Morality According to a Cognitive Interpretation: A Semantic Model for Moral Behavior

Sara Dellantonio and Remo Job

Abstract. In recent years researches in the field of cognitive psychology have favored an interpretation of moral behavior primarily as the product of basic, automatic and unconscious cognitive mechanisms for the processing of information, rather than of some form of principled reasoning. This paper aims at undermining this view and to sustain the old-fashioned thesis according to which moral judgments are produced by specific forms of reasoning. As critical reference our research specifically addresses the so called Rawlsian model which hinges on the idea that human beings produce their moral judgments on the basis of a moral modular faculty "that enables each individual to unconsciously and automatically evaluate a limitless variety of actions in terms of principles that dictate what is permissible, obligatory, or forbidden".[25, p. 36] In this regard we try to show that this model is not able to account for the moral behavior of different social groups and different individuals in critical situations, when their own moral judgment disagrees with the moral position of their community. Furthermore, the critical consideration of the Rawlsian model constitutes the theoretical basis for the constructive part of our argument, which consists of a proposal about how to develop a semantic, quasi-rationalistic model to describe moral reasoning. This model aims to account for both moral reasoning and the corresponding emotions on the basis of the information which morally relevant concepts consist of.

1 Introduction

Contemporary cognitive research identifies three fundamental models that describe moral behavior and in particular the processes leading to moral

Sara Dellantonio · Remo Job
Department of Cognitive and Education Sciences, University of Trento,
Rovereto, Italy
e-mail: sara.dellantonio@unitn.it,remo.job@unitn.it

judgments [25, p. 45]. The first model is called *Humean* and is based on
the idea that human beings are equipped with an innate moral sense, of an
emotional kind, that drives judgments regarding right and wrong. The model
foresees that an individual's perception of a morally relevant situation auto-
matically triggers a positive or negative emotional response that leads to the
moral judgment of the situation: the perceived situation will be considered
as right in the case of a positive emotion and as wrong in the case of a nega-
tive emotion[1]. In fact according to this view moral pronouncements are not
"judgments" in the proper sense of the word but *immediate and spontaneous
intuitions that assume for the subject the appearance of evident and obvious
truths.*

A second model may be defined as *rationalistic* and is based on the op-
posite idea that moral statements are not conceived as intuitions, but as
judgments produced by conscious reasoning on explicit knowledge at the dis-
posal of the subject[2]. This reasoning may be intended as based on utilitarian
principles, but it may also be interpreted in different ways. One of the main
alternative views is a *Kantian* one, according to which moral reasoning can
neither be utilitarian (since morality is what allows humans to overcome par-
ticular, egoistic utility to reach an universalistic good) nor can it consist of
the capacity and/or in the willingness to learn and to follow legal rules (since
there is not necessarily a coincidence between what the law stated and what
is right from a moral point of view)[3].

A third model – qualified as *Rawlsian* – hinges on the idea that human
beings produce their moral judgments on the basis of "an evolved capacity of

---

[1] The contemporary psychological research about the Humean model became es-
tablished in the 1980s mainly through Martin Hoffman [27, 28, 29] and involves
authors such as Kagan [31]; Frank [15]; Schweder and Heidt [51]; Wilson [60];
Haidt [18]; Haid and Bjorklund [19].

[2] The rationalistic model originates mostly from Kohlberg's researches on morality,
which are based on Piaget's developmental theory: see Kohlberg [33, 34, 35].

[3] The description we give here of this second model does not correspond entirely
to Hauser's classification of the main views on morality currently discussed in
the field of the cognitive debate. In fact, Hauser speaks of a Kantian model,
which actually consists of a hybrid view of moral judgment according to which
the perception of a morally relevant situation produces both conscious reasoning
(intended mostly as utilitarian reasoning) and emotional reactions. According
to this view, emotion and reasoning interact in the production of moral judg-
ment; if their assessments diverge, the production of a judgment requires that
one gets the "upper-hand" over the other. First of all, we refuse to call this ratio-
nalistic model "Kantian" because Kant's perspective on moral judgment can be
connected neither with a utilitarian reasoning nor with emotional reactions. Sec-
ondly, to speak of a rational model allows us to present the positions in play first
of all in their most simple versions, avoiding hybrid views where both rationality
and emotions are in play. For a clarification of the difference between a purely
rationalistic model and a hybrid one and for a brief overview of the positions
belonging to this hybrid model see Hauser, Young, Cushman [26].

all human minds that unconsciously and automatically generates judgments of right and wrong" [25, p. 2]. More precisely: "all humans are endowed with a moral faculty – a capacity that enables each individual to unconsciously and automatically evaluate a limitless variety of actions in terms of principles that dictate what is permissible, obligatory, or forbidden" [25, p. 36]. According to this model the perception of a morally relevant situation is followed by an analysis of the situation on the basis of a moral module which decomposes it into its elements and that uses its own principles to evaluate them and to formulate moral judgments.[4] Once it has been expressed, the moral judgment causes an emotional reaction and a reasoning process aimed at justifying it *post hoc*. Even if this model is different from the Humean one in the way it describes the process that leads to the moral pronouncements, it shares with it the idea that morality is a kind of intuition based in this case not on emotive evidence, but on an *unconscious and automatic processing* of information[5].

Over recent years cognitive research has concentrated a lot of attention on the Humean and on the Rawlsian model at the expense of the traditional rationalistic model, which is mostly considered outdated[6]. These models seem indeed to be more compatible with the contemporary researches in the field of cognitive psychology, which favour an interpretation of moral behavior primarily as the product of basic, automatic and unconscious cognitive mechanisms for the processing of information rather than of some form of principled reasoning. Furthermore, the Rawlsian and some versions of the Humean model are characterized by another aspect which turns out to be extremely desirable from a cultural point of view. Indeed, if these models should turn out to be true, they would allow a tracing of a common fundament, of a biological nature, for the moral principles consisting of the specific cognitive mechanisms which drive human judgment about right and wrong. In this case, moral discourse need not be based on pure rational or sociological or even religious argumentation, but could become part of a scientific-experimental investigation of human cognition. In fact, as Hauser states, the identification of innate principles of this kind would determine "the range of both possible and stable ethical systems" [25, p. 54].

Even though this way to approach the problem of morality has become a major trend in the field of cognitive research, "mechanistic" explanations of the moral behavior often show large difficulties in explaining moral views of people in real situations outside of the experimental settings defined in the laboratory. Indeed, this paper aims at undermining the view that moral

---

[4] About the question of modularity and about the hypothesis that the central system is also organized according to (more or less rigid) modules see e.g. Samuels [50] and Carruthers [3].

[5] Among the most prominent authors that contributed to developing the so called Rawlsian Model there are Susan Dwyer [12, 13]; John Mikheil [37]; Gilbert Harman [24] and especially Mark Hauser [25].

