

Sarah Songhorian

**Neuroethics:
The relation between
philosophical reflection
and empirical research**

1. NEUROSCIENCE OF ETHICS: A METHODOLOGICAL CHALLENGE

The aim of this paper is to focus on the most relevant methodological challenge that neuroethicists have to face when dealing with the neuroscience of ethics (Roskies 2002), that is with the contribution empirical research in general and neuroscience in particular can make to our knowledge of moral issues. Before considering any substantial thesis in the debate, those who are interested in neuroethics have to try and answer a tricky question: how, and to what extent, can empirical analysis modify philosophical theories? That is, how can the way things *are* influence the way they *should* or *ought to be*?

If, for instance, neuroimaging studies would modify our conception of free will or of moral decision-making, one would have to understand first how this interaction is possible.

I will claim that the role of empirical findings is showing the most likely bases upon which certain constructs of the philosophical debate are built. Yet a worry can arise concerning the relation between empirical findings and theoretical analysis and it has deep routes into the philosophical debate. The interpretation of Hume's "is/ought" passage, the division between fact and values (for a discussion see Putnam 2002), and Moore's argument against the naturalistic fallacy (1903) have made any attempt to consider scientific discoveries into the philosophical debate hard.

G.E. Moore's diatribe against the naturalistic fallacy in 1903 set the stage for most of twentieth-century moral philosophy. The main protagonists over the next sixty years were intuitionists and emotivists, both of whom were convinced by Moore that empirical science is irrelevant to moral philosophy and common moral beliefs. Even in the 1970s and 1980s, when a wider array of moral theories entered the scene and applied ethics became popular, few moral philosophers paid much attention to developments in biology and psychology.

This isolation must end. Moral philosophers cannot continue to ignore developments in psychology, brain science, and biology. Of course, philosophers need to be careful when they draw lessons from empirical research. As Moore and his followers argued, we should not jump straight from

descriptive premises in psychology or biology to positive moral conclusions or normative conclusions in moral epistemology. That would be a fallacy. Nonetheless, psychology can still affect moral philosophy in indirect ways (Sinnott-Armstrong 2006, 339).

Firstly, empirical findings and theoretical accounts can bear reciprocal influences. On the one hand, empirical findings can be considered relevant for setting the limits of what we can actually obtain as far as morality and normativity are concerned. Given the individuals that we are and our limits, a better knowledge of our natural endowment can help building a theoretical framework to which we can actually conform. In order for a thing to be something we ought to do, we need to have the capacity and the possibility to make that thing actual. In the economic debate, learning that individuals do not decide with perfect rationality, got a new line of inquiry on limited rationality off the ground. Something similar can happen within the relationship of empirical findings to our view of morality. Discovering that emotions are relevant for our moral judgments and for directing our behaviors, for instance, should be enough to consider any theoretical framework that denies the role of emotions as unnecessary – if any such framework really exists. This is not to argue for a reduction of the normative onto the psychological or biological level; rather, to subscribe to a liberal form of naturalism as setting possibility conditions.

A convincing version of Liberal Naturalism would need to do justice to the range and diversity of the sciences, including the human and social sciences, and to the plurality of ways of understanding, including the possibility that some of these ways are non-scientific yet non-supernatural (De Caro, Macarthur 2010, 9).

Moreover, the influence of empirical findings on theoretical accounts should be indirect inasmuch as, as Sinnott-Armstrong says, we should not jump from them to draw conclusions for normative or moral accounts that rest on a different level of the analysis of the world; rather, we should first try to incorporate those findings into the vocabulary of the latter and see what aspects of our normative account can be confirmed or falsified by them.

On the other hand, as far as the influence of theoretical accounts on empirical findings is concerned, the former can certainly help in clarifying the concepts investigated empirically, in designing the experimental task, and in interpreting the outcomes and results.

Secondly, I think that there has been a misconception of Hume's divide and that several contemporary attempts to overcome it have done a bad service to their praiseworthy purpose, jumping too easily from description to prescription (e.g. Preston, de Waal 2002; Shamay-Tsoory 2011; Rifkin 2009; Baron-Cohen 2011). I believe Hume's concern was the absence of any explanation or justification for the passage from sentences containing "is" and "is not" to propositions connected with "ought" or "ought not" rather than a claim of impossibility of such passages.

