RICERCHE

The Methods of Neuroethics: Is the Neuroscience of Ethics Really a New Challenge to Moral Philosophy?

Sarah Songhorian $^{(\alpha)}$

Ricevuto: 23 novembre 2018; accettato: 5 marzo 2019

Abstract Within the otherwise lively debate on neuroethics, little attention has been devoted to the peculiar methodological issues and challenges it faces. My aim is to track down its methodological specificities. Firstly, I will investigate to which traditional debates neuroethics bears similarity and to what extent it actually represents a novelty in ethical thinking. While the ethics of neuroscience is akin to bioethics, the neuroscience of ethics seems akin to moral psychology. And yet they differ as far as the level of explanation of human moral behavior and reasoning is concerned. Thus, while the neuroscience of ethics and moral psychology share a family resemblance, they cannot be reduced to one another. Secondly, I will explore three different philosophical temperaments towards the role empirical findings can and should have in normative ethics and in metaethics. Prudential reasons would recommend openness to new sources of evidence without risking either reductionism nor neglect.

KEYWORDS: Neuroethics; Methods; Moral Psychology; Empirical Ethics; Levels of Explanation

Riassunto I metodi della neuroetica: la neuroscienza dell'etica è davvero una nuova sfida per la filosofia morale? – Nonostante il dibattito in ambito neuroetico sia molto vivace, poca attenzione è stata dedicata alla questione relativa al suo metodo specifico e ai problemi a cui essa va incontro. Il mio scopo è quello di identificare le specificità metodologiche di questa disciplina. Anzitutto cercherò di determinare a quali dibattiti tradizionali la neuroetica può essere avvicinata e in che misura invece questa costituisce una novità all'interno del pensiero etico. Mentre l'etica della neuroscienza ha delle affinità con la bioetica, la neuroscienza dell'etica appare maggiormente simile alla psicologia morale, sebbene tali discipline siano diverse fra loro per via del livello a cui spiegano il comportamento e il ragionamento morale degli esseri umani. Pertanto, nonostante la neuroscienza dell'etica e la psicologia morale rivelino una certa somiglianza di famiglia, non è possibile ridurre l'una all'altra. In secondo luogo esplorerò tre diversi atteggiamenti filosofici rispetto al ruolo che i risultati empirici giocano o dovrebbero giocare rispetto all'etica normativa e alla metaetica. Ragioni di prudenza suggeriscono di rimanere aperti verso nuove fonti di evidenza evitando tanto di ridurre l'etica a esse quanto di trascurarle del tutto.

PAROLE CHIAVE: Neuroetica; Metodi; Psicologia morale; Etica empirica; Livelli di spiegazione

E-mail: songhorian.sarah@unisr.it (⊠)



⁽a) Facoltà di Filosofia, Università Vita-Salute San Raffaele, via Olgettina, 58 - 20132 Milano (I)

NEUROETHICS IS AN INTERDISCIPLINARY FIELD in which philosophical insight interacts with data from cognitive neuroscience and other scientific areas. Since its origins in the early 2000s,1 there has been growing academic interest² in focusing on:

ethical issues raised by our increased and constantly improving understanding of the brain and our ability to monitor and influence it, as well as on ethical issues that emerge from our concomitant deepening understanding of the biological bases of agency and ethical decision-making.³

Despite this substantive interest in how neuroscientific data can inform our theories of moral agency and moral decision-making, there has been little attention to the peculiar methodological issues and challenges neuroethics has to face. Assuming that the issues and challenges neuroethics has to deal with are similar – or even identical – to those addressed by other empirical disciplines, the relevant peculiarities of the methods of neuroethics can be obscured by excessively underscoring this resemblance. For this reason, I will focus on the methods of neuroethics as a distinctive discipline.

The methodological challenges neuroethics has to face should not be conflated with those of neuroscience. It is one thing to take into account the methodology of neuroscientific research per se - e.g., learning how neuroimaging techniques work and to what extent experiments may fall below given standards -; it is another thing to assess and evaluate the methods related to, on the one hand, the ethical considerations that can - and should - be raised while doing cognitive neuroscience, and, on the other, the impact the latter can - and should - have on ethics. Certainly, one may claim that the methods of neuroethics are nothing more than an a posteriori combination of the methods of cognitive neuroscience and those of ethics, whatever these may be. And yet, this kind of answer begs the question in two different ways.

