The Emotion Challenge
Towards A Sentimentalist Account of Universal Moral Grammar

Mémoire de Master 2 de sciences cognitives

Sous la direction de
Elisabeth PACHERIE et Pierre JACOB

Institut Jean Nicod
Ecole des Hautes Etudes en Sciences Sociales
Juin 2009
CONTENTS

CONTENTS ............................................................................................................. 2
ACKNOWLEDGMENT ............................................................................................. 3
Introduction ........................................................................................................... 4
1) Preliminary Considerations .................................................................................. 4
2) Should Reason and Emotion Be Part of the Explanation of Moral Judgment? ...... 14
3) The Case for a Sentimentalist Universal Moral Grammar ..................................... 27
Conclusion ............................................................................................................ 32
Bibliography .......................................................................................................... 33
ACKNOWLEDGMENT

I first would like to thank my supervisors, Elisabeth Pacherie and Pierre Jacob, for their help throughout the writing of this paper. I would also like to thank the following people for being kind enough to read previous drafts and give me very useful feedback: Florian Cova, Alberto Masala, Alexandra Marcellesi, and Hady Ba. Last, I would like to thank all the people to whom I have exposed some of the ideas presented here. I hope I have managed to take into account at least some of their criticisms and doubts.
Introduction

The scientific study of moral judgment is generally divided into two main camps which have their counterpart in moral philosophy: the rationalist camp which claims that reasoning is at the basis of our moral judgments, and the sentimentalist camp which claims that emotion best accounts for the production of moral judgment. Recently, a number of theorists (Dwyer, 1999, Hauser, 2006, Mikhail, 2000, 2002, 2007) have argued that both views are mistaken and that our capacity to judge morally must have an innate basis, reasoning and emotion being, though perhaps recruited in its production, not necessary parts of the explanation of moral judgment. Drawing from an analogy between the study of language and the study of morality, they propose that humans possess a moral faculty constituted by innate principles that constrain the range of possible moral systems. According to them, we are indeed endowed with a grammar of action that assesses the causes and consequences as well as intentional structure of morally charged situations, triggering in turn a moral judgment. In this paper, I will first give a general overview of the Universal Moral Grammar (UMG) research program. I will then discuss the model UMG-theorists propose for the explanation of moral judgment production, and their argument for the rejection of reasoning and emotion as significant components of this explanation. In order to assess their model, I will propose a number of criteria that a given psychological process has to meet in order to be part of the explanation of moral judgment. Given these criteria, I will then turn to the study of reasoning and emotion. I will conclude that reasoning should not be part of the explanation of moral judgment, for it both isn’t necessary for the production of moral judgment and has been shown not to play a significant role in the production of most of our moral judgments. However, I will argue that emotion must be part of the explanation of moral judgment, despite UMG-theorists’ contention that it should not. I will conclude this paper by showing how emotion could be integrated into a UMG model of moral judgment.

1) Preliminary Considerations

I) What Should One Be Studying When One Studies Moral Psychology?

People behave wrongly for various reasons. It can be because of money, an unkind environment, certain beliefs, and many other factors that could be intervening at a given
instant. Unusual are cases of murder in which the killer has performed his action without being influenced by situational and historical factors (a troubled youth and a bad day and a tendency to be bothered by others during bad days and a wife not showing understanding, etc—the list can go on for quite a while). On the other end of the spectrum, people may also behave righteously because of various factors. I may help an old lady cross the street because I believe it is my duty to do so. I may also do it, perhaps unconsciously, because I want to impress the pretty lady who is on the opposite sidewalk (see Harman, 1999, Doris, 2004 on the influence of context on moral behavior). Consequently, nothing in my behavior can tell us for sure what causes it, whether it is genuine consideration for others, a self-interested motive, or both. It therefore seems that if we want to engage in a systematic study of moral behavior, we would have to find a way to distinguish between genuine moral behavior and moral behavior based on non-moral factors. In other words, we would have to eliminate the noise involved in morally charged situations, which is probably a case-by-case endeavor.

Psychology as part of cognitive science aims at universals, i.e. at processes that can be found in each normal human being in given circumstances. Moral behavior does not therefore seem to be the right kind of agenda for the kind of moral psychology that is currently done within cognitive science. The moral domain is constituted by many elements acting at various levels (intuitive, judgmental, behavioral…): emotions, beliefs, intuitions, judgments, principles, norms, rules are all more or less theoretical entities and distinctions that can be used when dealing with morality. At the judgmental level, we praise or blame others for some of their actions, and find some actions permissible, impermissible or obligatory. At the behavioral level, we actively engage in morally charged actions, sometimes without realizing it. There are two different aspects of morality here: on the one hand, we are the judge, the one praising or blaming; on the other hand, we are the one who is judged. Given the fact that moral/immoral behavior could never be fully explained by appeal to non-contingent factors, it is tempting to say that morality as a domain of scientific inquiry should be more concerned about the perspective of the judge (i.e., the production of moral judgment) than about the perspective of the judged (i.e., the execution of moral behavior). Nevertheless, this does not necessarily mean that it is vain to engage in the study of factors relevant to moral behavior. Indeed, there is without a doubt a lot of work being done on why and in what circumstances morally charged situations arise. For instance, research on cognitive biases has been able to pinpoint

---

1 I may indeed believe that it is in my interest to be kind to others.
certain recurrent processes involved in conflict creation (see for instance Pronin, 2006). Still, it goes without saying that this type of work cannot have the level of generality that is aimed at in cognitive science. This is analogous to the difference between theory and practice. In theory, I may have to work fifty hours straight in order to finish this paper, but in practice, we all know that this is impossible, given our knowledge about humans. More importantly, this is also analogous to the difference between what Chomsky called competence, the underlying principles and mechanisms of a faculty, and performance, how that faculty is actually effective.

Chomsky used this distinction to highlight the difference between our linguistic capacity, with which we are presumably innately endowed, and our use of it when we speak specific languages (e.g., English, French, Chinese, Arabic, etc.). Linguistic competence is what can be studied from a scientific point of view because it arguably is something that never changes, unlike linguistic performance, which is marked by the variability that we can all notice. Recently, this distinction has been argued to be a useful tool to study morality itself (e.g., Dwyer, 1999, Hauser, 2006, Mikhail, 2000, 2002, 2007). We will see in the next part how the analogy between morality and language has been made. For now, let’s look at the five main questions to be answered by a complete account of morality, questions which are asked by the proponents of this approach. As one can expect, these questions have their counterpart in Chomskyan linguistics:

(1) What constitutes moral knowledge?
(2) How is moral knowledge acquired?
(3) How is moral knowledge put to use?
(4) How is moral knowledge physically realized in the brain?
(5) How did moral knowledge evolve in the species?

Questions (1) and (2) are about moral competence, while (3) is about moral performance/behavior. Questions (4) and (5), though no less interesting, are, according to Hauser and others (Hauser, 2006, Mikhail, 2002), questions that will have to be asked after having answered (1) and (2), since we do not presently have a good definition of what exactly constitutes moral knowledge and how it is acquired. As Mikhail himself puts it, “we cannot profitably ask how moral knowledge evolved in the species or where it resides in the brain until what constitutes moral knowledge and how it is acquired are better understood.” (Mikhail, 2002, 4). Hence, the main goal of UMG-theorists—and the main goal of this
paper—is to explain moral judgment, that is give a precise formulation of the processes that must be at work from the perception of a moral action (input) to the formation of a moral intuition or the expression of a moral judgment (output), for achieving such a goal would give us an answer to (1) and some important constraints for an answer to (2).