[6] For an overview of the most important difficulties identified by this model see for example Thomas [55].

behavior can be seen as the product of a form of intuition produced by some
unconscious and automatic processing of information and to sustain, on the
contrary, the old-fashioned thesis according to which moral positions express
judgments produced by specific forms of reasoning. This thesis does not imply
that considerations of psychological or of cognitive nature no longer play any
role in the understanding of moral behavior, but it does mean that moral
behavior has to be explained on the basis of *conscious and non-automatic
high-level cognitive mechanisms related with the thought and with the human
capacity to reason and to judge.*

As critical reference our research specifically addresses the Rawlsian model[7].
We individuate two kinds of principles that many of the authors, who are in
agreement with this model, believe belong (among others more controversial)
to the moral faculty: the one consists of some innate moral contents, the other
concerns the capacity to distinguish strictly moral principles from merely con-
ventional norms. Furthermore, we argue that, if we assume the existence of a
moral faculty based on these kinds of principles, we will not be able to account
for the moral behavior of different social groups and of different individuals in
critical situations, when their own moral judgment disagrees with the moral
position of their community.

The critical consideration of the Rawlsian model constitutes the theoretical
basis for the constructive part of our argument, which consists in a *proposal
about how to develop a semantic, quasi-rationalistic model to describe moral
reasoning.* This model aims to account for both the moral reasoning and
the corresponding emotions on the basis of the information which morally
relevant concepts consist of.

2   The Rawlsian Model and the Linguistic Faculty

The Rawlsian Model Moves from a Structural Hypothesis which is congenial
to the functional architecture of the cognitive system as it is conceived by the
classic cognitive science, since it is based on a faculty interpreted as a module

---

[7] The reason why we prefer to consider this model instead of the Humean one is
firstly that the position of the Rawlsian model is more univocally and decidedly
oriented towards a mechanistic view of moral behavior. Indeed, even though –
as noticed before – both the Humean and the Rawlsian models tend to give an
account of moral behavior in terms of unconscious and automatic processing of
information, Humean models (which are by the way also very variegated in the
specific positions they maintain) attribute in general to those mechanism a weaker
role as regards the determination of the end results of the moral output. Indeed,
while the Humean model just assumes that human beings are characterized by
an innate affective constitution, the Rawlsian model makes a much more binding
presupposition, assuming the existence of an innate moral modular faculty which
works according to specific fixed principles and on specific and fixed information.
A Humean model that does not imply any mechanistic view of moral judgments
is for example the one proposed by Prinz [45].

and explains moral judgments in analogy with Chomsky's grammaticality judgments. Its denomination is due to the fact that it was John Rawls who first explicitly drew an analogy between language and morality and who hypothesized the existence of a moral faculty based on the linguistic faculty proposed by Chomsky.[48, p. 55]

Actually, the possibility of using the model of Universal Grammar in order to determine the functioning of a hypothetical modular moral faculty has also been put forward by Chomsky himself:

> The acquisition of a specific moral and ethical system, wide ranging and often precise in its consequences, cannot simply be the result of "shaping" and "control" by the social environment. As in the case of language, the environment is far too impoverished and indeterminate to provide this system to the child, in its full richness and applicability. [...] it certainly seems reasonable to speculate that the moral and ethical system acquired by the child owes much to some innate faculty. [7, pp. 152–153]

This idea met with big success both in psychological and in philosophical research and it has been further developed by authors like Susan Dwyer [12, 13], John Mikheil [37], Gilbert Harman [24], and especially Mark Hauser (2006) who tried to elaborate, in a concrete fashion, a modular Rawlsian model of moral competence.

As Chomsky's quotation already suggests, among the main arguments these authors appeal to in order to argue for the existence of a modular moral faculty analogous with the linguistic faculty there is the so-called "poverty of stimulus argument". Originally provided by Chomsky[8], this argument is based on the observation that children learn language early and easily, even though its rules are never taught to them and the stimuli available to them to reconstruct these rules on their own are extremely poor, fragmentary and asystematic. Since the experience the child can rely on is not – according to Chomsky – sufficiently rich to justify such a fast and easy learning process, we must assume that humans are endowed with an innate linguistic faculty that organizes and completes the available experience in a way that allows the learning of language. The authors who maintain the Rawlsian position apply the same argument to moral competence. They claim that the everyday experience children can rely on in order to learn the moral rules of their own group is partial and underspecified. Therefore, in order to account for the fact that moral rules are learned precisely and quickly in early childhood, we must assume the existence of a specific faculty that organizes and completes children's experiences to allow the learning of the moral rules of their group[9].

In order to give an idea about the way this moral faculty is supposed to work, these authors refer again to Chomsky and to his *Principles and*

---

[8] About this see in particular Chomsky [4, pp. 2–11], and Chomsky [5, pp. 5–6].

[9] Among the most important authors which appeal to the "poverty of stimulus argument" in relation to moral learning there are Mikheil [37]; Dwyer [12]; Harman [24]; Mahlmann [36]; Mikheil [38]; Nichols [42].

Parameters Theory [7, p. 62 ff.]. According to this theory (whose general structure in the field of linguistics has actually been developed in various ways) the Universal Grammar is made on the one hand of universal principles which are common to all real and possible languages, and on the other hand of universal parameters, which complement these principles for some particular aspects. The idea behind parameters is that some basic characters of the grammar of natural languages are variable, even though only a *limited and defined* set of variations is possible. Parameters have the function of defining these possible variations. One of the hypothesized parameters governs, for example, the ordering of subject, verb and object in the phrase which is supposed to vary in different languages according to a very limited number of options. According to Chomsky's theory even a minimal linguistic experience is sufficient for the child to set the parameter in the right manner depending on the language he is learning and so to organize from that moment on further stimuli according to that parameter.

The existence of parameters accounts for the fact that children learn the particular grammar of their own language in its specificity and difference to other languages. In the same vein, the hypothesis that the moral faculty is also made of parameters is introduced to explain the fact that children sharing common universal moral principles are able to learn the particular variation of these principles adopted by their own community. Once the grammar of a particular language has been learned on the basis of the principles and parameters, the children will be able to use it to produce spontaneously, and without reflection, well-formed sentences in that language. In the same way, once the moral grammar of a group has been learned, this will allow the member of the group to produce *spontaneously and without conscious reflection* moral judgments that reflect that grammar and its underlying principles.

3   The Problems with This View

As the idea of the Universal Grammar proposes that our linguistic competence is driven by innate unconscious, operating principles, the idea of a Universal Moral Grammar entails a view of moral judgments, according to which they are the automatic and immediate product of unconscious moral principles. Moral positions are no longer conceived as a form of judgment, resulting from complex, principled reasoning, as common sense used to think, but they are understood in analogy to language as the result of a creative mechanism which produces immediate moral intuitions before and independently from conscious thought.

