignment with "human nature", or its stronger assertion of the "will to power" of the agents performing them, or any such justification, having little to do with what we traditionally understand as morally relevant considerations. We will deal more carefully with such an objection in Section 5 (on skepticism), suffice to say now that they are all wrong in a fundamental (self-defeating) sense.

Rather, the lesson to be extracted is that how we judge other people's actions requires more information than we are typically willing to gather, as they will surely have justifications we ignore, act under pressures we know nothing of and negotiate and compromise between opposite intentions that can be as legitimate as our own. In summary, that you should "be kind, because you don't know what other people may be going through" (or "what battles they may be fighting"). Attentive readers may identify in this feature the origin of a strong impulse in ethics, at least since the modern era, towards universalism (and making abstraction of one's own situation) as a regulative ideal, a motif we will explore in

greater detail in Section 4 (on deontology, which elevates universalism to the preeminent position of being the keystone of practical philosophy).

2) Our symbolic, hence necessarily imprecise, language: it is highly risky to contend for any given behavior that it is uniquely human. As we devote more time and effort to studying our fellow animals we keep on finding them exhibiting more and more traits we long believed were distinctively ours: building complex structures, using tools, identifying themselves in a mirror, adorning their bodies with different shiny implements or culturally transmitting habits. These are all things we have found other species also do. After such caveat, so far, we have evidence of symbolic language, of using symbols to represent vaguely bounded areas of reality, only between *homo sapiens* (although whales and elephants may be quite close to having a similar capacity).

That seems to give us a good indication that symbolic language is evolutionarily quite a big deal (and quite a recent one). Indeed, upon reflection, it provides us with unparalleled flexibility to coordinate our activities, plan ahead and devise completely novel ways to adapt successfully to our environment. An environment that, given that flexibility, is more and more determined by other people, other language users, and less and less by brute, unresponsive nature; for us human nature and culture, rather than being in opposition, are inextricably intertwined. But such flexibility comes at a price: it belongs to the very essence of words to be imprecise. For symbolic language, lack of precision is not a glitch, but a design feature. However, the fact that we can never univocally determine once and for all to what aspects of reality each word exactly applies has driven thinking people nuts for ages, and some of the most brilliant minds have (unsuccessfully) strived mightily to try to overcome what they saw as a correctable imperfection. And in no field has their lack of success been more felt than in Ethics. Let's consider two vignettes, containing an ethically loaded description of the preliminaries of a certain action:

Vignette A

Johnny has always shown, ever since he was a little boy, a great capacity for moral indignation. Raised under the yoke of an unjust (and internationally condemned) foreign occupation of his ancestral land, he witnessed almost daily the hopelessness and humiliations that his countrymen were subjected to.

Being a courageous and upright young man, he joined the underground resistance movement that struggled to gain for his people the minimum of freedom and dignity that every human being should be granted (as reflected in the International Declaration of Human Rights of the UN). There he gained the confidence of his leaders as he was resolute, faithful and disciplined, a good comrade and a loyal soldier.

The latest mission assigned to his cell is to travel to an undisclosed location near the border with a piece of military equipment (a missile launcher) provided by foreign sympathizers and fire it towards a military garrison in a village at the other side (belonging to the occupying power). Although Johnny recognizes that the missile's guiding technology is somewhat crude and there are minor chances the civilians in the village may be hit, he still thinks it is an admissible risk. Villagers near the frontier are known for their expansionist sympathies, and their enmity towards Johnny's people, and he has seen too many innocents killed among his brethren to give the plight of the villagers too much weight.

Vignette B

Ali has always been, since a very young age, a rebellious and unruly brat. Brainwashed by his family and friends to blame all the ills that befell them on evil foreigners, instead of on the shortcomings of the corrupt politicians they keep on electing, he has grown bitter and resentful.