On the one hand, there is no universally accepted method of ethics and, thus, the controversy would simply be moved one step forward while remaining substantially unsolved: for one would still have to determine which one of the different proposed methods in ethics is the most suitable for neuroethics. On the other hand, the methods of cognitive neuroscience and those of ethics rest on very different grounds and it is not obvious how they can inform each other: the former is mostly concerned with how the techniques actually work, the latter concerns our knowledge (or lack thereof) of what should count as good or right. Not only do these domains of inquiry not seem necessarily related, but also their combination, however performed, would require a nontrivial passage from description to prescription that should, at least, be argued for.4 Thus, even trying to avoid the question of the nature or existence of a specific method of neuroethics, one would still have to investigate which method, among the possible methods of ethics, is the most suitable for neuroethics, or to ask how cognitive neuroscience and ethics can relate to each other while being so different. In the end, then, it seems that methodological issues akin to those one has tried to avoid by saying that we can simply combine the methods of ethics and those of cognitive neuroscience, come back through the window after being thrown out the door.

With the need for a proper consideration of the methods of neuroethics in mind, my aim in this paper is twofold. First, I will try and see whether there are some specificities of neuroethics, as far as its methodology is concerned, that make it, also from a purely methodological perspective and not only from a substantive viewpoint, a truly novel research field in ethics. To do so, I will consider the definition put forth by Roskies in 2002 of two sub-fields within neuroethics: the ethics of neuroscience and the neuroscience of ethics. I will contend that each of them bears some similarity to more traditional debates: specifically, the ethics of neuroscience is akin to bioethics, and the neuroscience of ethics is akin

to moral psychology. Besides this family resemblance, however, I will claim that the neuroscience of ethics should nonetheless be considered as distinct from moral psychology insofar as the explanations these two disciplines provide for human behavior and human decision-making processes are at different levels. Lacking a proper distinction between these levels of explanation, one could be unduly prone to believe that moral psychology can be reduced to the neuroscience of ethics.

The investigation into the proper relationship between these two latter domains of inquiry may lead one to wonder if there is any prospect for a form of naturalist reductionism in the study of moral cognition. As my second aim, thus, I will deal with the issue of the extent to which empirical findings in general can have consequences for normative ethics or for metaethics – an issue not necessarily specific to neuroethics – with reference to the traditional distinction between facts and values, or "is" and "ought" in ethical reflection.

As a *caveat*, since the aim of this paper is to investigate primarily methodological issues, I will as far as possible avoid any substantial assumptions which may indirectly affect methodological considerations.

Ethics of neuroscience and neuroscience of ethics

In a very influential article, Adina Roskies proposed a distinction between two main branches within neuroethics that would be welcomed by most scholars.⁵ She defined (1) the ethics of neuroscience and (2) the neuroscience of ethics as follows:

(1) the ethical issues and considerations that should be raised in the course of designing and executing neuroscientific studies and (2) the ethical and social impact that the results of those studies might have, or ought to have, on existing social, ethical, and legal structures.⁶

In a nutshell, on the one hand, (1), i.e. the

ethics of neuroscience, can be conceived of as analogous to bioethics. It thus raises mostly the same issues that are raised by bioethics but in relation to the innovative technological possibilities afforded by cognitive neuroscience. Informed consent and incidental findings, for instance, take on specific features when applied to neuroscientific research. Just to provide an example, experiments in cognitive neuroscience mostly use volunteers whose brains are scanned while performing a task in order to see the typical functioning of the human brain: these subjects volunteer for the sake of science and the acquisition of further knowledge about brain functioning. And yet, it may happen that, while scanning a supposedly healthy subject with no behavioral or clinical symptoms, the researchers find some abnormality that may or may not have consequences for the future health of the subject. The obvious ethical issue is whether such information should be provided to the subject or not - just as much as it happens that DNA testing shows a subject's susceptibility to some pathologies that may or may not actually develop in her lifetime.