I cannot forbear adding to these reasonings an observation, which may, perhaps, be found of some importance. In every system of morality, which I have hitherto met with, I have

always remark'd, that the author proceeds for some time in the ordinary way of reasoning, and establishes the being of a God, or makes observations concerning human affairs; when of a sudden I am surpriz'd to find, that instead of the usual copulations of propositions, *is*, and *is not*, I meet with no proposition that is not connected with an *ought* or an *ought not*. This change is imperceptible; but is, however, of the last consequence. For as this *ought*, or *ought not*, expresses some new relation or affirmation, 'tis necessary that it shou'd be observ'd and explain'd; and at the same time that a reason should be given, for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it. But as authors do not commonly use this precaution, I shall presume to recommend it to the readers; and am persuaded, that this small attention wou'd subvert all the vulgar systems of morality, and let us see, that the distinction of vice and virtue is not founded merely on the relations of objects, nor is perceiv'd by reason (Hume 2007, 302).

It is this sort of precaution that Hume recommends: if it is possible to move from a descriptive sentence to a prescriptive one, the change has to be justified (for an extremely detailed analysis of Hume's law, see Celano 1994). If justification is not at hand, an unwarranted jump, as Sinnott-Armstrong describes it, is in place. Interpreting Hume's quotation as claiming that it is impossible to go from empirical propositions to normative ones, neuroscience of ethics would be impossible. However, it seems an extreme interpretation of Hume's concern that is not grounded on what he actually claims. He just recommends precaution and that is what the neuroscience of ethics should conform to in order to avoid unwarranted jumps from prescription to normativity and yet to be able to use the most recent advances in the knowledge of our moral psychology to better define our range of possibilities as far as normative moral theories are concerned.

2. EMPIRICAL FINDINGS AND ETHICAL THEORY

There are two ways in which one can think empirical findings have an influence on moral theory. The former concerns the support that certain knowledge of our moral psychology and functioning can provide for a specific metaethical theory. The latter regards a connection between the description of our endowment and normative¹ and substantive theories, telling us how we actually should behave.² As far as the latter is concerned, one can hold that there is no direct influence of empirical findings, but only an indirect one

¹ In the following pages, I will use "normative theory" as indicating the domain of morality that concerns norms, rules, duties, or rights. At this level, the interest is on what we ought (or have right) to do. It refers to substantive and particular moral propositions, such as "stealing is wrong", "altruism is right", and the like.

² I will consider here metaethics and normative and substantive theories as two separate levels of moral reflection in order to investigate their relation to empirical findings and yet I will not enter into the debate about the relation between metaethics and substantive theories. The argument presented here is compatible both with the view that they have reciprocal influences and with the one that claims they do not.

based on the idea that what we ought to do (or have a right to do) has to be practically feasible. Moreover, a particular reading of reflective equilibrium³ can help here understand to what extent one can derive normative conclusions from empirical data without properly deriving “ought” from “is”.

As Greene underlines in his recent article (2014) – which replies to many of the objections his studies have raised (Greene 2007; 2005; Greene *et al.* 2009; 2004; 2001; Greene, Haidt 2002) – experiments, of the kind he himself conducted

[...] identify factors to which our moral judgments are sensitive. This information may be combined with independent normative assumptions concerning the kinds of things to which our judgments ought to be sensitive. This combination can lead us to new, substantive moral conclusions. In other words, scientific information can allow us to trade in difficult “ought” questions for easier “ought” questions, and thus advance ethics (Greene 2014, 711).

His subsequent example helps clarify this insight. If one wonders whether juries make good judgments, one can refer to several studies claiming for certain biases in their decisions – as, for instance, the fact that they are sensitive to race. Combining this insight coming from empirical studies with a normative assumption that juries should not be sensitive to race and several other biases, one can derive that at least sometimes juries make bad judgments. From this factual conclusion one can consequently derive that those biases that actually influence juries’ decisions should be minimized by several strategies – for instance, including racial differences in the juries themselves. Three aspects of this way of solving normative questions that are still open are noteworthy. First, this kind of derivation does not represent properly an “is/ought” passage inasmuch as a normative assumption is present also in the premises. Second, the derivation represents