Such bitterness and resentment have led him to associate with a terrorist organization widely condemned by the concert of nations, which has repeatedly shown its disregard for human life and dignity, starting with that of its own members, who are frequently (and barbarically) executed on the slightest suspicion of collaboration with the enemy (which many times is used as excuse to hide internal power struggles). Ali has himself happily participated in some of those executions, proving to be a reliable

and docile hand of his fanatical superiors.

However ruthless his actions may have been up to this date, his latest assignment requires an additional degree of recklessness and disregard for human life, as he is asked to lead a commando action to shoot a missile at a village close to the border, causing as many victims as possible within the civilian population. The village has a medium-sized school where children play unsuspecting of the wanton destruction intended upon them.

As my astute readers will have surely noted, both vignettes intend to describe exactly the same event, under a very different light (and any informed reader probably can guess without much help the time and place where it is situated). They have been depicted in a somewhat extreme, even cartoonish way, as I do not want to direct my readers towards favoring one interpretation over the other. What I want is to highlight the powerful effect of the language we choose when presenting each situation, the same power we are inadvertently subjected to when thinking alone about such issues ourselves. Focusing on the familiarity of the main characters (are they an "us" we can identify with, like Johnny, or clearly a "them" like Ali?), on their agency or lack thereof (are they "capable of moral indignation" or "brainwashed"?) and in the salience we give to potentially mitigating circumstances (do we recognize that he has seen many injustices, or focus instead on his docility abetting evil deeds?) we may end up assigning completely opposite moral valence to the acts of others… or to our own. The key message here is that such variability of moral valence is not due to some preventable, correctible feature of our language. It is absolutely foundational, as without such imprecision it would not be serviceable to speak about morality and ethics in the first place.

To sum things up, because of human nature and how language works, the answer to the very foundational questions that Ethics is concerned with, like how to live, and how a life well-lived looks like, have to be answered collectively. To reason ethically we need to involve (imaginatively or actually) the different human beings affected by the outcomes of our reasoning. We have to grant them a dignity similar to the one we assume ourselves. We have to give their interests and values the same weight we give to ours. How to behave ethically is not something one can decide alone.

A final point to consider is that we necessarily inherit a good deal of such knowledge, so we have to listen not only to what people alive today may have to say, but to what other people long since departed had to say in their own times. The engagement with tradition, in the form of the refined thoughts of the most salient ethical thinkers, will be a constant in most of the remainder of this book.

Definition of Ethics

Given the scope we have identified, and the constraints imposed on its expression (due to unavoidable bias towards our own benefit and the ambiguities and limitations of language), I propose the following definition of the subject we are dealing with:

> **Ethics is the discipline devoted to finding rationally how humans should live, and to justify and explain what a good life consists of**

Let's then begin our journey by unpacking the proposed definition, as some of its terms are complex enough as to merit some extra consideration:

- Ethics is a **discipline**: That is, an **organized**, internally **coherent**, body of **knowledge**. People may voluntarily choose to engage with such body of knowledge, and thus make it a **practice** (in Alasdair MacIntyre's sense, we will return to him later on): they will accept the authority of acknowledged previous masters of the discipline, and struggle to adapt it to their current time, discarding what they think is no longer relevant and developing those parts they think are insufficiently understood. Some authors have called it a **science**, and depending on what you understand by that term it may be a valid part of the definition: if you take science, like R. G. Collingwood did, to be a collective enquiry to answer some question (in this case "how should I live?") while being willing to subject the evidence you bring to bear in such a quest to public scrutiny (so other people can judge and eventually validate your tentative answers) then yes, ethics is indeed a science (and so is Economics, Politics and History, which was exactly Collingwood's concern). If you take science, like Karl Popper did, to be a set of propositions that can be experimentally falsified, then no, ethics is definitely no science (and neither are Psychoanalysis, Astrology or… most of Economics or Psychology) but is not worse off for it.