On the other hand, (2), i.e. the neuroscience of ethics, seems to be the subset of neuroethics that is "truly novel".⁷

The neuroscience of ethics refers to the impact of neuroscientific knowledge upon our understanding of ethics itself.8

Its core interest is in the knowledge that can be acquired by these new technologies as far as our moral abilities are concerned. Just to provide an example of the kind of knowledge that can be acquired by scanning typical subjects' brains, consider the research launched by Joshua Greene and colleagues on the neural bases of moral judgment. They showed which brain areas are involved in judging the appropriateness of a certain course of action by using primarily moral dilemmas as stimuli. It is easy to see that, if this is really to be seen as some novelty that the new field of neuroethics is to bring to ethics, then it actually shows some family resemblance

to traditional moral psychology, since the latter can be described as exploring «human functioning in moral contexts». ¹⁰ It could thus be argued that what is really new about neuroethics is not its *aim* – i.e. investigating how we actually make decisions and pass judgments in moral circumstances (as opposed to dealing with how we ought to do those very same things) –, for this is shared by moral psychology as well; rather, the novelty would be seen to lie in the *methods* it uses and in the *level of the explanation* that it intends to provide.

Not just moral psychology with new tools

Because of their shared aim, moral psychology and the neuroscience of ethics are often conflated in contemporary debate and it is not trivial to draw a line between them. Should, for instance, experiments on implicit attitudes be considered to pertain to moral psychology or to the neuroscience of ethics?

Furthermore, by focusing on specific issues and topics – just to give an example, consider the study of implicit attitudes –, one risks overlooking the existing methodological differences between these two fields and losing some relevant aspects of both. In fact, when considering particular issues, one might end up disregarding the relevance of methodological aspects: if one's aim is that of understanding implicit attitudes, it may be natural to take into account all available evidence regardless of whether their source refers to one method or the other.

Thus, even though both moral psychology and the neuroscience of ethics may aim at describing and explaining human ways of behaving, judging, and deciding, they do so by appealing to different levels of explanation. As David Marr pointed out, one can distinguish between a computational, an algorithmic, and an implementational level of explanation for a certain phenomenon. In order to introduce these levels, Marr 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. The second level – i.e. the algorithmic and representational level – on the contrary, is described by Marr as a way to realize the process, it specifies *how* to do it. Thus,

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. [...] 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 9.¹²

Finally, the last level is the implementational one; in the case of the cash register, it is the specific machine that physically embodies the algorithm. Various implementations of the algorithm are, nonetheless, possible. The cash register, for instance, can be mechanical or electronic. Moreover, human beings themselves are capable of performing the same operation a cash register does – namely, summing up quantities –, even though they are very different as far as their basic mechanisms are concerned.

Once the distinction between the computational, algorithmic, and implementational levels of explanation 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 in-

volves issues that are rather independent of the other two.¹³

The three different levels are of course, according to Marr, «logically and causally related». ¹⁴ And yet:

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. [...] In attempts to relate psychophysical problems to physiology, too often there is confusion about the level at which problems should be addressed.¹⁵

That the three levels are logically and causally related can be exemplified, for instance, by the fact that, when one is designing a device, there might be some implementational constraints guiding the choice for a certain algorithm to be implemented, or it might be the other way around. However, as Marr claims, the three levels are to some extent independent from one another.16 One can advocate a computational theory of vision without entering into details about how it is represented or implemented, just as much as one can defend a theory of addition without dealing with its implementation in any given system nor with its representation in Arabic or Roman numerals.

By distinguishing between different levels of explanation, 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 a low-level and purely implementational description.

Neuroimaging studies can, for instance, reveal how typical humans' brains are activated when we face moral dilemmas. However, if moral computers or cyborgs became possible, the implementational level of explanation would certainly differ, while the representational one might remain identical. The same could be said of inhabitants of other planets who might be wired differently from us. Keeping representational and implementational explana-

tions 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 higher levels of explanation – say, representational and computational.