The possibility to identify an automatic mechanism for the production of moral intuitions has, according to some, the reassuring effect to make all humans appear essentially similar and moral. A proof for that is the fact that the analogy between language and morality (often called "*linguistic analogy*") lets us take for granted that we can speak of a "moral competence", while

the opposite idea of "moral incompetence" appears paradoxical. However, it should not be disregarded that these reassuring aspects are the outcome of a naturalistic conception according to which our moral sense is part of the human nature and its judgments are entirely driven by our cognitive mechanisms. As Jesse Prinz points out:

> Recently researchers have begun to look for moral modules in the brain, and they have been increasingly tempted to speculate about the moral acquisition device and innate faculty for norm acquisition akin to the celebrated language acquisition device promulgated by Chomsky [...]. All this talk of modules and mechanism may make some shudder, especially if they recall that eugenics emerged out of an effort to find the biological sources of evil. Yet the tendency to postulate an innate moral faculty is almost irresistible. For one thing, it makes us appear nobler as a species, and for another, it offers an explanation of the fact that people in every corner of the globe seems to have moral rules. Moral nativism is, in this respect, an optimistic doctrine – one that makes our great big world seem comfortingly smaller. [46, p. 367]

These reflections show unequivocally that embracing a mechanistic view of morality has an important cultural and social impact. Still, we are not so much interested in the actual desirability of this view; rather we aim at investigating the general plausibility of the thesis that moral positions are the product of a modular moral faculty. In this regard we need to consider *what the unconscious principles this module is supposed to work with are, and whether these principles permit accounting for moral behavior as it manifests itself in the everyday life of subjects and of social groups.*

In order to address this aspect we need to distinguish two different groups of principles. The *first* group consists of principles that cannot be considered specifically moral and that are supposed to explain how the moral faculty structures the information in terms which are suitable for a moral evaluation, for example identifying in the stream of perception specific actions or events, their direct or indirect consequences or understanding whether someone caused them or is responsible for them.[10] The *second* group consists of principles that are specific to the moral competence and that drive directly the decision about what is permissible, obligatory, or forbidden. In order to carry out any further reflection about the moral faculty, we need first of all to examine this second group of principles.

As far as this second group of principles is concerned, the literature is characterized by hypotheses which are also very different from each other. Some of them recur more often and are considered more important. (A) First of all many authors agree that among the working principles of the moral faculty there are at least some *innate moral content* like the prohibition of murder, harming, stealing, cheating, lying, breaking promises, and committing adultery[11]. Since these principles are evidently subject to a lot of exceptions all

---

[10] See for example Hauser [25, pp. 8, 21, 41, 45–48, 166–182].
[11] See for example Hauser [25] who refers back to Mikheil [37].

the time, they are usually interpreted in a parametric way. This means that, even though they are considered to be universal, they are also supposed to consist of some parametric variables that assume a specific value just in the "moral community" a person happens to grow up in[12]. The most important variable in this respect is the determination of who is worthy of moral consideration and who is not. Susan Dwyer describes this problem in a very clear manner introducing the notion of "*schweeb*".

> Let us [...] define a schweeb as "creature with the highest moral status". A very basic principle of all possible [internalized] moralities might be "Schweebs are to be respected" or "Given the choice of saving the life of a schweeb or saving the life of a non-schweeb, always save the life of a schweeb." [13, p. 249]

Since human groups never identified univocally or permanently who qualifies as a "*schweeb*" (or, in less imaginative juridical words, as a "person", i.e. as a subject who is recognized as having specific rights), authors that maintain a Rawlsian view consider this notion to be a parameter in the sense that all groups decide and impose to each member who is a "*schweeb*" according to its own moral rules. In this way it becomes possible to explain why each culture prohibits, for example, the killing of some people even though it permits the killing of some other people according to criteria which resemble the logic of the ingrouping and outgrouping.

(B) A second type of principle, which is usually considered characteristic of the moral faculty, expresses a capacity. Specifically the capacity to distinguish situations which have to be considered properly moral and need to be evaluated using moral rules, from other kinds of situations where they may face a form of disrespect for some social rules but without any real moral issue being in play. This principle is considered analogous to one assumed in relation to the linguistic faculty which is supposed to account for the capacity children show to select just auditory inputs of a linguistic kind distinguishing them from other kinds of sounds. In the case of the moral faculty it is maintained that people need to distinguish on the one hand authentically moral situations and violations – which are mainly associated with harming – and on the other, situations and violations which deserve social blame (like going to the office in pyjamas) or which cause revulsion (like licking the lavatory seat) without being morally relevant[13]. In addition to Hauser, Susan Dywer also attributes to this capacity a great importance for the definition of a moral faculty. She maintains that the poverty of stimulus argument applies to it. If children can begin to distinguish properly moral situations at a very early stage of their cognitive development, without explicit instruction and even though they do not seem to have enough experience to reconstruct the

---

[12] See Dwyer [12, p. 177].

[13] The examples mentioned here have also been investigated experimentally. See below in this section for reference.

criteria commonly used to identify properly moral situations, then we need to admit that this capacity has to be innate[14].

The discussion of this issue began with the publication of Elliot Turiel's studies according to which moral norms needed to be sharply differentiated from other kinds of norms for being:

1. *objective* (in the sense that their prescriptive power is not supposed to depend on extrinsic authorities);
2. *general* (in the sense that their prescriptive power is not perceived as limited to a particular group, place or time, but it is considered to extend to any group, place or time); and
3. *important* (in the sense that their violation is perceived as something extremely serious that harms the wellbeing of the people who experience it)[15].

Furthermore these investigations go together with others which aim to show that non-moral disapproval is experienced as very different from moral disapproval even when it comes with deeply negative emotions evoked by moral violations[16].

The issue which needs to be faced in relation to this principle and to the previous one is whether they have to be considered plausible and, furthermore, whether it is plausible to assume that they are part of something like a moral faculty.

(B) Let us consider this question starting from this second principle according to which – to sum up briefly – humans are already able at a very early stage of their cognitive development to determine in a universal and transcultural manner which situations require a moral evaluation, since they are connected with harming, and which others need to be evaluated through milder norms concerning good taste and customs. If we could establish that humans do have a capacity like this, this would allow us to define shared and indisputable moral issues valid for every cultural group in any time, distinguishing them from other kinds of situations which do not deserve consideration from a moral point of view. However, as Jonathan Haidt and his colleagues showed, the identification of a distinction of this kind which binds the morality issue with harming, is only a mystification resulting from the naive Western laity of educated classes. In reality people or groups characterized by strong systems of values derived for example by religion, or even

---

[14] See for example Dwyer [12, 13]. For critics to the application of the poverty of stimulus argument to this competence see Prinz [46, pp. 392–395]. See also Dwyer's reply to Prinz in Dwyer [14].

[15] Among the most important studies which embrace the thesis of the psychological realty of the distinction between moral and conventional rules see for example: Turiel [57] 1983; Nucci [43]; Turiell; Killen, Helwig [59]; Smentana, Braeges [53]; Smentana [52]; Tisak [56]; Nucci [44]; Turiel [58].