- Ethics is committed to **rational** inquiry: it aspires to provide **reasons** to act in one way rather than another. Those reasons, necessarily conveyed in plain **language**, should ideally be able to **convince** any speaker, that is, to inspire him to **voluntarily** follow the proposed prescriptions. Being supported by language, ethics is addressed to other free human beings: the behavior of animals, or plants, or machines can be good or bad for us, but it cannot be judged as ethical or unethical. Ethics presupposes other people, endowed with the same rationality as we are ourselves, and whose interests are as worthy of consideration as our own. There is no distinctly ethical way of living on a deserted island, and it is debatable if there can ever be ethical prescriptions for "Supermen", for Gods or Beasts alone in a universe of their own where they have to respond to nobody. Being conveyed in everyday language, Ethics has to deal with the unavoidable ambiguity and imprecision that comes with the medium we use to communicate with each other. Although it has been attempted to develop a perfectly clear and rigorous "philosophical language" in which to develop unmistakable, unambiguous precepts that every listener would understand in the same manner, such attempts have generally ended in abject failure.

- Ethics is concerned with how we should **live**: it deals with very general, overarching precepts. It does not consider every little detail of our lives, like how we should tie our shoelaces, or if we can wear a plaid trouser with a striped shirt, or in what order we should greet our guests at a formal dinner. It does not deal with other more practical, similarly immediate issues, like how to design a building (that's what architecture does) or how to organize a company to maximize its earnings (the area of Business Administration) or how to solve a set of linear equations (that would be the domain of Mathematics). However, it purports to inform and to oversee all of those fields and more, providing guidelines on how much time to devote to each of them in different circumstances, or how to balance our pursuit of them with other demands on our time, like caring for an ill relative or playing with our kids or hanging out with friends. Like an orchestra conductor, Ethics aspires to be the final arbiter on how all our different interests should be organized and prioritized. There is no "more ultimate" authority to tell us how we should balance the demands of Ethics with those of any other

field; by definition, Ethics is the final arbiter, with no further court of appeals.

- But, even while being so general, Ethics aspires to be **substantive**: for it to be meaningful, to successfully guide us in how to live, it has to have **content**, to correctly identify what that good life that it leads us to consists of. It can (and indeed, it should) tell us that having friends is better than having money and thus, *in certain situations*, spending more time consoling a bereaved friend is ethically superior to spending more time at work to make more money; that being fair takes precedence over obtaining the maximum possible benefit and thus, *in certain situations*, we shouldn't push to reduce the price of a business transaction, even if we have the upper hand in the negotiation; that every person, regardless of rank or station, deserves to be treated with dignity and respect, and has certain inalienable rights and thus, *in almost all situations*, it is ethically unacceptable to abuse, heap scorn onto, or publicly belittle a subordinate in the job, no matter how badly he may have screwed up. Such content must, of course, be up for scrutiny. That's what we mean by it being justifiable and explainable: Ethics should prepare us to justify and explain to our fellow humans why we act as we do, what our reasons are, and how we think acting on those reasons leads all of us to a better life, to a life more worthy of being lived.

Values

We have already covered a good deal of what ethics is about, talking schematically about types of behavior and the response they cause in us, be it of blame or praise. We have tangentially noted how such behavior is "socially" valued, which is to say the criteria for that valuation (as praiseworthy or blameworthy) are provided by the society, the group of people with a common language and common traditions and institutions we live in. As we already said, a creature living forever outside of a society, a creature with no interaction ever with other sentient, intelligent beings, has no need for ethics, and it is questionable if his behavior can be ethical (or not) in any meaningful way.

Before we start discussing how such socially-constructed (or socially-discovered, depending on your previous epistemic and ontological commitments) values have evolved, a word is called for about them, as in

a certain understanding of the discipline of Ethics its central goal is precisely the study of values themselves, and its principal aim to determine what is valuable and why. Considering what values *are* will take us on a brief detour through the historical genesis of computer science, whose connection with our main topic will become apparent only towards the end. Thus, bear with me patiently.