II) The Linguistic Analogy

1) Universal Grammar: The Basics

According to Chomsky, humans are endowed with an innate set of rules or principles that enable normal children to acquire a language. ‘Universal Grammar’ (UG), as linguists call it, is the theory about such principles. UG claims that there are three main components to learning a particular language: there are (i) the innate principles that every child is endowed with and that constraints language acquisition, (ii) the environment in which the child happens to be, and its particular language, which triggers (iii) specific parameters in the child’s grammar. In other words, learning a language is permitted by the interaction between the child and her environment, i.e., between innate and learned elements. The innate principles constrain the range of possible languages, the existence of parameters explains linguistic variability, and the environment is there for a specific language to be acquired.

A distinction that is often made within linguistics is between operative and expressed principles. The first are all the underlying principles driving linguistic competence; they operate below the conscious level, and are hard to be discovered through introspection. It takes many years of studying language to come up with principles that are close to those of UG. The second are all the ‘grammatical rules’ as used in the everyday sense. They are (or could be) conscious, and drive our linguistic performance, i.e., the particular way we use language in everyday life. Operative principles are universal, while expressed principles need not be. Why posit such operative principles? Because of the simple fact that when we ask people whether a sentence like “Colorless green ideas sleep furiously” is grammatical, they answer ‘yes’ without being able to tell why, regardless of who they are and where they come from. The linguist’s job is therefore to uncover the principles underlying people’s acceptance of certain kinds of sentence and not others, i.e., find out what kind of knowledge constitutes UG. As Hauser et al. (2008) nicely put it, “When it comes to language, what we think we know pales in relation to what our minds actually know.” (109)
A useful distinction to be made here is between (1) UG, (2) I-language (‘Internal language’), and (3) E-language (‘External language’). (1) is the set of principles at the basis of language acquisition, (3) is the public expression of a particular language, and (2) is mainly a state at a given developmental time whose “growth’ [is] determined by the interaction between the individual’s innate endowment of UG and the linguistic environment to which she is exposed” (Dwyer, forthcoming, 11). A way to study (1) has been to study (2), for the latter clearly is evidence for the former—at least this is what ‘biolinguists’ have concluded (Dwyer, ibid). It is not surprising, then, to realize that the main methodology used by linguists has been studying people’s intuitions about particular sentences, i.e., whether they judge them as acceptable or not. Note that we should here distinguish between grammaticality judgments and acceptability judgments (Dwyer, ibid). When facing a particular sentence, people do not judge whether it is grammatical; rather, they judge whether it is acceptable to utter it. Grammaticality is a theory-dependent notion that has nothing to do with what is going on when people face a particular sentence. Indeed, they judge whether they could utter it, not whether it accords with any principles. The principles are part of the linguist’s theoretical suitcase, and are not in any way what people have in mind at the moment in question. In other words, people do not “bear any epistemic relation to whatever rules and principles a theoretical linguist might find it helpful to reference in an articulation of a given signal’s grammatical status. (...) And, crucially, an I-language is a state of a speaker’s mind/brain; it is not something that could be the object of her knowledge or belief, tacit or otherwise” (Dwyer, ibid, 14). An I-language is what causes acceptability judgments, but is not itself the object of the judgment.

The main argument that has been provided in favor of UG is the so-called ‘argument for the poverty of stimulus’ (PoS), supposed to answer what has been called the ‘perception problem’, that is the problem of knowing how children acquire the ability to discriminate between linguistic stimuli and other kinds of stimuli through experience. This argument claims that a child could not learn her language solely based on experience, for the nature of the available data would render language acquisition impossible for an individual without the kind of endowment postulated by UG. In particular, the quasi-absence of negative data (i.e., data about what can’t be said) is evidence for the thesis according to which there must be something to the human mind that still enables children to perform the relevant kind of discriminations. Although this argument has been subjected to many critiques (see for instance Pullum & Scholz, 2002), we will here assume that it is correct.
2) The Case for Moral Grammar

In *A Theory of Justice* (1971), John Rawls writes:

> A useful comparison here is with the problem of describing the sense of grammaticalness that we have for the sentences of our native language. In this case the aim is to characterize the ability to recognize well-formed sentences by formulating clearly expressed principles which make the same discriminations as the native speaker. This is a difficult undertaking which, although still unfinished, is known to require theoretical constructions that far outrun the ad hoc precepts of our explicit grammatical knowledge. A similar situation presumably holds in moral philosophy. (47)

Following Rawls’ analogy, a number of theorists have used the research program within generative linguistics as a tool, a heuristics for studying our moral capacity (Dwyer, 1999, Hauser, 2006, Mikhail, 2000, 2002, 2007). We have seen earlier that we can also distinguish between competence and performance in the case of morality. In this section, we will cover two main arguments in favor of what has been called ‘Universal Moral Grammar’ (UMG). The first argument is the argument for the existence of a set of moral principles, a ‘moral grammar’, operating automatically below the conscious level. The second argument tells us that a child could not possibly acquire her moral faculty through experience alone.

*The Argument for a Moral Grammar*

How can we determine that unconscious principles are operating when we are judging morally? The only way to discover such principles is, it seems, to make people produce moral judgments and see whether they know by virtue of what principles such judgments were produced. Recent work on moral intuitions appears to lead to the conclusion that there indeed are unconscious principles at the basis of at least some of our moral intuitions, which suggests that reasoning and justification need not be at work in moral judgment production. For instance, Haidt (2001) has provided evidence for the thesis that people generally do not have access to what principles direct their judgments, and in cases in which they happen to be aware of this limitation, they become extremely dumbfounded—hence the name of this effect, ‘moral dumbfounding’—by their inability to change their mind. Haidt therefore suggests that moral intuitions are generated by rapid, automatic, effortless, emotional processes that he calls ‘intuitions’. The emotional component of this mechanism seems to be confirmed by
neuroimaging studies (e.g., Greene et al., 2001, Greene & Haidt, 2002) and in studies of patients (Ciaramelli et al., 2007, Greene, 2007, Koenigs et al., 2007, Mendez et al., 2005). We can conclude from such studies that we have strong evidence for the existence of unconscious processes that may reliably be translated into distinct principles. Nevertheless, the existence of unconscious principles is neither evidence that they are innate, nor evidence that they are part of a ‘universal grammar’ of the sort postulated by UMG-theorists. Other evidence is needed in order to draw such a conclusion. So far, the studies mentioned show that there is a clear dissociation between judgment, mainly intuitive, and justification, characterized by conscious reasoning (Hauser et al., 2007). People know that P is wrong, but they do not know why, or at least they do not know why in the transparent way that is assumed, it seems to me, by reasoning-based—or so-called ‘rationalist’—theories of moral judgment (e.g., Piaget, 1932, Kohlberg, 1969). This effect (i.e., not knowing where our moral principles come from), and the patterns of answers to scenarios, are found across cultures, which is prima facie evidence for the innateness thesis.

What we need to know now is the precise nature of the processes involved in people’s moral judgments, if it is not reasoning. Emotion seems to best explain our production of moral judgments (Nichols, 2002, 2004, Prinz, 2006, 2007, Greene et al., 2001). However, unless we know exactly what role emotion plays in the process and when, we do not have reason to say that emotion constitutes the whole explanation (though see Prinz, 2007). Let’s see how UMG-theorists argue for their particular position against the available models. The argument they make seems to have the following structure: there is a pattern in people’s responses to scenarios that cannot be accounted for by the available models, and that can be accounted for by UMG. This argument is weak, for, as we will see, it is possible to come up with alternative hypotheses that account for the pattern just as well. In order to understand the pattern in question, let’s consider two of the scenarios that were given to participants in a cross-cultural study (Hauser et al., 2007):

Ned is taking his daily walks near the train tracks when he notices that the train that is approaching is out of control. Ned sees what happened: the driver of the train saw five men walking across the tracks and slammed on the brakes, but the brakes failed and they will not be able to get off the tracks in time. Fortunately, Ned is standing next to a switch, which he can throw, that will temporarily turn the train onto a side track. There is a heavy object on the side track. If the train hits the object, the object will slow the train down, thereby giving the men time to escape. Unfortunately, the heavy object is a man, standing on the side track with his back turned. Ned can throw the switch, preventing the train from killing the men, but killing the man. Or he can refrain from doing this, letting the five die.