[16] About this aspect see for example: Zahn-Waxler, Radke-Yarrow, Wagner, Chapman [61]; Nichols [40], Nichols [41, pp. 23–25].

just less educated people, do not share the same indissoluble association between morality and harming, but they do see as morally binding norms related to other kinds of facts. They also see as morally relevant norms which well-educated Westerners often consider to be "matters of custom" (like the prescriptions related to specific foods, clothing or festivals) or "matters of good taste" (for example the ones that seem repugnant or inappropriate like licking the lavatory seat or cleaning it with the national flag or masturbating with a dead chicken)[17]. Furthermore, people of low social status or coming from conservative, strongly religious backgrounds tend to also consider morally relevant lifestyle choices like having a relationship with a person of the same sex. For these reasons the existence of a universal principle that determines which actions are subject to a moral judgment and which are not, seems mostly a form of wishful thinking that does not find any correspondence in the way different cultures or even different groups belonging to a same culture think.

(A) An analogous reflection can also be carried out in relation to the supposed *innate moral contents* of the moral faculty, which according to Hauser and the pioneering study of Mikheil that he and others rely on, include things such as murder, harming, stealing, cheating, lying, breaking promises, and committing adultery[18]. As mentioned previously, the authors who introduced these principles are aware of the fact that their application in different cultures or groups is liable to many exceptions and they try to solve this problem interpreting them in a parametric way. According to this interpretation it is the cultural context which determines to whom they apply and who is an exception to them. Universal principles assume therefore the following form: "Murder/harming/stealing from/cheating/ etc. members of the group X is forbidden (morally wrong)", while X is determined contextually, depending on which creatures are "schweebs" (i.e. deserve the highest moral status) in the considered culture. Still, even though we give a parametric interpretation of these principles, at least some of them already appear at first glance unlikely and derived by a naive projection of the values of our contemporary Western society back in time, or elsewhere in space. This is decidedly the case for example of adultery, since adultery could be considered an innate principle just if we admit that the original condition of human communities is monogamy instead of polygamy, which anthropology considers to be false[19].

---

[17] All the proposed examples including the following one are taken from famous experiments of Heidt and colleagues. See for example Haidt, Koller, Diaz [23]; Heidt [18]; Heidt, Joseph [22]; Haidt, Graham, [21]. For a general report of these studies which takes Heidt's part see for example Kelly, Stich [32].

[18] See Hauser [25, p. 48]; Mikheil [37].

[19] Domestic groups have taken very different forms during the human history: before nuclear family became established, humanity went through a condition of promiscuity and subsequently of matrifocality, polygyny, and polyandry. About this see for example [49].

If adultery can be considered a borderline case which has been incautiously included in the list of the universal principles, the point is that the whole idea of the universal contents proposed by the Rawlsian account seems to be highly problematic. Let us consider this aspect discussing the case of principles which, at first glance, appear shared and transcultural like murder or harming. A good way to pose the problem is suggested by Sripada and Stich [54, p. 282], according to whom it does not make sense to appeal to universal principles as being valid in every culture like "murder is wrong", since these principles express just purely analytical sentences, i.e. sentences which are true in virtue of their meaning only. The fact that murder is wrong is, for example, already implicit in the meaning of "murder" which can be expressed in terms of "killing someone in an impermissible way" and which does not say anything about what kind of killing counts as murder and what does not (think for example about executions which are considered admissible or inadmissible depending on the political systems). In this sense "murder" is just an empty form of moral discourse which can be arbitrarily filled with all kinds of killing which the judging subject considers inadmissible. The truly moral problem does not therefore concern the question about whether all cultures agree to condemn murder, but rather the question about the criteria they use to decide which forms of killing are admissible and which are inadmissible. This way to pose the problem suggests a different "linguistic analogy" from the one embraced by the Rawlsian model, which is based on semantics instead of syntax (to which Chomsky's thesis primarily applies). It suggests that *what we really need to understand in order to face the moral problem is under which condition a subject categorizes an event as an act of murder, harming, stealing, cheating etc. rather than as an act of a different kind which can be considered as morally legitimate.*

Even though this semantic interpretation of the linguistic analogy opens up new questions related to the fact that there is not a unique recognized theory of categorization, and that therefore the reliability of the semantic model we propose depends on the reliability of the semantic theory we choose, in our opinion it still permits us to approach the moral problem in a more productive way. Firstly, it permits us to individuate more clearly the role and the weight of the cultural influence on moral judgment. Secondly, tracing back the moral problem to semantics helps us to avoid, in part, strong assumptions about the existence of automatic, innate cognitive mechanisms and restores the function of reasoning in moral choices. Affirming the idea that reasoning does play a function in moral choices allows us to account for some very relevant aspects of the moral discourse which have been mostly ignored by the Rawlsian model.

Indeed the views which conceive moral positions as intuitions – i.e. as an automatic output of a mechanism which has assimilated the cultural conventions of the culture of origins (think about the Universal Grammar as a means to learn the specific grammar of the native language) – identify somehow the moral rules with the rules imposed by a culture, while at times making a

morally right choice means distancing oneself from culturally imposed rules. A well-known and often debated example in this direction is the case of Germans during the Third Reich who disagreed with race laws issued by their own government, even though they belonged to the Aryan community and believed in the nationalistic ideals of the German Right. This phenomenon, which has often been defined as "tacit dissension" is of great importance for the moral psychology since it shows the need to assume that humans are also capable of autonomous, subjective moral choices, which are separate from the "corporative moral".

In fact, dissent can hardly be explained in the theoretical framework of the Rawlsian model which stresses the automaticity of moral judgment in analogy with the automaticity of the process that lead us to produce grammatically well-formed sentences in our language. But, if in the case of grammar we can assume that sentences are well-formed because they correspond with the grammatical rules of our languages, in the case of morality it is possible that something is judged as morally wrong by some member of a group even though the group itself accept is as right. Considered more generally, the problem is how to account for the possibility to qualify an act as morally wrong or as morally right independently from the rules that a particular culture adopts at a certain time. This problem can also assume a different form which is of immediate relevance for the Rawlsian model: even though a cultural group determines who counts as a "schweeb" in that group (for example the Aryan, in the case of the Third Reich), a moral model needs to account for the fact that sometimes someone can disagree with the criteria adopted by the group and recognize that other people (for example the Jew) deserve to be regarded as "schweeb". The general point here is that the Rawlsian model does not account for phenomena like these concerning subjective autonomous moral judgments.