Humans have spent enormous amounts of ingenuity and effort to design and build machines that could free us from most repetitive tasks required to obtain the commodities necessary for our well-being ever since the beginning of our species. The first successful attempts provided us with the earliest tools (like the lever, which allowed us to multiply the strength we could apply at a single point, the potter's wheel or the inclined plane) and have advanced ever since, epitomized in the days of the Industrial Revolution in sophisticated contraptions like the steam engine, that could substitute more traditional sources of power without ever getting tired or frail with age. In those days, some bold thinkers (like the German Gottfried Wilhelm Leibniz, or the French René Descartes) already toyed with the idea of unburdening themselves not just of the tasks in the "physical" world that required until then considerable exertion and discomfort, but also those that took place in the "mental" world, supposedly performed by our higher faculties, but demanding no less strain and effort.

Beyond somewhat crude aids to numerical calculation (the abacus was probably known in ancient Sumer more than 2000 years BC, and tables to aid in the resolution of trigonometric problems have also been found dating from that time), it was only when technology had already mastered the manipulation of electricity, after the second World War, that we start seeing a fruitful effort to simulate some of the workings of the human mind. The widespread substitution of printed circuits for transistors (supported by vacuum tubes) in the last quarter of the 20th century allowed for a level of miniaturization and an increase in the density of computations that could be performed by tiny machines that seemed finally to put within humanity's reach the old dream of having those machines take over the unpleasant task of having to think for ourselves. A brief overview of how such endeavor has proceeded will show itself to be highly illuminating in our current research.

A defining characteristic of the program for achieving a "general purpose artificial intelligence" (GP AI) has been, since its inception, to consider as the most salient feature of said intelligence whatever we were

able to replicate in algorithms and that could in turn be executed by a computer (or, to be more technical, by a universal Turing machine). Thus, when all computers could do was compute (hence their original denomination) the prevailing opinion in the emerging field was that thinking was an especially convoluted way of adding, multiplying, dividing and subtracting, with some logical operations that could also be reduced to arithmetic operations thrown in for good measure.

So it came to be that computers got good at performing arithmetical operations beyond their designers' wildest dreams, which theoretically could allow them to decide on the validity of any statement of number theory and break into new areas of knowledge that mere humans would be unable to reach (it's a pity Kurt Gödel had shown a few decades before that such decidability was essentially impossible to achieve), but nobody seemed to be able to use such ability to advance an inch closer to the vaunted GP AI. Indeed, the field stagnated and a lot of lofty expectations were dashed.

However, one of the concepts developed in the era right before the stagnation (recursive neural networks) came to the rescue, as given enough data, it could be used to simulate another feature of our cognitive system that until then had been untapped: the ability to categorize and construct hierarchies of ever more abstract nested categories. This approach underlies current initiatives of what has been dubbed "machine learning" or "deep learning", that has experienced an exponential increase in interest, funding and popularity in the second decade of the 21st century, although the outlines of its functioning were sketched in the very influential book *On Intelligence*, published by Jeff Hawkins in 2004 (with the tellingly ambitious subtitle *How a New Understanding of the Brain will Lead to the Creation of Truly Intelligent Machines*). The core of this approach is that categorization is all that the brain does, so once we got deep enough neural networks, regardless of what physical device supports them, that can categorize as deeply and recursively as our brain can, we will have an actionable and functioning GP AI.

Predictions from then on run all the gamut from the eminently sensible to the downright bonkers (the arrival of the "Singularity" beyond which all we can predict is either Sha-la-la land or the annihilation of the human species, downloading our minds to imperishable silicon devices that will make death obsolete and the like), although some voices have recently started sounding the alarm that the field is entering a new phase of stagnation and most of the ambitious claims of what is just around the

corner will again disappoint the wild expectations that are being created. The reason why is precisely what allows us to reconnect with our original argument (about what values are).