³ “In searching for the most favored description of this situation we work from both ends. We begin by describing it so that it represents generally shared and preferably weak conditions. We then see if these conditions are strong enough to yield a significant set of principles. If not, we look for further premises equally reasonable. But if so, and these principles match our considered convictions of justice, then so far well and good. But presumably there will be discrepancies. In this case we have a choice. We can either modify the account of the initial situation or we can revise our existing judgments, for even the judgments we take provisionally as fixed points are liable to revision. By going back and forth, sometimes altering the conditions of the contractual circumstances, at others withdrawing our judgments and conforming them to principle, I assume that eventually we shall find a description of the initial situation that both expresses reasonable conditions and yields principles which match our considered judgments duly pruned and adjusted. This state of affairs I refer to as reflective equilibrium. It is an equilibrium because at last our principles and judgments coincide; and it is reflective since we know to what principles our judgments conform and the premises of their derivation. At the moment everything is in order. But this equilibrium is not necessarily stable. It is liable to be upset by further examination of the conditions which should be imposed on the contractual situation and by particular cases which may lead us to revise our judgments” (Rawls 1971, 18). As far as moral theory is concerned, Rawls claims that: «we investigate the substantive moral conceptions that people hold, or would hold, under suitably defined conditions. In order to do this, one tries to find a scheme of principles that match people’s considered judgments and general convictions in reflective equilibrium. This scheme of principles represents their moral conception and characterizes their moral sensibility» (Rawls 1999, 288). The kind of reflective equilibrium I am thinking of here is an equilibrium that takes people’s principles and considered judgments into account, together with some facts about the world and about ourselves.

a simplification of the problem posed. Moreover, the conclusion itself refers to an empirical outcome – that is, they actually make bad judgments at least sometimes –, from which one can derive strategies to achieve what was included in the normative assumption – that is, juries should not be sensitive to racial biases. Balancing empirical findings and normative assumptions makes the derivation indirect and represents an application of reflective equilibrium.

As this example illustrates, we can reach interesting normative conclusions by combining interesting scientific facts with uninteresting normative assumptions. However obvious this may seem, some mistakenly assume that empirically based normative arguments are empty or question-begging if they rely on nonempirical normative assumptions. The above example suggests a more general principle: An empirically driven normative argument is non-question-begging if the normative assumptions it requires are less interesting (i.e., less controversial) than its normative conclusion. I am not claiming one can derive a moral “ought” from nothing but a scientific “is”. Rather, my point is that moral psychology matters for ethics, that it is “normatively significant”. Moral psychology matters, not because it can generate interesting normative conclusions all by itself, but because it can play an essential role in generating interesting normative conclusions. A natural objection to this example is that the work done by the science, while not insignificant, is *normatively* insignificant. The science does not challenge anyone’s values. Instead, it simply alerts us to an application of the values we already have (Greene 2014, 711-712).

In order to answer this latest worry, Greene analyzes the example of consensual adult incest. In case we learn, from scientific data, that our disapproval of incest derives from a negative emotional response which aim is to avoid the risk of genetic diseases in the offspring, should we rely on this response in all possible contexts? He considers the example of Joe and Jane, two siblings separated in early childhood who grow apart and eventually fell in love. When Joe and Jane discover that they are siblings, they decide to remain together and take all possible precautions against the risk of producing a genetically ill offspring (i.e., vasectomy). According to Greene, with this example in mind and taken for granted that our disapproval of incest depends only on the emotional response against producing ill offspring, one should review the general condemnation against all possible cases of incest.

Having made this assumption, and having learned something from science, we may now conclude that we ought not condemn all incestuous behavior – an interesting normative conclusion. This example – a classic *debunking* explanation – is notable because it genuinely challenges some people’s moral values. Indeed, such arguments can change people’s minds, if you give them enough time to think (Greene 2014, 712).

As controversial as the example may be, this still represents one possible way in which empirical findings have a role in shaping and changing our normative assumptions in specific cases. No proper “is/ought” passage is met because normative assumptions are also present in the premises of this kind of derivations.

As far as the role of neuroscience is concerned, Greene – contrary to what one may think based on his previous works – is explicit in claiming that findings from imaging studies do not play any special role in ethical thinking, but they are rather just one among many different sources of empirical research that might play the role depicted above (Greene 2014, 716). To this extent, the neuroscience of ethics does not differ radically from moral psychology. The neuroscience of ethics has been considered revolutionary, however, if Greene is right, it does not play any special role. Neuroscience can only provide a new tool to shed some light onto the moral domain.

Finally, as simple as it may seem, our normative account of morality should consider empirical findings inasmuch as they illuminate the basic capacities we are endowed with. If, as it seems likely from empirical studies (Greene *et al.* 2001), emotional responses are relevant for our production of moral judgments, one cannot think of morality as totally non-emotional without providing an account that is not viable for the kind of individuals we are. That is not to say that our epistemological access to morality is itself normative, but rather that, given the way in which we become acquainted with morality, we should prescribe behaviors that are at least possible. Knowledge of our implementational and representational endowment restricts possible theories to viable ones. If I ought to behave altruistically, I must be able to do so. This has to be regarded as a naturalistic attempt inasmuch as it restricts possibilities to what is naturally feasible for humans. A further problem might arise if one focuses only on what is actually praised or blamed, rather on what is praise-worthy or blame-worthy. A naturalistic account has to include something more than what we actually do if it claims to be normative and to be able to establish duties (or rights) as opposed to simply describing what we already do.