4   Which Theory of Concepts?

In the field of cognitive research, when categorization is addressed, people tend to confuse two different levels of the analysis. On the one hand there is the categorization in the sense of the *prelinguistic* cognitive procedure that "puts together" similar instances, forming the "conceptual units" children need in order to learn their first language; and on the other hand there is the categorization in the usual linguistic sense. In the first case the point at issue is the way in which infants or a hypothetical Robinson Crusoe who grew up alone on an island – or even animals – recognize some instances as similar to each other and group them together in "conceptual units", which do not depend on any linguistic, conventionally determined category. In the second, the matter under discussion are the forms of categorization carried out by means of the language, when adult subjects group instances together on the basis of similarities suggested by the semantics of the language they learned. The

difference between these levels of analysis can be further clarified by introducing two notions of semantics: an *Externalized Semantics* (*E-semantics*) and an *Internalized Semantics* (*I-semantics*)[20]. The notion of E-semantics accounts for the semantics in its conventional and public dimension, which is characterized by rules whose aim is to assure the possibility of intersubjective communication. The notion of I-semantics addresses instead the issue of which information people use (internally, i.e. in their mind) to carry out categorizations and to understand the linguistic meaning (as it is codified in the E-semantics).

These two notions of semantics can clearly not be considered completely independent from each other, but they necessarily merge together through linguistic learning: It is mostly for this reason that many studies think it is not necessary to distinguish between them. To explain this aspect it is useful to think about what happens when an infant learns his first language: in order to learn linguistic meanings, the infant must already be able to carry out some form of categorization which allows him to formulate a hypothesis about the possible use of the words he hears. Nevertheless, once he has learnt his first language, the specific concepts of *that* language (of *that* E-semantics) retroact on his previous prelinguistic way to categorize, determining a new way for him to see the world. I- and E- semantics are connected through a double binding: the I-semantics set the fundamental criteria of the E-semantics, in the sense that it would not be possible to learn E-semantics if it was not compatible with the I-semantics; this compatibility with the I-semantics is what the E-semantics of all existing languages must have in common. However, once it has been learned, the E-semantics affects the I-semantics of subjects, which at least to some extent become conformed to it. The thesis we will try to support in the following is that a cognitive account of morality must rely on the relation between *I-semantics* and *E-semantics*.

Let us focus first of all on the categorization at the level of the *E-semantics*. At this level we think that a theory with great explanatory power is the so called *theory-theory* which – on the basis of a Quinean view on language – interprets the semantic systems as complex and highly structured theories abut the world whose elements are interlinked and determined by each other[21]. According to this view people's beliefs are articulated bodies of knowledge, which work like theories: concepts express single elements of these theories and are identified on the basis of the rule they play in the whole of people's beliefs[22]. What this view suggests is that when a subject learns a language (an E-semantics), he acquires the belief system of which this language is an

---

[20] This distinction is diffusely discussed mostly by Ray Jackendoff who develops it on the basis of Chomsky's analogous differentiation between I-Language e E-Language: see for example Chomsky [6], §2.2 e §2.3 and Jackendoff [30, p. 22]

[21] See for example Quine [47]. In the field of the cognitive science the so-called theory-theory has been introduced and developed first of all by Carey [1, 2] and Murphy, Medin [39].

[22] About this aspect see specifically Carey [2].

expression. Since concepts are the constituents of beliefs, they depend on the belief system they are connected with and vary according to it.

As already pointed out by several authors on the basis of arguments very different from each other, this semantic theory cannot account for all aspects of categorization. One argument for this has already been mentioned: if we want to account for the possibility to learn a first language, we need to admit that children are able to produce on the basis of their experience some primary internal categorization – a primary I-semantics – which they can use to make hypotheses about the possible groupings underlying the language they are learning. Once they produce some primary grouping strongly related to the perceptive experience, linguistic learning proceeds: this prelinguistic "conceptual core" can be specified in different ways making more and more precise and abstract differentiation through the adding of features carried by the linguistic meanings (the E-language) and by the beliefs about the world it expresses[23].

This way to interpret the relation between I- and E-semantics implies already a very precise theory of concept according to which the conceptual system of adult subjects (described by the I-semantics in the form it assumes after linguistic learning) will be characterized by a double structure. Concepts are supposed to consist on the one hand of a core corresponding to the primary grouping produced by the prelinguistic categorization, and on the other hand of a periphery of cultural features imported from the E-semantics which cover and specifies this core.

Differently from the periphery of the concept, the core cannot be seen as an articulated and interlinked body of knowledge, as the *theory-theory* suggests. We think that the explanation of the core needs to rely on a different semantic theory of prototypical character. We cannot enter here into details about the kind of theory of concepts we consider to be best in order to explain how the core works[24]. For the aims of this work it is sufficient to point out how – according to our view – the relationship between E- and I-semantics imposes the embracement of a form of *Dual Theory able to account for both the components of concepts, the cognitive core connected with the perceptual dimension of instances and a theoretical periphery made of more complex features added to the core by language.*

5   A Semantic Model for the Moral Judgment

The view on concepts we proposed here is just a draft of a theory of concepts. Still the elements we brought into the discussion are enough for the aim of showing how moral issues can be approached using a semantic theory. In order to introduce this aspect it is useful to go back to the conclusions we

---

[23] For an articulated critique to this view in relation to the problem of language acquisition see Dellantonio [8, Ch. IV].

[24] About this aspect see Dellantonio, Pastore [9]; Dellantonio, Pastore [10].

drew from Sripada and Stich's thesis: once an act is categorized as "murder" it has already been judged as morally wrong, while *admissible* voluntary acts which make a person die are not categorized as "murder", but otherwise (e.g. as executions, as self-defence etc.). What a theory of concepts can help to understand in this regard is first and foremost *when (under which conditions) a form of killing is categorized as "murder"*.

The concept of "murder" has indubitably at its core the idea of life and of the loss of life; this is the idea of death. However, not any death is linked to the concept of "murder"; "murder" implies the idea of that particular form of loss of one's life which is caused by an external voluntary intervention. Furthermore, "murder" differs from "killing": any living being (such as a mosquito) can be killed and the act of killing does not need to be done by a human being (for example a rock falling from a cliff can kill someone); "murder" denotes, on the contrary, only the killing of people by people. A murder can be characterized by different degrees of intentionality (and can be qualified for example as voluntary, involuntary or justifiable); however it describes an avoidable action carried out by a human being at the expense of another human being which leads to the loss of life. Since human beings perceive life as something extremely desirable and positive, whereas they perceive the loss (of anything) as something negative, the core of the concept of murder already carries a very "negative value", which is heightened even more by the fact that the loss caused by a murder is perceived as voluntary and evitable.

This observation permits the addition of another element that has been omitted until this point concerning the emotional dimension of moral judgments. If we admit that some perceptive elements of concepts are characterized by "values" which can assume both a positive or a negative sign, we may well suppose that these values comes along with corresponding emotions which match them in intensity and orientation. In respect to the example we are considering this means that if an act is categorized as "murder", *then this act is perceived as having a negative value which comes along with a corresponding negative emotion.* Furthermore, since this value and the corresponding emotion are connected with the elements of the conceptual core – i.e. with the loss of the life by means of a voluntary act – then we can suppose that all concepts characterized by this core have the same negative value coming along with the same negative emotions even though they can cover this core with a periphery made of further cultural features, which are supposed to justify or dignify this act (think about concepts like "execution" or "human sacrifice").