I don't pretend to know how distant or close we are to producing a true GP AI, as opposed to a computer program that can beat a human player at whatever game with well-defined rules, like Go or Jeopardy!, or at recognizing cats in YouTube videos, but I do know that to get there one piece is conspicuously missing: values. I don't mean "values" in the sense of "program some rules in the putative AI so it does not end maximizing the number of paperclips in the universe and dooming us all" (just Google it) or "ensure Asimov's three laws of robotics are properly implemented". I mean that you can make the program execute as many clever instructions as you wish (you may even make it give itself its own instructions so it gets better and better with each iteration), but you cannot (as of now) make it "care" about following them or not. So, all it can do is execute them (why wouldn't it, not caring one way or the other?) without being any "wiser" (any more conscious) in the process. We are (or will soon be) stuck again in our effort to create an AI (general purpose or not), to give machines a mind, because we truly don't have a clue about how to make them "mind" about the reality that surrounds them or about themselves, about how to *value* differently certain states of the world, and certain representations of such states in their own perceptual apparatus.

Maybe values end up being another way we humans have of categorizing reality, and by giving a computer program enough instances of things we value we may end up having machines which are better than us at valuing things… Maybe, although it is not clear at all how those instances may be assigned a numerical weight that can be processed by an algorithm in an internally consistent way, but I seriously doubt it. There seems to be something about values that doesn't lend itself easily to breakdown into smaller, simpler component parts subject to algorithmization or quantification. The fact that we find not just "states of the world" (how things are), but also our relationship to such states (the emotional coloring we bring to bear when contemplating them, having or not some responsibility in bringing those states about) as inherently more valuable than others seems to be stubbornly resistant to correlation with any particular physical feature of such states.

That difficulty to correlate "what there is" (physical, material, reality in front of our noses, let's say) and how we *value* "it", how we think "it" should be, is something that was already noticed by David Hume in the

XVIII century (who termed it "the is-ought divide", a divide that he was much better at identifying than at bridging) and still salient enough in the XX century for one of the most notable moral philosopher of the age (G. E. Moore, in a legendary work titled *Principia Ethica*) to talk of the "naturalistic fallacy" of trying to assign the goodness of an action (which he explicitly identified with its moral value) with any objectively measurable physical feature of the action, recognizing goodness as a "primitive fact, not subject to further analysis", that is, not subject to being broken down into smaller pieces that could be understood better in isolation.

The defining feature of those evanescent, complex entities we call values is that they are, for those holding them, both *conflicting* (we can value different things, and acting towards the realization of one of them typically involves ignoring, when not downright opposing, the materialization of the rest) and, because of that conflicting nature, *costly*: when we devote our finite, limited resources to the pursuit of one value we hold dear we are sacrificing our energy, our attention and our means. That is indeed the litmus test of our commitment: are we truly willing to do what it takes to bring it about? If we are not, it is likely we don't really value it that much. Maybe it is a second-order value, although not yet something we immediately value itself. Something we would like to appreciate or to have an affinity towards, maybe by training our sensibility and educating our taste, maybe by taking the necessary steps to acquire the knowledge (or the physical disposition and abilities) to be able to enjoy it afterwards, but not yet enjoyable and thus not yet valuable to us at that particular moment. That shows how valuing is inextricably bounded with being conscious, with perceiving an emotional state we recognize as indisputably ours (an indisputability that has been questioned of late by some theoreticians of consciousness, which find that "privileged access" to our own emotions and appraisals of what surrounds us problematic). Let us not forget that "mind" is both a substantive (what the brain creates, or the manifestation of the spirit, as in "he has such a brilliant mind") and a verb ("to mind": to care about, to direct our attention towards something, like in "mind your own business"). Valuing states of the world, being able to order them based on how important they are for us, and thus accepting to sacrifice other alternative states to achieve them, is something that seems strongly to require us being fully conscious of ourselves and of the role we can play in that world as something distinct from it. Something we are very far from being able to replicate in any kind of inanimate substrate.