The naturalist [...] supposes that no duty can be established on grounds independent of what men do in fact do; and, when what men do in fact do shows a duty to make excessive demands, the duty is declared no duty at all (Brown 1950, 276).

If one focused only on what is actually praised or blamed in a certain society at a given time, then morality would be relativistic and no effort would be necessary for humans to judge and act morally. I think that is not the case. Naturalism can have different forms. The one provided by Brown is neither the kind I am interested in nor the primary interest of neuroethics in general. Among several other problems, if this naturalistic perspective were the only one possible, morality would be nothing more than a generalization over actual behavior. There would be nothing special about moral reasoning and moral disagreement – it would be only a disagreement about what we actually do or do not.

Moreover, several accounts can avoid the denial of a role to personal engagement and deliberation in a reductionist manner. It is not the case that, given certain basic

abilities, one is *determined* to behave in a specific way. Referring to empirical findings in this context is only a way to test which theories are more viable and to clarify the set of possibility conditions for an account.

As to the connection between empirical findings and metaethical theories, there is no proper passage from factual premises to normative conclusions inasmuch as metaethics does not deal with ethical substantive “ought”. In Joyce’s words:

Even if there were an *a priori* prohibition on deriving evaluative conclusions from factual premises, this need not stand in the way of *metaethical* implications being drawn from factual premises, for a metaethical claim is not an ethical “ought” claim; it is more likely to be a claim about how we use the word “ought” in ethical discourse – which is a perfectly empirical matter (Joyce 2008, 371).

If any sense of “ought” must be applied to metaethics, it is a linguistic one, a second-order reflection concerning how we should apply moral concepts, under what conditions those concepts are properly used.

In so far as some metaethicists offer prescriptions about how the word “ought” *ought* to be used, metaethics sometimes steps beyond the descriptive. Even in such cases, however, metaethicists are still not pushing *ethical* “ought” claims (Joyce 2008, 371).

As for normative conclusions, there is a sort of balance between empirical findings and linguistic normative assumptions that rest on the premises’ level before any conclusion about how we should apply the word “ought” can be made. So, in this respect, there is no illicit passage from empirical premises to normative conclusion, even when the normative concept is a linguistic one.

A further role empirical findings might play in choosing a metaethical account resembles one of the roles identified earlier. If empirical data support a role for emotions, as well as for higher order rationalization, in our moral judgment, then a hybrid account of ethics can better explain the way in which we actually learn about morality, they can illuminate our epistemological access to moral properties. If data from neuroimaging studies underline that both emotions and rational engagement are simultaneously present when a moral deliberation occurs, then the metaethical theory that has to be favored is one among those that recognize both an emotional and a rational component to moral deliberation. Any theory that advocates moral deliberation to be uniquely emotional or uniquely rational should be considered unfeasible. It is not the case that a single experiment, or even a set of significant ones, can prove a thesis on the metaethical level, however it can reduce the spectrum of theories that can be considered feasible by falsifying other theoretical options – for instance, those that deny completely a role for either emotions or reasoning.

3. THE POWER OF EMPIRICAL FINDINGS

Another noteworthy consideration regards the power of empirical finding *per se*. As Sinnott-Armstrong said, we should not jump from empirical evidence to normative conclusions, not only because of a certain philosophical caution relative to Hume's law, but also for independent reasons. Firstly, empirical data are most of the times correlational studies – with some exceptions like behavioral genetics –, so that interpreting their conclusions as providing a comprehensive explanation of our behavior is an overestimation of their powers. As Hume rightly pointed out, precaution is the key. On the one hand, correlational studies do not identify causes; and, on the other, different theoretical approaches can be used to interpret and make sense of exactly the same set of data. Secondly, given the different vocabularies that science and philosophy adopt, it is not always easy to understand exactly what empirical findings can actually contribute to. If one is not a reductionist and, thus, if one does not believe that there are correspondence rules between philosophical and scientific statements (Nagel 1961), then some reflection is needed before claiming that certain data support – they do not demonstrate in any case – any philosophical thesis. These two further difficulties explain why, on the one hand, several findings are taken to support antithetic perspectives and, on the other, why, even though we have gained some further knowledge about how we judge, there is still room for philosophical enquiry and debate. Philosophical and empirical enquiry should cooperate to pursue a better understanding of human ethics.