Is it morally permissible for Ned to throw the switch?
Oscar is taking his daily walk near the train tracks when he notices that the train that is approaching is out of control. Oscar sees what has happened: the driver of the train saw five men walking across the tracks and slammed on the brakes, but the brakes failed and the driver fainted. The train is now rushing toward the five men. It is moving so fast that they will not be able to get off the track in time. Fortunately, Oscar is standing next to a switch, which he can throw, that will temporarily turn the train onto a side track. There is a heavy object on the side track. If the train hits the object, the object will slow the train down, thereby giving the men time to escape. Unfortunately, there is a man standing on the side track in front of the heavy object, with his back turned. Oscar can throw the switch, preventing the train from killing the men, but killing the man. Or he can refrain from doing this, letting the five die.

Is it morally permissible for Oscar to throw the switch?

People are more inclined to answer ‘yes’ to the question in the second scenario than to answer ‘yes’ to the question in the first scenario, which suggests that the emotional valence of the action in question (for instance, see Greene’s, 2001, distinction between personal and impersonal actions) is not necessarily what is causally responsible for people’s moral judgments. Here, in contrast, it seems that the particular causal structure of the situation is crucially responsible for the difference between the two patterns of answers. This suggests that a capacity that evaluates the causes and consequences of action, along with its intentional structure, is both causally responsible for at least some of our moral judgments and at the basis of some of the principles underlying our moral intuitions. In Hauser et al.’s pattern of response, for instance, the principle that seems to be at work is what has been dubbed by philosophers the ‘principle of double effect’ (PDD), which can roughly be expressed in the following way: “it may be permissible to harm an individual for the greater good if the harm is not the necessary means to the greater good but, rather, merely a foreseen side effect” (Hauser et al., 2007, 3).

The advantage of postulating UMG now seems more limpid. First, we have reasons to say that there are unconscious processes that are necessary for the production of moral judgment. Second, we have reasons to say that these processes are translatable into distinct (abstract) principles. Third, with the prima facie universality of PDD, we have at least some reason to believe that some principles are innate. Fourth, with the structural aspect of moral judgment production, we have reasons to use the idea of a moral grammar in the explanation of our moral judgments.

The Argument for the Poverty of the Moral Stimulus

The second argument put forth in favor of UMG is the argument according to which, since it is implausible that a child would acquire her moral knowledge through experience alone, there must be an innate set of principles that fills in the gap between the stimuli the child is in
contact with and her mature moral faculty. The use of PoS in the case of morality has been advanced as an answer to the same kind of perception problem that was raised with language. How does the child come to know the moral norms of her society? Does she start from scratch, taught by her parents that doing so is right and doing so is wrong? Does she infer from the reactions of others that some particular behaviors are morally permissible and others impermissible? How come she can distinguish between moral norms and other kinds of norms such as the norms of etiquette? Overall, is it possible that she acquire her moral norms through experience alone?

Although it may be possible to imagine an (almost) entirely empiricist story (e.g., Prinz, 2007, 2008), nothing seems to provide compelling reason supporting the thesis that experience alone is sufficient for the acquisition of a moral faculty. Indeed, in order for an empiricist account to work out, it would have to tell us that there are experiences that would be both necessary and sufficient for a child to acquire her moral faculty, since in order for the latter to attain the level of specificity that it appears to display, it would have to be created by a number of clearly defined and not contingent (or accidental) processes. If we find out that every child (of a given culture at least) is taught explicitly in a specific way (by means of ‘ought’ statements, for instance), that every child witnesses the exact same emotional reactions to what will, in the future, be for her morally charged situations, or that every child is punished in a specific way every time she does so and so, etc., then we would have strong evidence for the empiricist story, since we would find the same kind of processes in every child.

However, no strong evidence of this sort seems to be available to the empiricist, or so UMG-theorists argue. For instance, Prinz (2007, 2008) has proposed that the acquisition of the distinction between moral norms and conventional norms is caused by differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at. As Dwyer (2008) has highlighted, there is too much variability between caretakers for the differentiated emotional reactions from caretakers. In spite of its apparent appeal, this proposal fails to secure the kind of generality that a plausible empiricist approach should aim at.

\[2\text{ For an presentation of the moral/conventional distinction, see Turiel (1983)}\]
Explicit moral instruction is real, Dwyer contends, but it is not in any way a universal phenomenon. After all, many of the moral norms we respect were never taught explicitly to us, and even when they were, they were not taught very differently from conventional norms (the word ‘ought’ is used in both kinds of teaching). But, how, then, did we acquire the ability to discriminate between moral norms and conventional norms? The UMG-theorist’s answer is that there must be something to the child’s mind/brain that contributes to her acquisition of moral norms in various environments, and given the variability of teaching styles. To counter this thesis, the empiricist would first have to propose an alternative explanation in which the processes of acquisition are not just found in some cases, but rather are found in all cases of normal moral agents. She would then have to argue that such processes are sufficient for the acquisition of a moral faculty. According to UMG-theorists, no viable alternative proposal will be found. In this paper, I will not try to argue for an alternative empiricist story of the moral faculty, or argue against the research program’s assumption (see Prinz, 2008, for a complete critique). Rather, I will assess the explanation that UMG-theorists give for the production of moral judgment, and will argue that this explanation is not complete and that another ingredient, emotion, must be included.

III) Today’s Agenda: Explaining Moral Judgment

Earlier, I have argued that the idea of a universal moral grammar is not absurd and should be taken seriously. There are, however, several ways we can approach this hypothesis. First, we can think of the linguistic analogy as claiming that the moral faculty should be taken as the exact analog of the linguistic faculty, or at least understood within the same framework, that of the one the generative linguistics research program. This strong interpretation of UMG (‘strong-UMG’) entails the existence of a number of necessary conditions for moral judgment production: (1) UMG must be modular (with domain-specific components), (2) it must have a grammatical structure, (3) it must be compositional, and (4) it must be productive. I will not understand UMG in this strong sense in this paper. Rather, I will approach it in a weaker sense (‘weak-UMG’) according to which the analogy between the moral faculty and the linguistic faculty should be taken as a mere heuristics used to understand morality, in the sense that advances in generative linguistics may be of some theoretical and conceptual use for the study of moral judgment. In this sense, a perfect analogy is not assumed, and differences between the two types of faculties would not necessarily undermine the entire research program. This approach will enable me to put forward a new model of the moral
faculty that does not annihilate the idea of a universal moral grammar. Indeed, I think that the explanation UMG-theorists give for the production of moral judgment might be compatible with rival models, and that a theory combining them may account for moral judgment as well, or perhaps even better, as I shall argue.

Explaining the process leading to moral judgment amounts to giving a precise account of moral competence (or moral faculty). Besides the set of innate principles that are argued to be essential for the acquisition and performance of morality, i.e. the grammar of action that is at work in moral judgment, what kind of processes can we consider as essential to the moral faculty? It seems that a good way to determine whether a process is part of moral competence is to determine whether it is specific to it, which seems to be an important methodology within nativist research programs. However, as we will see, a better methodology may be to determine what kinds of processes are necessary for moral judgment, for even UMG-theorists (Hauser, 2006) have granted that some mechanisms involved in moral judgment, such as action analysis, are not in themselves specific to morality, although they are necessary for moral judgment to be produced. It is indeed commonplace in the generative tradition to distinguish two questions that must be raised about a given mechanism in terms of which we seek to explain a particular faculty:

1) Is the mechanism necessary for the faculty, or is it optional (i.e., what the faculty produces can be produced without this mechanism)?