Which is not necessarily a bad thing. I want to finish this section, after these musings attributing the failure (so far, one has to be humble about what the future may bring) of our attempts to produce thinking machines to the fact that valuing things or situations is not something that we ourselves fully understand, not something we fully know "how it works". Which is surprising, as we are eminently valuing creatures that have created a world of shared values, that devote inordinate amounts of our lives to debating, disputing, arguing about and discussing values, that are willing to lay down our own lives (or to take that of others) for values… but we don't know and cannot articulate what those mysterious values consist of, what is it about them that makes them so important, or how we could conscript machines to help us with assessing and balancing them.

To help young students define for themselves what they find valuable and why is precisely the purpose of this book (or of any book on practical philosophy worth its salt), and what these latest considerations highlight is that such research admits no shortcuts. What great minds in times very different from ours have considered valuable is then the first step, the first piece of input, in determining what values we want to live by, as choosing values is (in some traditions) the foremost responsibility of the will. Towards that first step, that requires we take a look at a time very distant from our own, we turn then to the next section.

Selected Bibliography

Anscombe, Gertrude Elizabeth M. (1958) "Modern Moral Philosophy", *Philosophy,* Vol 33, Nº 124, pp 1-19.

Appiah, Kwame Anthony (2008) *Experiments in Ethics,* Harvard University Press Collingwood, Robin G.

—. (1994) *The Idea of History,* Revised edition, with lectures 1926-1928, edited with an introduction by Jan Van Der Dussen, originally published in 1946

Dancy, Jonathan (2004) *Ethics Without Principles,* Clarendon Press of Oxford University Press

Dworkin, Ronald (2011) *Justice for Hedgehogs,* The Belknap Press of Harvard University Press

Gazzaniga, Michael S. (2005) *The Ethical Brain*, Dana Press

Gibbard, Allan (1990) *Wise Choices, Apt Feelings. A Theory of Normative Judgment*, Clarendon Press of Oxford University Press

Hare, Richard M. (2003) *The Language of Morals,* Clarendon Press of Oxford University Press, originally published in 1952

Harman, Gilbert (1977) *The Nature of Morality. An Introduction to Ethics*, Oxford University Press

Harman, Gilbert & Jarvis Thomson, Judith (1996) *Moral Relativism and Moral Objectivity*, Blackwell Publishing

Joyce, Richard (2006) *The Evolution of Morality,* Massachusetts Institute of Technology Press

Korsgaard, Christine M. (1996) *The Sources of Normativity,* Cambridge University Press

MacIntyre, Alasdair (2007) *A Short History of Ethics. A History of Moral Philosophy from the Homeric Age to the Twentieth Century*, Second Edition, University of Notre Dame Press, originally published in 1966

Mackie, J. L. (1990) *Ethics. Inventing Right and Wrong,* Penguin Books, originally published in 1977

McKeever, Sean & Ridge, Michael (2006) *Principled Ethics. Generalism as a Regulative Ideal*, Clarendon Press of Oxford University Press

Moore, G.E. (1988) *Principia Ethica,* Prometheus Books, originally published in 1902

Nagel, Thomas (1978) *The Possibility of Altruism,* Princeton University Press, originally published in 1970

Nussbaum, Martha C. (2001) *Upheavals of Thought. The Intelligence of Emotions*, Cambridge University Press

O'Neill, Onora (1996) *Towards Justice and Virtue. A Constructive Account of Practical Reasoning*, Cambridge University Press

Prinz, Jesse J. (2007) *The Emotional Construction of Morals*, Oxford University Press

Rawls, John (1999) *A Theory of Justice. Revised Edition,* The Belknap Press of Harvard University Press, first edition originally published in 1977

Raz, Joseph (1999) *Engaging Reason. On the Theory of Value and Action*, Oxford University Press Richardson, Henry S.

(1994) *Practical Reasoning About Final Ends*, Cambridge University Press

Ross, David (2002) *The Right and the Good,* Clarendon Press of Oxford University Press, edited and w an introduction by Philip Stratton-Lake, originally published in 1930

Sen, Amartya (2009) *The Idea of Justice,* The Belknap Press of Harvard University Press