4. A NON-REDUCTIONIST NATURALISM

Subscribing to a naturalistic account of ethics does not amount to committing oneself to reductionism or to the impossibility of finding a place for normativity (Prinz 2007, 2). There are four understandings of naturalism according to which one can advocate for it without incurring these two risks (Prinz 2007, 2-3). As far as naturalism is concerned, one can be interested in metaphysical, explanatory, methodological, or transformation naturalism. All these forms of naturalism have several different versions. I will not deal with them here specifically, in particular I will not discuss transformation naturalism. My aim is just to show that one can advocate a non-reductionist naturalistic account.

As far as metaphysical naturalism is concerned, one can simply claim that no supernatural entities exist.

It is the view that our world is limited by the postulates and laws of the natural sciences. Nothing can exist that violates these laws, and all entities that exist must, in some sense, be composed of the entities that our best scientific theories require. This is a metaphysical thesis; it concerns the fundamental nature of reality (Prinz 2007, 2).

One may wonder how metaphysical naturalism can avoid being reductionist, on the one hand, and maintain the existence of moral properties, on the other. As the latter

worry is concerned, our epistemic difficulties in accessing moral properties – as opposed to other real entities – should not matter when dealing with metaphysics, with their existence. Moreover, if one considers what entities actually exist from a scientific perspective – say, atoms and electrons – the epistemological issue reveals itself trickier than expected also as far as scientific entities are concerned.

The naturalist offers to save realism but eliminate the mystery: the domain of moral value is to be seen as simply a part of the familiar natural world, known about in just the familiar, broadly empirical ways we know about the natural world. While for other views there is a gulf to bridge between the domain of prosaic natural facts and the domain of values, the moral naturalist seeks to collapse this distinction and reveal value to us as straightforwardly part of the domain of natural fact (Lenman 2014).

Collapsing such a distinction does not, however, necessarily commit to reductionism. Moral properties can emerge or supervene onto natural facts and they can be analogous to natural facts in many relevant respects. Moral properties are properties of the world we live in, the natural one. We access them as we access other secondary qualities. They emerge or supervene on natural properties similarly to the way in which any other secondary property – as colors or emotions – supervene or emerge on them.

A minimal version of metaphysical naturalism refers simply to the fact that we live in a natural world and the experiences we have are all in a sense natural (Lecaldano 2010, 25). From such a perspective there is no need to postulate supernatural entities to account for a version of moral realism. Moreover, avoiding reductionism is possible through the concepts of emergence or supervenience.

Metaphysical naturalism entails a kind of explanatory naturalism. If everything that exists is composed of natural stuff and constrained by natural law, then everything that is not described in the language of a natural science must ultimately be describable in such terms. This is not equivalent to reductionism in the strong sense of that word. Strong reductionists say that the relation between natural science and “higher-level” domains is deductive. We should be able to deduce higher-level facts from their lower-level substrates. Antireductionists deny this. They think, for example, that there are higher-level laws or generalizations that could be implemented in an open-ended range of ways. Regularities captured at a low level would miss out on generalizations of that kind. [...] The explanatory naturalist does not need to claim that low-level explanations are the only explanations (Prinz 2007, 2).

One can, thus, endorse explanatory naturalism without being committed to reductionism.

To this extent, it will be useful to consider Marr’s distinction of different levels of explanation (Marr 1982, 24-25). Marr distinguishes between a computational, an algorithmic, and an implementational level. In order to introduce these levels, he uses the example of a cash register. The first and more abstract level of explanation is that of a computational theory, which aims at understanding *what* the device does and *why*. Following Marr’s example, the cash register sums. The *what*-question is answered by a theory of

addition. The *why*-question is answered by considering the reason for choosing that particular operation – addition – instead of other possible options. It is answered by a series of constraints.

In order that a process shall actually run, however, one has to realize it in some way and therefore choose a representation for the entities that the process manipulates. The second level of the analysis of a process, therefore, involves choosing two things: (1) a *representation* for the input and for the output of the process and (2) an *algorithm* by which the transformation may actually be accomplished. For addition, of course, the input and output representations can both be the same, because they both consist of numbers. [...] If the first of our levels specifies what and why, this second level specifies *how*. For addition, we might choose Arabic numerals for the representations, and for the algorithm we could follow the usual rules about adding the least significant digits first and “carrying” if the sum exceeds 9. Cash registers, whether mechanical or electronic, usually use this type of representation and algorithm (Marr 1982, 23).