2) Is it unique to the faculty, or is it shared?

In the case of language, Hauser et al. (2008) conclude, “[w]e can see, then, that the faculty of language is comprised of several different types of cognitive mechanisms: those that are unique versus those that are shared and those that are necessary versus those that are optionally recruited.” (110)

In the next section, I will deal with two main candidate processes that have been argued to be necessary for moral judgment—reasoning and emotions—and that have been at the basis of two influential models of moral judgment—rationalism and sentimentalism—and assess how UMG-theorists have explained their occurrence.

2) Should Reason and Emotion Be Part of the Explanation of Moral Judgment?

I) Two Conditions
What Do We Need To Look For?

At the end of the previous section, we made two important distinctions between processes that are part of a given faculty:

1. Processes that are unique to the faculty (faculty-specificity) vs. processes that are shared with other faculties (faculty-generality)

2. Processes that are necessary for the faculty vs. processes that are optional

Of course, the elements of (1) can be combined with the elements of (2): faculty-specific processes can either be necessary or optional and faculty-general processes can either be necessary or optional. Given the fact that what we seek to do here is providing a general theory of moral judgment (see section 1 part I), i.e. a theory that can account for cases of moral judgment produced by normal adults, it seems that we first would have to be concerned with what is necessary for moral judgment. Indeed, any theory of moral judgment must start with the question of what it takes for a moral judgment to be produced by normal adult humans, and then, having an answer to this question, we may seek to determine if any of the processes necessarily involved in moral judgment are specific to the moral faculty. In other words, there is a clear difference between (i) explaining moral judgment (i.e., finding all the processes that are necessarily involved in its production), (ii) discovering what are the processes involved in moral judgment that are unique to morality, and (iii) determining what processes that are specific to the moral faculty are optional. (ii) is too ambitious to be our primary goal, for it relies on a precise answer to (i), and (iii) is even more ambitious in that we would need an answer to both (i) and (ii) in order to start asking it. At this time, it seems that the debate over the problem of the production of moral judgment revolves around the first question, so we need not concern ourselves with the other two questions in the rest of this paper.

We now can provide a criterion that any process must satisfy in order to count as part of the explanation of moral judgment:

(a) **Necessity condition:** For a process X to constitute part of the explanation of a psychological phenomenon or behavior Y, X must occur for Y to occur.³

³ I need not concern myself here with problems of multiple realizability, although they are problems that will need to be faced in future research.
However, this condition is problematic, for it accepts too much. Many mechanisms are necessary for moral judgment, and some of them are already known: mechanisms of perception (a person in a coma could not make a judgment about anything), the processing of information (a person without a working memory could not possibly remember what is going on), the pumping of blood by the heart (a dead person cannot do anything), and so on. Appealing to such processes is not really explaining moral judgment but expressing the background conditions that make it possible. What we need to determine are all the mechanisms that are necessary for moral judgment in a relevant way. Condition (a) is therefore too general to provide by itself the relevant constraints for a good explanation of moral judgment.

Methodological Chunking

The explanation that UMG-theorists give for the production of moral judgment is based on a computational problem: we have some inputs (an observable action), something is going on in the head (the computations that we seek to determine), and then we have an output (a moral judgment). We have seen in the previous section that, according to UMG-theorists, there must be something already in the brain for a moral faculty to be possible (this is the PoS argument). They propose that, as in the case of language, there is a moral grammar constituted by principles that is at the basis of morality. As a result, they explain moral judgment by appeal to a mechanism (action analysis) that assesses the causes and consequences as well as intentional structure of the action. But can action analysis (coupled with moral grammar) account by itself for the production of moral judgment?

To answer this question, we first need to specify the computational tasks that would have to be explained in order for moral judgment to be accounted for. This is what I call ‘methodological chunking’: we derive from the input and the output of a faculty all the particular tasks that must occur in between, given a set of assumptions and observations, which will in turn have to be explained. For instance, UMG-theorists have put forward action analysis as the best available candidate for the explanation of moral judgment because it seems to best explain certain tasks that would have to be performed for moral judgment to be produced. When I perceive someone being hurt by someone else, there must be something in my mental representation of the action (other than the mere representation of colors and

4 Note that these conditions need not be biological. The existence of the world, and of its processes, is, albeit physical, necessary for moral judgment.
shapes) that makes me see it as wrong: a causal process involving somebody’s intention to harm and a consequence, the actual harming of somebody else. Assuming PoS, this process can be explained by the working of an action analysis mechanism that triggers a particular norm in my moral grammar.\(^5\)

Nothing in this description tells us that I must have eyes, a heart, or that the world must exists, in order to form a moral judgment. Instead, it shows that very specific, non trivial (or relevant), tasks must be performed in moral judgment production, which gives us important constraints in our theoretical endeavors. These tasks are non trivial in the sense that, if we discovered what would make them happen (what processes perform them), this would have a significant impact on our understanding of moral judgment; for instance, it would enable us to make clear predictions in cases of selective cognitive impairments. Of course, the tasks involved are hypothetical constructs that can be revised or rejected if our observations do not support their existence. However, as we will see, it is possible to discover that certain tasks are hard, perhaps impossible, to put aside, and that they must be (by definition) somehow accounted for by a very narrow range of possible processes. For now, I wish to introduce a second condition that a cognitive process has to fulfill in order to be part of the explanation of moral judgment:

(b) **Relevancy condition:** A process X is relevant if it can perform a task Z that has been identified as a necessary step for the understanding of a psychological phenomenon or behavior Y.

This last condition is minimal in the sense that it amounts to saying that we should seek to find processes that may perform the tasks that have been determined by the theoretician to be relevant for the explanation of the phenomenon she is interested in. In other words, a process is relevant if it performs a task that we deem relevant.

Having set the two conditions that must be met by a given process to be included in the explanation of moral judgment (the necessity and relevancy conditions), I now turn to two of the processes that have been argued to explain moral judgment, and that constitute the main opponents to UMG.

---

\(^5\) Dwyer (2008) makes a distinction between moral *principles*, which are the principles that are involved in the acquisition of one’s moral grammar, and moral *norms*, which are the principles that are part of the acquired moral grammar. In a sense, moral principles are innate while moral norms are not.
II) Reasoning

Some theorists (Dupoux & Jacob, 2007, Sterelny, 2008) have recently argued that, since reasoning and explicit moral beliefs are not, according to UMG-theorists, part of the explanation of moral judgment, UMG is ill suited to account for moral judgment. It indeed seems that, in everyday life, many of our moral judgments are not just intuitive and automatic processes followed by eventual post-hoc rationalizations, but are rather based on some kind of reasoning process. After reading books on the mistreatment of animals, I may come to judge that it is wrong to eat meat. I may also come to believe that active and passive euthanasia are equally right, thus ignoring the act/omission distinction, another possible principle for UMG. Like a linguistic grammar, a moral grammar is presumably fixed when it comes to maturity, and no explicit belief should influence people’s acceptability judgments. However, this does not seem to be the case. Explicit beliefs and justifications are often taken as inputs in processes of deliberation leading to moral judgment, and can even help us to adjudicate between conflicting intuitions. Background information affects moral evaluation, which shows that moral grammar is not “encapsulated in Fodor’s sense” (Dupoux & Jacob, 2007).