To clarify this idea, showing how not only the features of the core but also the features of the periphery contribute to determining the value of a concept and the emotion that comes along with it, it is useful to consider two further examples, namely the derogatory concepts of "Jew" and "nigger". The concepts of "Jew" and "nigger" have a common core, which is actually shared by all concepts defining human beings to a particular race or descent. This core

is made of perceptible characteristics belonging to all human beings such as their "physical form" (legs, arms, head etc.) and their salient behavioral traits (like the movements and the reactions to the situations which are typical just of humans)[25]. This common core must not be over-interpreted, in the sense that we cannot attribute to it any complex and morally binding characteristics, since it is supposed to consist just of perceptible features. Remaining as neutral as possible, we can say that the features this core is composed of are those which allow us to recognize other people as *conspecifics*[26].

In the hypothesis we are considering the categorization of subjects as "Jews" or "niggers" is due to a specific cognitive operation which couples to itself a neutral perceptive core with a periphery of negative cultural features. Whereas in the case of "nigger" this cognitive operation is carried out setting apart an element of the core relating to a specific distinctive somatic trait (*the dark skin* of this conspecific) and associating the negative features to those specific element, in the case of "Jew" the core does not have any distinctive perceptible characteristics[27], so the negative features typically associated with this concept in its derogatory use are just linked to other non-perceptible feature of the periphery like the race, the descent or the religion.

What the discussion of these examples suggests is on the one hand the cognitive system of humans allow them to recognize spontaneously other humans as conspecifics, i.e. as similar to themselves and as belonging to the same class. Still, on the other hand, the fact that the core of a concept can be associated with any kind of cultural peripheral features rules out the conclusion that – since someone is perceived as conspecific – he must also be perceived as a "person" (i.e. as a subject with binding rights). Indeed, as history teaches us, when derogatory cultural features are added to the core, it can occur that other humans are seen as conspecifics of inferior dignity. Nevertheless, even though it is possible to denigrate specific human subjects or groups through all kinds of negative features, according to the view we are putting forward it is impossible to categorize a human being as something totally non-human. The idea that concepts are characterized by a perceptive core which, in this case, consists of human features, imposes restrictions to the possible categorizations and therefore also opposes semantic relativism according to which everything can be seen as everything, depending on the feature we attribute to it.

This remark has important consequences with regard to moral judgment. Indeed if we admit (on the basis of an independent argument which cannot be

---

[25] About this aspect see Dellantonio, Pastore [9].

[26] About this see also Dellantonio, Pastore [9]; Dellantonio, Pastore [11].

[27] The "distinctiveness" of the features is always measured in relation to the characteristics of the group who carried out the categorization.

presented here[28]) that the recognition of something as a conspecific triggers off empathic responses of some kind, then we can claim that an empathic response will be always triggered off when an instance is categorized through a concept whose core is constituted by the feature "conspecific", even in case of derogatory concepts like "Jew" or "nigger". Still, in the case of derogatory categorization, such a response is suffocated by other opposite reactions connected with the negative features of the periphery.

If we consider the example of "nigger" from this perspective we can conclude that this is characterized by a positive value of the core "conspecific" which is connected with a corresponding positive emotional reaction and by a negative value of peripheral features which is connected with a corresponding negative emotional reaction. In the case of "nigger" the negative value of the peripheral features and the negative emotions that come with it preponderate over the core and over its positive value. The fact that we categorize someone through the derogatory concept "nigger" does not imply any moral judgment by itself, but if a moral reasoning concerns someone categorized as "nigger", the negative value of the concept will influence the overall result of the reasoning.

The argument we are proposing here suggests that the concepts subjects use when they reason about morally relevant situations can be decomposed into the features they consist of and that each of these features can be evaluated as being positive or negative and associated with a corresponding emotion. Reconsider the examples proposed previously: life carries a positive value, the loss of life carries a negative value, greed (typically associated with the prejudice against Jews) carries a negative value, and lack of intelligence (typically associated with the prejudice against black people) carries a negative value. If we determine how a subject categorizes the elements of morally relevant situations and then describe the features of the concepts he uses, then we can weigh up the positive and negative values of these features and have a precise indication of the way in which he perceives the situation from a moral point of view, namely if he is more inclined to see it as morally right or morally wrong. Moreover, since the value of the features goes along with corresponding emotions, the analysis of the value of the features will also give a reliable indication of the emotions triggered off by the situation. Indeed, since reasoning is carried out on the basis of propositions, which consists of concepts, if morally relevant concepts could be described in detail in terms of their features and of their specific value obtained, by weighing up all positive and negative values of the single features these features and values could be used to develop a *semantic model of moral reasoning*. Moreover, weighing up all positive and negative values of the single features will also allow us to determine the emotional orientation of the subject toward the conceptualized objects. *Such a model could therefore be instrumental in*

---

[28] An argument of this kind could for example be developed on the basis of the connection between moral sense and perspective taking: see Gordon [17] and Goldman [16].

establishing whether a situation will be judged and perceived – cognitively and emotionally – as morally right or morally wrong and why.

For this model it is of great importance to take into consideration the distinction between a core and a periphery of concepts. Indeed, even though we do not believe that the features of the core should be weighted more than the features of the periphery, the features of the core still take in some sense priority over the periphery, since the core remains stable over time and has to be considered universal, while the periphery is culturally determined. In fact, it is because of the intersubjective and intercultural stability of the core that the semantic model we are proposing becomes capable of explaining why sometimes people can disagree with their own "corporative moral", which is produced by the conventional norms of the group, they belong to. To clarify this let us return to the examples of "nigger" or "Jew". Even though a group conveys to its members that all black people have to be categorized as "nigger" or that all people of a certain religion have to be categorized as "Jew"(meant in a derogatory sense), still the member of this group cannot avoid perceiving "niggers" and "Jews" as conspecifics. The fact that they are always identified as conspecifics can be used to overcome prejudice since it ensures the possibility to re-categorize them through different peripheral features: it ensures, for example, the possibility to consider *on the basis of a conscious principled reasoning* that – since they are conspecifics – they must have the same basic biological features white people have and must therefore be considered "persons". Once you re-categorize black people or Jews as "person", the overall reasoning about the morality of something like racial laws, extermination, slavery etc. changes completely. Furthermore, in our view the same change also affects the emotions involved in the consideration of these situations explaining how it happens that in certain cases reflection drives or changes the course of emotions[29].