The last level is the implementational one; in the case of the cash register it is the specific machine that embodies physically the algorithm. As Marr underlines, the algorithm that a child uses to add two numbers might well be the same used by the cash register, but obviously the implementation is rather different. Some devices might be more apt for certain algorithms. So that when *designing* a program, it might be that the choice of either the implementation we want to use or the algorithm that best suits our purposes commits to certain choices at the other level. This, obviously, holds only when the aim is creating a device and not when the aim is *explaining* an existent device that was not produced by us – like in the case of vision, Marr’s concern, or any other cerebral device.

Once a distinction between different explanatory levels has been proposed, though, a question concerning the relations between them may arise. Marr explains how he believes the three levels relate to each other as follows:

[...] there is a wide choice available at each level, and the explication of each level involves issues that are rather independent of the other two.

Each of the three levels of description will have its place in the eventual understanding of perceptual information processing, and of course they are logically and causally related. But an important point to note is that since the three levels are only rather loosely related, some phenomena may be explained at only one or two of them. This means, for example, that a correct explanation of some psychophysical observation must be formulated at the appropriate level. In attempts to relate psychophysical problems to physiology, too often there is confusion about the level at which problems should be addressed (Marr 1982, 25).

The three levels are logically and causally related – as mentioned above, when designing a device some constraints from the implementational level might guide the choice for a certain algorithm to be implemented, or it might be the other way around – yet, they

are to some extent independent from one another. One can advocate a computational theory of vision without entering details about how it is represented or implemented.

Distinguishing between different levels of explanation and maintaining explanatory naturalism, one can account for the fact that generalization on higher levels – representational or computational levels – can provide explanatory elements that would be missed if one focuses only on the low-level description. Moreover, maintaining different explanatory levels can also account for multiple realizability at the implementational level. For instance, if the description of the way we judge morally rests on a representational level of analysis, one can claim that there might be different possible implementations for it to occur. Neuroimaging studies can enlighten humans' activations when they face moral dilemmas. However, in case computers or cyborgs became possible, the implementational level of explanation would differ, while the representational one could remain identical. The same can be said of inhabitants of other planets who might be wired differently from us. Keeping representational and implementational explanations apart can account for different possible implementations of the same representation: just as much as, in Marr's case, the algorithm "addition" can be implemented by certain operations of a person's mind, but also by some operations of a cash register. These two different devices constitute differences in the implementation, but not in superior levels of explanation – say, representational and computational.

Methodological naturalism claims that:

If all facts are, in some sense, natural facts (according to metaphysical naturalism), then the methods by which we investigate facts must be suitable to the investigation of natural facts. Philosophers sometimes claim to have a distinctive method for making discoveries: the method of conceptual analysis. If metaphysical naturalism is true, this cannot be a supernatural method of discovering supernatural truths. [...] Conceptual analysis is, like all legitimate investigatory tools, an empirical method (Prinz 2007, 3).

Finally,

Each form of naturalism has implications for normativity. Metaphysical naturalism entails that moral norms, if they exist, do not require postulating anything that goes beyond what the natural sciences allow. Explanatory naturalism entails that we can ultimately describe how any moral norm is realized by natural entities. Methodological naturalism entails that we should investigate norms using all available empirical resources tools. Transformation naturalism entails that we must investigate norms from within our current belief systems, and, as a result, the norms we currently accept will influence our intuitions about what norms we ought to uphold. If we choose to change our norms, we cannot do so by adopting a transcendental stance that brackets off the norms we currently accept (Prinz 2007, 3).

If these non-reductionist naturalisms are correct, then:

If the world includes facts about what ought to be, those facts must be explicable in terms of how things are. Every ought must supervene on an is (Prinz 2007, 4).

According to Prinz, naturalism does allow to infer an “ought” from an “is”.

The debate on Hume’s law, as well as the one concerning naturalism, should be analyzed in much more detail than it has been done here. However, the intention was just to show that naturalism is not necessarily committed to reductionism and to point at possible interactions between different research fields that can – and should, in my opinion – cooperate to achieve a better understanding of moral behavior and moral conscience, resting on different explanatory levels. I believe the representation of these problems provided here could guide further investigations and represent a way to achieve such cooperation.

5. CONCLUSION

One concrete example of an over-interpretation of data coming from empirical research is the role that has been attributed to empathy in the moral domain. Several authors have considered it as the basis for our moral capacity jumping too easily from the description of a neural device we are endowed with to normative conclusions (e.g. Preston, de Waal 2002; Shamay-Tsoory 2011; Rifkin 2009; Baron-Cohen 2011). They have claimed that since we are to some extent endowed with empathic abilities all that is necessary in order to act and judge morally is that empathy is activated. However, being empathy simply a device that enables the resonance of others’ emotions, there is nothing intrinsically moral about it. Empathy can provide the natural and embedded basis for our ability to understand those who are similar to us, but it will not be moral until we somehow decide that we are going to take the emotions of others as relevant for our moral judgment and decision-making and to give a precise direction to this consideration. As long as empathy is just an unreflective affective reaction, it is far from being morally relevant, and tells us little on how we should behave towards others. Before adopting and endorsing a more impartial and detached consideration, empathy is simply amoral (Carrasco 2011; Bloom 2014).