At any rate, moral reasoning, it seems, cannot be reduced to an epiphenomenal process. So how could the UMG-theorist account for its existence? Several possibilities are open to the UMG-theorist. First, she could argue that moral reasoning is always a post hoc activity. Needless to say, virtually nobody would argue such a thing. Second, she could admit that some moral reasoning is at work in the production of some moral judgment, and that reasoning is important for morality at broad, but that this activity, though real, is not essential to moral judgment production, making it in turn an epiphenomenal process in the case of moral judgment, but not in the moral domain at broad. One can indeed imagine a moral creature that has never come to reason over or find justification for her moral intuitions. Her moral judgments are full-fledged moral judgments without ever being caused or influenced by explicit moral beliefs and justifications. Although her moral life may not be as broad and flexible as ours, she would still be considered as a moral creature, and the capacity to discriminate between moral actions is at the very least the capacity that would explain why.

Moreover, UMG-theorists do not claim that the unconscious principles of UMG are all there is to moral cognition. As Dwyer and Hauser (2007) say, “[t]he fact that part of our moral psychology depends on explicit beliefs is not counterevidence against an intuitive, unconscious component, viz. a moral grammar” (1). Indeed, what is important here is not the specific content of moral judgments and beliefs, but their abstract structure. UMG-theorists
are interested in the latter, even though the former is obviously important for a complete account of moral judgment. “It is the abstract structure of these statements, as opposed to their content, that carries the signature of the [moral] faculty” (ibid).

Still, in order to convince us that moral reasoning should not be taken as part of the explanation of moral judgment, UMG-theorists need to provide a positive account of its role in moral cognition. One tentative proposal is that moral reasoning, like reasoning about beliefs, is a social enterprise: it is one of the tools we use to convince others that our beliefs are true. This is probably what Haidt’s (2001) study is supposed to show: in many cases, “when faced with a social demand for a verbal justification, one becomes a lawyer trying to build a case, rather than a judge searching for the truth.” However, as we have seen, moral reasoning is sometimes more than just post hoc rationalization. The very existence of moral change, defined by the rejection of old moral beliefs and the acquisition of new ones after, among other things, a process of deliberation (e.g., reflecting on the cruel treatment of animals), is evidence that moral reasoning is probably more than that. In the case of dilemmas, for instance, as Dupoux and Jacob (2007) have rightly noticed, moral reasoning plays a crucial role in that it helps us adjudicate between the various options offered to us. Moreover, moral change, though not systematic, is made possible by the conscious articulation of beliefs (moral or otherwise), desires, and justifications. What can the UMG-theorist say about such phenomena?

The UMG-theorist need not deny the role explicit reasoning plays in our moral life in general, and the instantiation of certain of our moral judgments in particular. Neither does she have to put explicit reasoning on the side of moral performance, though it may certainly be deeply important for moral behavior. Reasoning is presumably a general-purpose capacity that enables us, among other things, to articulate the logical structure of propositions or beliefs and to make inferences. Its role in moral cognition may be just that, to help us gather knowledge about the world, knowledge about our moral beliefs, and knowledge about the various contexts we may face at any given moment. So, yes, explicit reasoning is an important component of moral cognition. But it may be neither specific to it nor necessary. It may rather be recruited ‘when needed’. A future line of research would therefore be to determine exactly when (i.e., in what circumstances) reasoning comes into play in the production of moral judgment.

Before turning to emotions, I would like to point out a further ‘reasoning’ problem for the UMG-theorist. If adult moral competence is fixed, then why is it still possible, by means of metacognitive processes such as reasoning, to alter or ignore some of the innate principles
that constitute it? Why is it possible that, after reflection, I decide that active euthanasia is not worse than passive euthanasia, and may even be the better option, transgressing in turn the act/omission distinction? A possibility is that the act/omission distinction is not part of UMG. But this is a dubious solution, for the distinction appears as well suited to be part of UMG as is the principle of double effect. Another more plausible possible explanation is that the act/omission distinction is still part of our moral competence after such a change in moral beliefs, but the context of performance makes it possible for the distinction to be ignored. Moreover, we may even find it emotionally costly to transgress such a norm, even though we know that there is nothing reasonable about it in the present situation. This is an empirical matter that will need to be dealt with. One last possible explanation is that the situation has been reframed in such a way that the act/omission distinction is not operative anymore. Perhaps focusing on the suffering of the victim is enough for us not to ‘care’ about the act/omission distinction. We could also hypothesize that there is another potential principle of UMG that the situation is calling for, such as ‘A long time of suffering is worse than a short time of suffering’, and that focusing on the relevant features of the situation may make us respond to it differently. Although this is all speculation, this gives us reasons not to reject UMG in its current version, even though the analogy with the linguistic faculty appears weaker than expected.

III) Emotions

1) Does Science Show that Emotion is Necessary for Moral Judgment?

In the past few years, a considerable number of studies have highlighted the importance of emotion for morality. Many moral psychologists and philosophers have come to believe that emotion is somehow at the basis of morality in general, and moral judgment in particular (e.g., Nichols, 2002, 2004, Prinz, 2007). Indeed, facing the overwhelming evidence about the emotional dimension of moral judgment, such researchers have argued that our emotional system is necessary for its production, or at least necessary for the production of many of our moral judgments. In this part, I will examine these claims, and see whether emotion can be said to be part of the explanation of moral judgment.

First, let’s see whether emotion could be sufficient for moral judgment, i.e., whether there is evidence suggesting that an emotion alone could trigger a moral judgment. It seems that we
have such evidence. In an experiment, Wheatley and Haidt (2005) gave subjects, who had had been hypnotized to feel disgust every time they would see the words ‘often’ and ‘take’, vignettes displaying some cases of moral transgression and moral ‘excellence’, and asked them to rate the actions at hand. Participants experiencing a flash of disgust when seeing the target-words in the vignettes not only rated the moral transgressions more severely but also surprisingly rated the morally admirable acts as morally wrong, which indicates that “moral judgments may be grounded in affectively laden moral intuitions” (780). In other words, “[t]his suggests that a negative feeling can give rise to a negative moral appraisal without any specific belief about some property in virtue of which something is wrong” (Prinz 2006, 31).

This effect is evidence that disgust can somehow modulate moral judgment (see also Haidt, 2001, Nichols, 2004, Schnall et al., 2008, Yan, 2008).

Now we may wonder whether the apparent sufficiency of emotion for moral judgment shows anything about its role other than the fact that it can direct, or perhaps alter, our attention so much that we are inclined to judge actions morally even though they are not really moral or immoral actions. Because something is disgusting, we may be inclined to judge it as wrong by looking at certain of its features and not others. Emotion, especially negative emotion, has the power of narrowing our attention, and so could be recruited as a ‘funnel’ in order for us to apply the right kinds of principles to the right kinds of contextual features and their relationships. Something more is needed. We also need to determine whether emotion could ever be said to be necessary (in a relevant way) for moral judgment.