We do not have an answer for the question about why only certain people but not others in certain cases produce categorizations which are different from the ones conveyed by their own culture. The processes that are going on in these cases are most probably very similar to the ones that lead to the introduction of new categorizations in scientific theories and therefore to the change of previous scientific theories. Still, the matter we are actually interested in is just the kind of reasoning involved in the re-categorization of morally relevant concepts. This reasoning is indeed not just purely abstract and based on deliberate reflections about principles or rules, but it is a much more "embodied" form of thinking in which the subject is lead to change perspective and see things differently on the basis of information he already has by decomposing it and recomposing it differently. According to this interpretation moral reasoning is not a form of reflection completely open or free from any cognitive constraint, on the contrary it describes a cognitive procedure

---

[29] These reflections are similar to the tradition of research about dehumanization and moral disengagement which becomes established in the 1970s and is still one of the strong points of social psychology.

of information processing[30]. Nevertheless, it remains a form of reasoning in the sense that it is a non-automatic and conscious processing of information.

6 Concluding Remarks

The central thread of the discussion of this paper can be traced back to the thesis widely shared in the contemporary cognitive research according to which "conscious moral reasoning often plays no role in our moral judgments, and in many cases reflects a post-hoc justification or rationalization of previously held biases or beliefs" [25, p. 25]. The model of moral behavior mostly committed to the idea of a moral mechanism of a psychological kind which – once it is filled with cultural information – automatically produces some sort of moral intuition is the Rawlsian one. We address this model first of all in order to make a critical point about the fact that – if we accept the idea of a moral mechanism – we cannot any longer provide for a plausible explanation of the concrete dynamics of moral judgments in the case of groups whose culture is very different from the one of well-educated Westerners, nor in the case of controversial situations in which people must decide whether to accept as moral the norms of their own community or whether to refuse them.

In the second part of the paper we propose an alternative cognitive theory about morality, according to which our moral positions are judgments stemming from particular forms of conscious reasoning. In the hypothesis we consider these forms of reasoning consisting of specific operations on the concepts and specifically on the features these concepts are made of. In fact, our analysis suggests the possibility to delineate a semantic model of moral reasoning based on the features of the concepts used to categorize the elements of morally relevant situations, which also includes a description of the emotions that come along with moral judgments.

Differently from the Rawlsian model, the semantic model we propose does not conceive moral positions as being entirely determined by cultural and cognitive factors. On the contrary it suggests that the same situation can be categorized in various ways and that the moral judgment of a subject can change depending on how it has been categorized. Even though members of a group are lead to see some particular categorizations as more immediate and natural than others, conscious reasoning may allow in certain conditions to reconceptualize differently specific elements of morally relevant situations by breaking down and recomposing in a new way the perceptual and cultural features of the concepts we initially used to categorize them with. This view has the advantage of supporting both the idea of a psychological procedure underlying moral thinking, and a rationalistic conception of morality

---

[30] This conclusion implies that we still need a cognitive explanation of moral behavior and that we cannot come along with a purely cultural theory of morality. About this aspect see Dellantonio, Pastore [10, pp. 139–178].

according to which moral judgments are produced by conscious and principled reasoning.

References

1. Carey, S.: Conceptual Change in Childhood. MIT Press, Cambridge (1985)
2. Carey, S.: Knowledge acquisition: Enrichment or conceptual change? In: Carley, S., Gelman, R. (eds.) The Epigenesis of Mind: Essays on Biology and Cognition, Hillsdale, pp. 257–291. Lawrence Erlbaum Associates, Mahwah (1991)
3. Carruthers, P.: Moderately massive modularity. In: O'Hear, A. (ed.) Mind and Person, pp. 67–90. Cambridge University Press, Cambridge (2003)
4. Chomsky, N.: Recent Contributions to the theory of innate ideas. Synthese 17, 2–11 (1967)
5. Chomsky, N.: Reflections on Language. Pantheon Books, New York (1975)
6. Chomsky, N.: Knowledge of Language. Its Nature, Origin and Use. Praeger, New York (1986)
7. Chomsky, N.: Language and Problems of Knowledge: The Managua Lectures. MIT Press, Cambridge (1988)
8. Dellantonio, S.: Die interne Dimension der Bedeutung. Externalismus, Internalismus und semantische Kompetenz. Peter Lang Verlag, Hamburg, New York (2007)
9. Dellantonio, S., Pastore, L.: What do concepts consist of? The role of geometric and proprioceptive information in categorization. In: Hanna, P., McEvoy, A., Voutsina, P. (eds.) An Anthology of Philosophical Studies, pp. 91–102. ATINER, Athens (2006)
10. Dellantonio, S., Pastore, L.: Teorie morali e contenuto cognitivo. Cognitivismo, postmoderno e relativismo culturale. In: Meattini, V., Pastore, L. (eds.) Identità, individuo, soggetto, pp. 139–178. Mimesis, Milano (2009a)
11. Dellantonio, S., Pastore, L.: Struttura categoriale e categorizzazione. Un'ipotesi sull'origine della rappresentazione semantica. In: Dellantonio, S., Pastore, L. (eds.) Percezione, rappresentazione e coscienza, pp. 195–230. ETS, Pisa (2009b)
12. Dwyer, S.: Moral competence. In: Murasugi, K., Stainton, R. (eds.) Philosophy and Linguistics, pp. 169–190. Westview Press, Boulder (1999)
13. Dwyer, S.: How good is the linguistic analogy? In: Carruthers, P., Laurence, S., Stich, S. (eds.) The Innate Mind. Culture and Cognition, vol. 2, pp. 237–256. Oxford University Press, Oxford (2006)
14. Dwyer, S.: How not to argue that morality isn't innate: comments on Prinz. In: Sinnott-Armstrong, W. (ed.) Moral Psychology. The Evaluation of Morality: Adaptations and Innateness, vol. 1, pp. 407–418. MIT Press, Cambridge (2008)
15. Frank, R.: Passions within Reason: The Strategic Role of Emotions. Norton, New York (1988)
16. Goldman, A.I.: Simulating Minds. The Philosophy and Psychology, and Neuroscience of Mindreading. Oxford University Press, New York (2006)
17. Gordon, R.: Sympathy, simulation, and the impartial spectator. Ethics 105, 729–742 (1995)
18. Haidt, J.: The emotional dog and its rational trail: A social intuitionist approach to moral judgment. Psychological Review 108, 814–834 (2001)