The innate capacity is, thus, necessary for the sympathetic consideration of other people – or compassion, as Paul Bloom (2014) calls it –, but it is not a sufficient condition to make that relationship moral: having such a mechanism is totally compatible with making immoral decisions and judgments – it is, for example, totally compatible with some degree of *Schadenfreude*.

Schadenfreude – derived from the German from *Schaden* (adversity, harm) and *Freude* (joy) – means taking enjoyment in another’s pain or misfortune. *Schadenfreude* is the flip side of empathy: similar cognitive abilities are used (e.g., perspective taking, awareness of others’ emotions), but the result is not sympathy or a desire to help another person but rather a sense of pleasure when another person experiences pain or distress. [...] *Schadenfreude* is related to envy; brain imaging studies show that when a person feels envy, brain regions associated with physical pain (i.e., the anterior cingulate cortex) are activated. When a person one envies

experiences a misfortune, triggering feelings of *Schadenfreude*, the ventral striatum (a brain region associated with pleasure and reward) is activated [Takahashi *et al.*, 2009] (Reevy *et al.* 2010, 237, italics in original).

If we lack such a automatic mechanism for sharing other people's emotions, it is much more complicated to understand them, though it is not impossible by means of a more complex and mentalizing mechanism that cannot provide an immediate and emotional sharing, but an intellectual comprehension through the attribution of mental states to others. There are other ways to perform pro-social behavior that do not entail empathy (Slote 2010, Prinz 2007). Similarly, there are other more mentalizing ways to understand others that do not entail empathy. And yet, even when the mechanism is in place, it does not necessarily follow that the behavior or decision taken upon it would be morally good. *Schadenfreude* represents the counterexample of this inference: empathy – that is, the capacity to resonate others' emotions is in place and yet the behavior or decision that follows is an immoral one.

The neural circuits give us a possibility. The fact of having such a circuit does not mean that it is necessary to develop it in a desirable way: the realization of oneself as a sympathetic, compassionate, and caring being is totally open. Empirical findings about our empathic abilities only illuminate the set of possibility's conditions of the emergence of a moral conscience and moral capacities (Boella 2008, 43). Empathy has here been used only as an example of what inferences can and cannot be drawn from empirical findings. The existence of a certain mechanism requires our theoretical approaches to take it into consideration, that is to grant some room to it, and yet it cannot determine, by its existence *per se*, moral decisions or moral behaviors.

The neuroscience of ethics has, thus, to face a serious methodological challenge when trying to derive normative, substantive conclusions from empirical findings. As it has been shown, in particular with the example of empathy, it is not the case that empirical research cannot illuminate our understanding of the moral domain, yet, as Hume claims, anytime one goes from an "is" to an "ought" such passage has to be explained and justified. Data coming from neuroscience, as much as from any other empirical domain, can play an indirect role in modifying our theoretical understanding of morality. What they cannot do is prove one specific theory, their role is simply that of limiting possible theories to feasible ones. In the case of the research on empathy, for instance, what the huge amount of data collected can tell us is that empathy plays a role in our understanding of others, so that a theoretical approach that denies completely such a component in morality does not count as a feasible one. However, it is not by data alone that we can decide among several theoretical options that recognize a role to empathy.

Understood in this way – that is, as a combination of empirical research and philosophical insight to enlighten our actual possibilities in the moral domain –, the neurosci-

ence of ethics can avoid being reductionist without committing to non-naturalism. Our ethical questions will not find a definite answer in empirical data, but simply another set of information to be interpreted and used to better understand the constraints we are subject to. This account of the interaction of empirical findings with theoretical enquiry and of the role data can actually play should restrict the fear and criticism to neuroethics that have come from those who fear a reductionist move.