Let’s summarize some of the findings that have been gathered in favor of the thesis that emotion is causally responsible for our moral judgments. An affective mechanism is necessary for normal social behavior and decision making (Damasio, 1994). An individual lacking certain emotions (i.e., a psychopath) is virtually a bad moral reasoner (Nichols, 2002, 2004). People sometimes make moral judgments based on their intuitions/emotions/’gut feelings’ and are unable to justify convincingly their response (Haidt, 2001). Emotion alone (such as disgust) can trigger moral judgment (Wheatley & Haidt, 2005). An activation of the affective system in the brain co-occurs with many, maybe all, of our moral judgments, even those which are triggered during moral dilemmas (Greene et al., 2001) (see Prinz, 2006 for a review). Facing all these findings (and others), we may reasonably believe that emotion actually plays a crucial role in moral judgment. But can we infer from them that emotion is causally necessary for moral judgment? UMG-theorists seem to believe not. I think they are right on this point.
The UMG-theorists are right when they say that the available neurobiological data that have been gathered in favor of the necessity thesis is insufficient because “the activity of emotional circuits provides only correlational data, showing that emotions are associated with moral judgments.” (Huebner et al., 2008). Correlation does not imply causation; neither does it imply necessity. Furthermore, the available data “(on their own) can never be used to infer causality, and because of the poor temporal resolution of neuroimaging, cannot be used to assess when emotions have a role or whether they are constitutive of moral concepts.” (ibid)

By the same token, studies of patients (VM patients and psychopaths) have not, as suggested by many, really proved that emotion is causally responsible for moral judgment. Take the psychopath, for instance. The psychopath has been shown to lack the capacity to distinguish between the moral and the conventional, and the usual explanation has been that the lack of certain emotions (the so-called ‘moral emotions’) is the direct cause of this deficit. However, another equally plausible hypothesis has been recently made available by UMG-theorists (Huebner et al., 2008). According to them, emotions are primarily involved in moral action. Without emotion, we may indeed not be motivated to act morally. If the psychopath lacks certain emotions that are important for moral action (doing morally praiseworthy actions as well as not doing morally blameworthy actions), then he may more easily engage in anti-social behavior. And anti-social behavior has been argued to play a role in the development of our moral psychology (Raine and Yang, 2006). The more we engage in morally blameworthy behavior, the less we are compelled by the supremacy of the norms involved, and the less we may discriminate between different kinds of normative domains (moral vs. conventional). It is therefore possible that, at the beginning, children with psychopathic tendencies could distinguish between the moral and the conventional, but that, after recurrent acts of violence, their capacity to make such a distinction disappears. Huebner et al. (2008) conclude that “if this is true, the moral cognition of psychopaths is deviant as a result of deviant inputs rather than as a result of a deficiency in moral processing.” (5) At the very least, therefore, this shows that the necessity thesis is not as obvious as many have thought. In order to argue for the necessity thesis, we may need to adopt another strategy.

2) Emotion as Output
The strategy I will adopt is not empirical, but philosophical. I am going to put forward a tentative proposal regarding the specific role that emotion could play. My point here is that, even though the data do not show by themselves that emotion is necessary for moral
judgment, we can still argue that emotion must play some kind of role, given the particular nature of moral judgment.

A problem for the UMG-theorist is to account for the value-laden dimension of moral judgment. Indeed, unlike language, morality is not only about what should be the case (how the world is supposed to be, given the principles we have), but also about what ought to be the case (i.e., the absence of human suffering, the presence of helping behavior, etc.). Language is principle-based. Morality is principle-based and value-laden. Values cannot be comprehended by mere rationality. If I like chocolate, I value its taste. If I love someone, I value this person. If I want the world to be a better place, I value the absence of suffering. Morality is about killing, raping, violence, incest, help, harm, charity, and so on. All these things are value-laden, and must somehow be comprehended as value-laden. Along with many philosophers (e.g., D’Arms & Jacobson, 2000, De Sousa, 2001, Johnston, 2001a, 2001b, Tappolet, 2000), I think that this comprehension could only be made possible by the presence of emotion. Emotion must therefore be central to moral cognition.

Although it is far from obvious that emotion should be explained by appeal to axiological properties (e.g., Teroni, 2007), it seems less contentious to ground values on human valuing, i.e. on the capacity for humans to apprehend the world in axiological terms. Indeed, while it may be possible to imagine an account of emotion that doesn’t appeal to any axiological dimension, it is hard to imagine an account of value that doesn’t appeal to certain human psychological responses such as emotions, affects, and other sentiments. This of course doesn’t mean that certain values cannot be acknowledged as values without the presence of an adequate emotional disposition—an anthropologist studying another culture indeed has an access, albeit external, to the natives’ values—but rather it means that a value is not a value for us without the presence of an adequate emotional dispositional. In other words, the individuation of a particular value is made possible by the presence of an adequate emotional disposition, and the explanation of this individuation cannot be carried out without appealing to the relevant disposition: it is because the natives have emotional reactions (or dispositions) towards certain facts of the world that the anthropologist can attribute the possession of values to them. In short, axiological properties cannot, it seems, be properly conceived without the use of emotion concepts, for the apparent reason that values can only be comprehended by emotion.

Even though there is a clear distinction between values and norms—the second, not the first, are about what is permissible, obligatory, and forbidden (Tappolet, 2000)—there are strong reasons to assume that normative concepts depend, at least in part, on axiological concepts.
(Ogien & Tappolet, 2008). First, both kinds of concepts differ from natural concepts (i.e., concepts about the natural-physical world, or concepts that are clearly reducible to them). They are the kind of concepts that are studied in so-called ‘value theory’ courses (mainly ethics and aesthetics). Second, it seems that norms always have background elements that correspond to or could be transposed into values. The norm ‘Do not hurt anyone’ may be seen (besides as the expression of an obligation) as the expression of a particular value, that is the absence of suffering. The normative domain is intrinsically value-laden (see also McGinn, 1997, Pettit, 1991), and therefore, if what I have said so far is right, it cannot be fully comprehended without emotion.

What could be the nature of this comprehension? If we establish that there are objective values—if we are realists about values—the conclusion drawn from my argument should seem natural. Indeed, if there are values that are independent of us, or at least features in the world that we are disposed to regard as value-laden (like secondary qualities; see McDowell, 1985), we would need to determine the means by which such a perception could be made possible (perhaps with a distinct moral sense). Emotion, by contrast with mere belief, would presumably be the best available answer (see Tappolet, 2000 for a defense of this thesis). However, we do not have to be realists about values in order for my ‘response-dependence’ thesis to be true. If we accept something along the lines of the Fregean distinction between sense and reference, and assuming that it is possible for an expression to have a sense but no actual reference, I propose the following argument:

(1) (i) **Moral illusion thesis**: If realism is false or unacceptable, moral terms may have a sense, but no reference, *even if we believe that they do have a reference*;

Or (ii) **Moral agnosticism**: We can be agnostic about the existence of moral properties, given the fact that we can understand the sense of a moral term without knowing (or being able to specify) its reference, if it has one;

(2) **Trivial premise**: ‘Right’ and ‘wrong’ are moral/value-laden terms that have a sense;

---

6 Although we value the absence of suffering, there are of course cases in which making one suffer is acceptable (for instance at the dentist). Expressed norms are therefore sometimes too general to account for the kind of principles that are part of UMG.

7 We can distinguish here two kinds of moral realism, a minimal version, claiming that moral properties, though truth-apt, can still be dependent somehow on the working of our mind, and a stronger one, claiming that moral properties do not depend on the working of the mind (they are ‘there’ even when nobody can comprehend them). Here I admit the plausibility of the minimal version of moral realism (see Sayre-McCord, 1986, for such a view), while assume the stronger one is not tenable. However, this depends on what we mean by ‘objective’ and ‘mind-independent’ moral facts (see Joyce, 2007 for a discussion).
(3) **Conceptual claim:** It is impossible to grasp moral/value-laden terms without cognitive processes that are capable of performing such a task;

(4) **Empirical/conceptual claim:** Emotions (or feelings or sentiments) are the only cognitive processes that we possess that are capable of enabling us to grasp moral/value-laden terms (since, *inter alia*, they are the only processes that we could clearly call ‘positive’ or ‘negative’);

(5) **Conclusion 1:** Therefore, ‘right’ and ‘wrong’ cannot be grasped without emotion.