19. Haidt, J., Bjorklund, F.: Social intuitionists answer six questions about moral psychology. In: Sinnott-Armstrong, W. (ed.) Moral Psychology. The Cognitive Science Of Morality: Intuition and Diversity, vol. 2, pp. 181–217. MIT Press, Cambridge (2008)
20. Heidt, J., Joseph, C.: Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. Daedalus 133(44), 55–66 (2004)
21. Haidt, J., Graham, J.: When morality opposes justice: Conservatives have moral intuitions that liberals not recognize. Social Justice Research 20, 98–116 (2007)
22. Haidt, J., Joseph, C.: The moral mind. In: Carruthers, P., Laurence, S., Stich, S. (eds.) The Innate Mind. Foundations and the Future, vol. 3, pp. 367–391. Oxford University Press, Oxford (2007)
23. Haidt, J., Koller, S., Dias, M.: Affect, culture, and morality, or is it wrong to eat your dog? Journal of Personality and Social Psychology 65, 613–628 (1993)
24. Harman, G.: Explaining Value. Oxford University Press, Oxford (1999)
25. Hauser, M.D.: Moral Minds. How Nature Designed our Universal Sense of Right and Wrong. HarperCollins Publisher, New York (2006)
26. Hauser, M.D., Young, L., Cushman, F.: Reviving Rawls's Linguistic Analogy: Operative Principles and the Causal Structure of Moral Actions. In: Sinnott-Armstrong, W. (ed.) Moral Psychology. The Cognitive Science Of Morality: Intuition and Diversity, vol. 2, pp. 107–143. MIT Press, Cambridge (2008)
27. Hoffman, M.L.: Development of prosocial motivation: Empathy and guilt. In: Eisenberg-Berg, N. (ed.) Development of Prosocial Behavior, pp. 281–313. Academic Press, New York (1982)
28. Hoffman, M.L.: Affective and cognitive processes in moral internalization: An information processing approach. In: Higgins, E.T., Ruble, D., Hartup, W. (eds.) Social Cognition and Social Development: A Socio-Cultural Perspective, pp. 236–274. Cambridge University Press, New York (1983)
29. Hoffman, M.L.: The contribution of empathy to justice and moral judgment. In: Eisenberg, N., Strayer, J. (eds.) Empathy and its development, pp. 47–80. Cambridge University Press, New York (1987)
30. Jackendoff, R.: Languages of the Mind. Essays on Mental Representation. MIT Press, Cambridge (1992)
31. Kagan, J.: The Nature of the Child. Basic Books, New York (1984)
32. Kelly, D., Stich, S.: Two theories about the cognitive architecture underlying morality. In: Carruthers, P., Laurence, S., Stich, S. (eds.) The Innate Mind. Foundations and the Future, vol. 3, pp. 348–366. Oxford University Press, Oxford (2007)
33. Kohlberg, L.: Stage and sequence: The cognitive-developmental approach to socialization. In: Goslin, D.A. (ed.) Handbook of Socialization Theory and Research, pp. 347–380. Rand McNally, Chicago (1969)
34. Kohlberg, L.: From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development. In: Mischel, T. (ed.) Cognitive Development and Epistemology, pp. 151–235. Academic Press, New York (1971)
35. Kohlberg, L.: The Psychology of Moral Development: The Nature and Validity of Moral Stages. Harper and Row, New York (1984)
36. Mahlmann, M.: Rationalismus in der praktischen Theorie: Normentheorie und praktische Kompetenz. Nomos Verlag, Baden-Baden (1999)
37. Mikheil, J.: Rawls' Linguistic Analogy. Ph.D. Thesis, Cornell University Press (2000)

38. Mikheil, J.: Comment on Sripada. In: Sinnott-Armstrong, W. (ed.) Moral Psychology. The Evaluation of Morality: Adaptations and Innateness, vol. 1, pp. 353–359. MIT Press, Cambridge (2008)
39. Murphy, G., Medin, D.: The role of theories in conceptual coherence. Psychological Review 92(3), 289–316 (1985)
40. Nichols, S.: Norms with feeling: Toward a psychological account of moral judgment. Cognition 84, 221–236 (2002)
41. Nichols, S.: Sentimental Rules: On The Natural Foundations of Moral Judgment. Oxford University Press, Oxford (2004)
42. Nichols, S.: Innateness and moral psychology. In: Carruthers, P., Laurence, S., Stich, S. (eds.) The Innate Mind. Structure and Contents, pp. 353–370. Oxford University Press, Oxford (2005)
43. Nucci, L.: Children's conceptions of morality, social conventions and religious prescription. In: Harding, C. (ed.) Moral Dilemmas: Philosophical and Psychological Reconsiderations of the Development of Moral Reasoning, pp. 137–174. Precedent Press, Chicago (1986)
44. Nucci, L.: Education in the Moral Domain. Cambridge University Press, Cambridge (2001)
45. Prinz, J.J.: The Emotional Construction of Morals. Oxford University Press, Oxford (2007)
46. Prinz, J.J.: Is morality innate? In: Sinnott-Armstrong, W. (ed.) Moral Psychology. The Evaluation of Morality: Adaptations and Innateness, vol. 1, pp. 367–406. MIT Press, Cambridge (2008)
47. Quine, W.V.O.: Two dogmas of empiricism. In: Quine, W.V.O. (ed.) From a logical point of view, pp. 20–46. Harvard University Press, Cambridge (1961)
48. Rawls, J.: A Theory of Justice. Belknap Press, Cambridge (1971)
49. Remotti, F.: Contro natura. Laterza, Roma Bari (2008)
50. Samuels, R.: Massively modular minds: Evolutionary psychology and cognitive architecture. In: Carruthers, P., Chamberlain, A. (eds.) Evolution and the Human Mind: Modularity, Language and Meta-cognition. Cambridge University Press, Cambridge (2000)
51. Schweder, R.A., Heidt, J.: The future of moral psychology: Truth, intuition, and the pluralist way. Psychological Science 4, 360–365 (1993)
52. Smentana, J.: Understanding of social rules. In: Bennett, M. (ed.) The Development of Social Cognition: The Child as Psychologist, pp. 111–141. Guilford Press, New York (1993)
53. Smentana, J., Braeges, J.: The development of toddlers' moral and conventional judgments. Merril-Parmer Quarterly 36, 329–346 (1990)
54. Sripada, C.S., Stich, S.: A framework for the psychology of norms. In: Carruthers, P., Laurence, S., Stich, S. (eds.) The Innate Mind. Culture and Cognition, vol. 2, pp. 280–301. Oxford University Press, Oxford (2006)
55. Thomas, L.: Morality and psychological development. In: Singer, P. (ed.) A Companion to Ethics, pp. 464–475. Blackwell, Oxford (2006)
56. Tisak, M.: Domains of social reasoning and beyond. In: Vasta, R. (ed.) Annals of Child Development, vol. II, pp. 95–130. Jessica Kingsley, London (1995)
57. Turiel, E.: The Development of Social Knowledge. Cambridge University Press, Cambridge (1983)
58. Turiel, E.: The Culture of Morality: Social Development, Context, and Conflicts. Cambridge University Press, Cambridge (2002)

59. Turiel, E., Killen, M., Helwig, C.: Morality: Its Structure, functions, and vagaries. In: Kagan, J., Lamb, S. (eds.) The Emergence of Morality on Young Children, pp. 155–244. University of Chicago, Chicago (1987)
60. Wilson, J.Q.: The Moral Sense. Free Press, New York (1993)
61. Zahn-Waxler, C., Radke-Yarrow, M., Wagner, E., Chapman, M.: Development of concerns for others. Developmental Psychology 28, 126–136 (1992)