REFERENCES

- Baron-Cohen S. (2011), *Zero Degrees of Empathy: A New Theory of Human Cruelty*, Penguin, London
- Bloom P. (2014), "Against Empathy", *Boston Review. A Political and Literary Forum*, <http://www.bostonreview.net/forum/paul-bloom-against-empathy> [15 February 2015]
- Boella L. (2008), *Neuroetica. La morale prima della morale*, Milano, Raffaello Cortina
- Brown S.M. jr. (1950), "Does ought imply can?", *Ethics*, 60, n. 4, pp. 275-284
- Carrasco M.A. (2011), "From sympathy to moral sympathy", *Adam Smith Review*, n. 6, pp. 7-29
- Celano B. (1994), *Dialettica della giustificazione pratica. Saggio sulla legge di Hume*, Torino, Giappichelli
- De Caro M., Macarthur D. (2010), *Naturalism and Normativity*, New York, Columbia University Press
- Greene J.D. (2005), "Cognitive neuroscience and the structure of the moral mind", in P. Carruthers, S. Laurence, S.P. Stich (eds), *The Innate Mind: Structure and Contents*, New York, Oxford University Press, pp. 338-352
- (2007), "Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains", *Trends in Cognitive Sciences*, vol. 11, n. 8, pp. 322-323
- (2014), "Beyond point-and-shoot morality: Why cognitive (neuro)science matters for ethics", *Ethics*, 124, n. 4, pp. 695-726
- Greene J.D., Cushman F.A., Stewart L.E., Lowenberg K., Nystrom L.E., Cohen J.D. (2009), "Pushing moral buttons: The interaction between personal force and intention in moral judgment", *Cognition*, vol. 111, n. 3, pp. 364-371
- Greene J., Haidt J. (2002), "How (and where) does moral judgment work?", *Trends in Cognitive Sciences. Web*, vol. 6, n. 12, pp. 517-523
- Greene J.D., Nystrom L.E., Engell A.D., Darley J.M., Cohen J.D. (2004), "The neural bases of cognitive conflict and control in moral judgment", *Neuron*, vol. 44, n. 2, pp. 389-400
- Greene J.D., Sommerville R.B., Nystrom L.E., Darley J.M., and Cohen J.D. (2001), "An fMRI investigation of emotional engagement in moral judgment", *Science*, vol. 293, n. 5537, pp. 2105-2108
- Hume D. (2007), *A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects* [1739], critical edition D.F. Norton, M.J. Norton (eds), New York, Clarendon Press
- Joyce R. (2008), "What neuroscience can (and cannot) contribute to metaethics", in W. Sinnott-Armstrong (ed.), *Moral Psychology Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, Cambridge, The MIT Press, pp. 371-394
- Lecaldano E. (2010), *Prima lezione di filosofia morale*, Roma-Bari, Laterza
- Lenman J. (2014), "Moral naturalism", in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/archives/spr2014/entries/naturalism-moral/> [15 February 2015]

- Marr D. (1982), *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*, New York, Freeman
- Moore G.E. (1993), *Principia Ethica* [1903], T. Baldwin (ed.), Cambridge, Cambridge University Press
- Nagel E. (1961), *The Structure of Science: Problems in the Logic of Scientific Explanation*, London, Routledge & Kegan Paul PLC
- Preston S.D., de Waal F.B.M. (2002), "Empathy: Its ultimate and proximate bases", *The Behavioral and Brain Sciences*, vol. 25, n. 1, pp. 1-20; discussion pp. 20-71
- Prinz J. (2007), *The Emotional Construction of Morals*, Oxford, Oxford University Press
- Putnam H. (2002), *The Collapse of the Fact/Value Dichotomy and Other Essays*, Cambridge, Harvard University Press
- Rawls J. (1971), *A Theory of Justice*, Cambridge, Harvard University Press
- (1999), *Collected Papers*, edited by S.R. Freedman, Cambridge, Harvard University Press
- Reevy G., Ozer Y.M., Ito Y. (2010), *Encyclopedia of Emotion*, Santa Barbara, Abc-Clio
- Rifkin J. (2009), *The Empathic Civilization: The Race to Global Consciousness in a World in Crisis*, Cambridge, Polity Press
- Roskies A. (2002), "Neuroethics for the new millenium", *Neuron*, n. 35, pp. 21-23
- Shamay-Tsoory S.G. (2011), "The neural bases for empathy", *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, vol. 17, n. 1, pp. 18-24
- Sinnott-Armstrong W. (2006), "Moral intuitionism meets empirical psychology", in T. Horgan, M. Timmons (eds), *Metaethics After Moore*, Oxford, Clarendon Press
- Slote M. (2010), *Moral Sentimentalism*, New York, Oxford University Press
- Takahashi H., Kato M., Matsuura M., Mobbs D., Suhara T., Okubo Y. (2009), "When your gain is my pain and your pain is my gain: Neural correlates of envy and Schadenfreude", *Science*, vol. 323, n. 5916, pp. 937-939