(6) **Conclusion 2:** The endorsement of a moral norm is impossible without emotion.

The same argument applies to any axiological concept: admirable, delicious, beautiful, and so on. The use of ‘emotion’ here is broad enough to admit both processes that are not strictly emotional but somehow connected to emotion (e.g., desires) and processes in which emotion plays a crucial role (e.g., motivation).

This is a modest argument in the sense that it does not claim to show how exactly values are comprehended via emotion, or whether emotions are themselves constituted of axiological properties. It merely shows that, whether or not there are objective moral facts, emotion has to be part of the explanation of moral values, and in turn the explanation of moral norms and judgments. This poses a challenge to the UMG-theorist: either she shows that the axiological dimension of moral judgment can be accounted for without appealing to emotion or she admits emotion in the explanation of moral judgment. My aim in the last section of this paper will be to show that we should opt for the latter option.

The UMG-theorist may reply that expressions such as ‘greater good’ are perhaps axiological terms in themselves, but refer to objective matters of fact. As I have said, even if there are objective moral facts, i.e., if realism is true, my argument still holds. They may also argue that anything could be made value-laden. Take for instance truth. Truth is generally considered as good, and we all value it. Should we conclude then that emotion *must* be part of the mechanism of the acquisition of knowledge? I think that the analogy between truth and morality is misleading. Truth is about what is the case, morality about what ought to be the case. The first can be acquired whether or not we value it, while the latter seems to need axiological properties as background elements. Assuming the suffering of others is a universal moral concept, a moral creature cannot be a moral creature without valuing the absence of suffering. One might object here that a moral norm is not necessarily constituted by

---

8 For instance, she could make a distinction between moral intuition, a kind of intuition that would be value-free, and moral judgment, the value-laden expression of an intuition. Nevertheless, I do not think such a move would solve the problem, for it is hard, though perhaps not impossible, to imagine what a value-free moral intuition would be like.
axiological properties, for it remains possible to find moral norms that we do not value or even that go against some of our values (for instance, I may value X but feel the obligation not to do it). My argument is not that we need to value a moral norm in order to follow it, but that we need to have access to the background axiological properties that make it a moral norm (and not a linguistic rule, say) in order to understand it as a moral norm. This presumably could only happen if an emotion (perhaps a moral emotion) was triggered by the moral grammar of the judge.

My point here is that emotion can still be argued to be necessary for moral judgment without having sufficient empirical data to support this thesis, and that a model integrating emotion is still well suited.

3) Emotion as Input?

Before turning to the final section of this paper, I would like to explore the possibility that emotion may play a role at the very beginning of the process leading to moral judgment, i.e. as input to the action analysis mechanism.

Let’s take Hauser et al.’s formulation of the principle of double effect: “it may be permissible to harm an individual for the greater good if the harm is not the necessary means to the greater good but, rather, merely a foreseen side effect” (Hauser et al., 2007, 3). Presumably, this principle is part of UMG. Of course, it is an abstract principle that should not be taken as a rule that is consciously entertained by the judge when she expresses her moral judgment. Nevertheless, there is in the sentence one expression that needs to make sense to us in order for the principle to be understood: ‘greater good’. In order for a stimulus to count as a ‘greater good’, we need to be able to grasp what it means for something to be better than something else. We might also need to know what counts as good and what counts as bad in a given situation (i.e., we need to be able to assign a valence—positive or negative—to features of situations). Moreover, we may need to have access to such axiological notions during morally charged situations. In other words, in order to produce a moral judgment, an individual may first need to have access to a basic (non-moral) axiological reference framework from which a basic (non-moral) evaluative judgment (or perhaps a set of general axiological concepts) would be taken as input by action analysis. The mistake that UMG-theorists seem to have made is indeed to omit the fact that the concepts ‘good’ and ‘bad’ are relative to what we value; they do not have an univocal sense, and therefore can partly explain why two people would judge differently the same situation. If such basic axiological concepts as, say, costs
and benefits (two paradigmatic basic axiological concepts) must be taken into account by action analysis, we can safely say that emotion must play a role not only as output to MG, but also as input to action analysis.\(^9\)

If, as I have suggested, emotion is necessary in order to comprehend value-laden events, then it is necessary for the comprehension of morally charged actions, and in turn for moral judgment itself. If emotion is indeed the only way to comprehend the values that must arguably be involved in the process leading to moral judgment, it therefore fulfills both the necessity and relevancy conditions laid out above: it can explain how a given task, the comprehension of values, could be performed if that task was postulated as a necessary step for a complete understanding of moral judgment.

In the final section of this paper, I will seek to provide a tentative proposal on how moral judgment may be produced that takes these considerations into account, but that preserves the idea of a moral grammar.

### 3) The Case for a Sentimentalist Universal Moral Grammar

#### I) Two Problems With Emotion: Causation and Differentiation

If emotion is so obviously important for the production of moral judgment, why have Hauser and his colleagues failed to incorporate it into their model? It seems that this denial is based on two major, somewhat related problems that a theorist rooting moral judgment in our emotional reactions has to face. First, if emotion is the direct cause of moral judgment, what causes it? In order for emotion to be caused by an external stimulus, there must be something akin to an analysis of what is happening that must be performed first, for emotion would otherwise be triggered randomly (I will call this problem the ‘causation problem’). UMG-theorists therefore think that the action analysis mechanism that they have in mind would do the right kind of job. Second, it seems that emotion cannot, as we have seen previously, account by itself for the different kinds of norms that exist in the child’s environment (let’s call it the ‘differentiation problem’). “Our emotions, Hauser says, can’t explain how the child
navigate the path between social norms in general and moral norms in particular. A child's experiences are insufficient to create the dividing line between general social or conventional transgressions and specifically moral transgressions. (...) What we need, therefore, is an understanding of the evaluative process that triggers emotion.” (2006, 30) The solution the UMG-theorist gives to the problem of the differentiation between the different kinds of norms is that there must be a distinct, emotion-free, evaluative process that must occur before any emotional activation. However, as we have seen, the kind of evaluative process that must occur during moral judgment cannot be accounted for without appealing to emotion. Or can it? Hauser thinks it can:

The idea I wish to introduce here is that we consider thinking about the origins of our sense of right and wrong by starting with the process of generating an expectation. Before human infants can run, climb, eat with a fork, discuss their impressions, and understand humor, they can form expectations about patterns of action in the world. (...) A different sense of expectancy arises in the moral domain when we think about what we expect others to do or what we should do. These are normative expectancies and they refer to obligations, promises, and commitments. (2006, 167-168)

If I follow Hauser’s ‘moral expectation’ hypothesis, the picture I would give of the production of moral judgment would look like this: (1) I perceive an action being performed (say someone being hit), (2) my (general-purpose) action analysis mechanism is triggered, and (3) it makes me think (or feel) that there is something weird about the situation (activation of a moral norm), (4) which generates the intuition that the action is morally wrong; as a result, (5) I may express my intuition by saying that the person who did the hitting has done something wrong. This causal chain, though schematic, is appealing to the UMG-theorist because it enables her to put all the weight of the explanation of the production of moral judgment on the working of action analysis (2), for it is presumably what enables us to activate moral norms (3), intuitions (4), and public judgments (5). Here, the action analysis mechanism gives as input to the Moral Grammar (MG) some information about the situation, and MG gives as output an intuition about this same situation.

One might be tempted to say that, if we accept my argument, emotion must somehow come before the activation of the moral norm (i.e., the working of MG) so as to bridge the gap between the structural analysis of the situation and the categorization of it as morally charged, for a moral norm cannot be endorsed (by an individual) without emotion. However, this is problematic. First, it is based on a confusion about the nature of the moral norms that are part of MG. Such norms are computational processes that enable the transition between descriptions of situations and their categorization as morally charged. They are not the kind of
moral norms that we have in mind when we engage in moral talk; in other words, they do not have an axiological component in the everyday (and relevant) sense. Therefore, we do not seem to need emotion before the activation of a moral norm because MG is where the transition between a purely descriptive analysis of an action and the categorization of this same action as morally charged is performed. The kind of evaluative process that UMG-theorists deal with when they talk about moral grammar is not the kind of evaluative process that we have had in mind in this paper: the second is impossible without emotion, while the first gives us abstract constraints on our judgments (whatever that might mean). Second, action analysis, though supposedly general-purpose, is supposed to provide all the necessary information about the situation (causal structure, intentions, mental representations, whether someone is suffering, etc.), which prevents us from saying that a structural analysis of a situation does not give us enough information for a moral norm to be triggered (and arguing that emotion must somehow directly pick out the relevant properties in the world). Overall, postulating MG seems to be the best strategy given the problems of causation and differentiation raised against emotion.

II) Bridging the Perception Gap

Given the empirical constraints that we have now at our disposal, I think that, in order to argue for the idea that emotion must be part of the explanation of moral judgment, we need to qualify the thesis according to which emotion is necessary for the endorsement of a moral norm. I would therefore like to make a distinction between the activation of a moral norm in MG and the endorsement of this norm by a particular individual. A moral norm in MG is akin to an abstract principle that is both unconscious and partly innate (i.e., constrained by the principles of UMG during the child’s development). The endorsement of a moral norm is, as I have argued, based on an emotional reaction to (or at the very least an emotional disposition towards) certain features of actions. Thus it must be partly conscious, for a distinct phenomenal character identifies it. Although we may not be aware of the moral norm that is at the basis of our judgment, it is hard to deny that there must be one, based on the way we ‘feel’ when we are facing a morally charged situation (i.e., on ‘how it is like’ to produce such a judgment). Nevertheless, the particular phenomenology of moral judgment does not by itself support the thesis that emotion must occur before moral judgment. As UMG-theorists would contend, moral experience is indeed real, but may still be caused by moral judgment, rather than the other way around.
Recall that, in order for a moral judgment to be performed, an individual must somehow comprehend the situation she judges as morally charged. In other words, there must be in the process something that bridges the gap between perceiving an action and perceiving the same action as morally charged, i.e. between action analysis and moral judgment. MG presumably performs the categorization of actions by taking descriptions and applying moral norms to them. A problem raises once we take into consideration the idea that emotion is necessary for the ‘perception’ of situations as morally charged: although MG’s moral norms enable categorization, this categorization is by itself insufficient for an individual to produce a moral judgment, for she hasn’t yet perceived the relevant situation as morally charged. Given the fact that emotion could not possibly occur before MG (for the reasons presented above), emotion must therefore occur between MG and the production of a moral judgment. The following model captures this intuition:

(Basic Emotion/Value $\rightarrow$) Action Analysis $\rightarrow$ MG $\rightarrow$ Emotion $\rightarrow$ Moral Judgment

Here MG produces an abstract moral norm that causes an emotion which in turn enables the judge to comprehend the relevant properties that make the activation of the norm appropriate (i.e., see the ‘moralness’ of the action). In a sense, the evaluative process comes to an end only when the judge has perceived the action as morally charged, i.e. only when the judge has an emotional reaction to it.\(^{10}\)

This new model is supported by various empirical data. First, it is in agreement with Greene et al.’s (2001) neuroimaging data: given what has been argued in this paper, the emotional activation found in this study probably comes before the judgment, rather than after. Moreover, data on psychopaths, though not conclusive in themselves (see section 2), now support the idea that emotion is part of the causal process leading to moral judgment. We can indeed hypothesize that, because the psychopath has an emotional deficit, the link between MG and moral judgment cannot be made, which in the long run may be detrimental to the moral grammar itself (since no genuine moral judgment would ever be taken as input in order to produce other moral judgments). To conclude, it seems that adding emotion to the explanation of moral judgment makes more sense than not postulating it, given both our intuitions about moral judgment and the empirical data we have.

\(^{10}\) This is why an alternative hypothesis would be that the emotional activation may be equivalent to a (not-yet-expressed) moral judgment. In this paper, I will not try to argue for one or another hypothesis.
III) A problem with the Model

In this paper, I have argued that an emotion must be triggered by MG in order for a moral judgment to be produced, for it is necessary for the perception of situations as exemplifying morally charged features. However, this thesis appears to go against the common observation that many of our moral judgments do not seem to be caused or accompanied by emotion. If I say out loud that killing is wrong, it is hard to believe either that an emotional reaction caused my judgment, or that my judgment was accompanied by emotion. At the very least, my judgment does not have the obvious emotional import that a judgment in context may have. Moreover, studies on VM patients (Mendez et al., 2005) have shown that such individuals, while suffering from an acquired emotional deficit, still manage to produce normal moral judgments, the only exception being their judgments when confronted to trolley dilemmas. This is at first sight strong evidence against the thesis according to which emotion must be triggered in order for moral judgment to be produced.

In order to respond to such an objection, I would like to distinguish between two kinds of moral judgments: on the one hand, there are full-fledged, partly emotional, moral judgments; on the other hand, there are moral judgments that are based on past full-fledged moral judgments and that are retrieved from memory by the judge. When I say that killing is wrong, I may not feel anything for the simple reason that I have condemned killing in the past, many times, and felt anger, disgust, and sadness in front of such acts of cruelty. In other words, it is because I have genuinely judged killing as wrong in the past that I can now genuinely say that I believe that killing is wrong. Put differently, my unemotional uttering of ‘Killing is wrong’ is indirectly caused by my past emotional condemning of murder. Emotion here is therefore still necessary.

This explains the fact the VM patients, though emotionally deficient, can still produce normal moral judgments: they can say that killing is wrong, for instance, by using pre-deficit memories of full-fledged moral judgments. Furthermore, and more strikingly, this explains the fact that the patients are unable to judge like normal humans in cases of trolley dilemmas: since they have never encountered cases like the trolley dilemmas before their accident, they never made full-fledged moral judgments about them, and therefore are unable to perform as normal humans do in such tasks. Studies on patients with damage in their prefrontal cortex (Damasio et al., 1999) also seem to confirm this hypothesis. While the patients who acquire an emotional deficit in their early childhood are unable to judge moral situations as normal
humans do, those who acquire their deficit later in life seem to be able to produce normal moral judgments in a wide variety of cases, the difference plausibly being that the first, contrary to the second, did not have the chance to store full-fledged moral judgments in their long-term memory. Of course, many of our moral judgments are not accompanied by emotions, but this doesn’t mean they are not still somehow caused by them.

**Conclusion**

My aim in this paper has been limited: to show that the model provided by UMG-theorists to account for moral judgment may be incomplete. By appealing to the widespread view according to which values are conceptually related to emotion, I have provided a good reason for UMG-theorists to take stock and reflect on the metaphysical problems that their explanation engenders. If they do not grant emotion any real explanatory role (by at best, to use Hauser’s phrase, making it part of moral judgment’s ‘support team’), they will be forced to deny either that values can be explained by appeal to our emotional reactions or that the experience of producing a moral judgment has a value-laden dimension. In other words, they will have to give us good reasons to believe that there is no gap (an explanatory gap, indeed) to be bridged between the mere perception of an action and the perception of the same action as morally charged: this is the Emotion Challenge.
Bibliography


Yan, A. (2008). *Proceedings of world academy of science, engineering and technology*, 30, 404-408