How does all this apply to our present concern? While moral psychology is interested in what we can call the representational level of how humans judge, decide, and act morally; the neuroscience of ethics deals with the neural mechanisms that undergo and implement such representations. If this way of understanding the difference between moral psychology and the neuroscience of ethics is plausible, then it is easy to see how the latter cannot be fully reduced to the former - nor the other way around. In other words, the neuroscience of ethics cannot be considered as being just moral psychology with new and more efficient tools. I, thus, agree with Schleim who claims that:

had the research been restricted to that descriptive question only, just offering an updated scientific account of how people of different kinds make moral decisions under certain circumstances, moral neuroscience might have just become a modernized form of moral psychology, promising to offer better explanations owing to its more direct access to the central organ of the mind, the human brain.¹⁷

However, what made the neuroscience of ethics particularly interesting and troubling – depending on the viewpoint – is precisely the attempt to bridge the gap between description and prescription. Again, in Scheim's words:

The seductive allure, from the neuroscientific point of view, and the provocation, from the philosophical point of view, instead consisted in the attempt to cross the border between the descriptive and the

normative, for example, by distinguishing morally justified ("rational") intuitions from unjustified ("irrational") ones, based on the brain areas, associated psychological processes, and evolutionary pathways putatively underlying them.¹⁸

My aim so far has been to show to what extent the methods of neuroethics are similar to those which can be found in other more traditional debates in ethics and to what extent they represent a novelty. It seems fair to claim that the ethics of neuroscience has a family resemblance to traditional bioethics and the neuroscience of ethics looks akin to moral psychology. I have focused primarily on the relationship between the neuroscience of ethics and moral psychology underlining both the elements that make them look alike and those that pull them apart from each other. Specifically, I have argued for the fact that the neuroscience of ethics and moral psychology concern - in Marr's terms - two different levels of explanation in accounting for how humans judge, decide, and act morally. Lacking a proper distinction between these levels of explanation, one could be unduly prone to believe that moral psychology can be reduced to the neuroscience of ethics or the other way around.

The investigation into the proper relationship between these two latter domains of inquiry may lead one to wonder if there is any prospect for a form of naturalist reductionism in the study of moral cognition. Up until now I have avoided dealing with the "seductive allure" and the "provocation" – as Schleim calls it – that the neuroscience of ethics represents for moral cognition and moral philosophy. In the next section, I will focus on my second aim in this paper and I will discuss whether empirical findings can have an impact on normative ethical theories or on metaethics, and to what extent (if any) they can do so.

The (not-so-new) challenge of empirical ethics

As Christen and Alfano point out, con-

temporary empirical research on morality is hugely cited in the humanities.¹⁹ In particular, recent empirical findings are used either in normative ethics or in metaethics.20 A similar role was always played by empirical data - e.g. moral psychological ones -, one difference being, though, that it was not clear how exactly they could provide such a contribution. From a philosophical viewpoint, the intricacies and complications of the relationship between empirical findings and normativity has extremely deep roots. The interpretation of Hume's is/ought passage,21 the division between facts and values,²² and Moore's argument against the naturalistic fallacy²³ constitute the main reasons why there seems to be philosophical resistance to the possibility of empirical findings having an impact on moral theorizing. As Sinnott-Armstrong puts it:

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*.²⁴

If we interpreted Hume's passage as the

claim according to which it is impossible to go (even non-deductively) from empirical propositions to normative ones,25 the neuroscience of ethics would have no impact whatsoever on normative accounts. However, this seems an extreme interpretation of Hume's concern - and definitely not the only one available. It is in fact possible that he was mainly concerned with the lack of argumentation in favor of the passage from "is" clauses to "ought" clauses rather than with a complete denial of the very possibility of such passage. And yet, even if a prohibition was actually in place, that would say nothing about the possibility of drawing metaethical implications from empirical premises, since metaethics is itself a descriptive enterprise.²⁶

Kauppinen distinguishes three "philosophical temperaments"²⁷ one can display as regards the question whether empirical research should have an impact and play a role in ethics. His account is particularly interesting for the present purposes as he applies these temperaments precisely to metaethics and to normative ethics. Table 1 summarizes Kauppinen's distinctions.

Let me now briefly summarize Kauppinen's description of the three temperaments to see how they can prove useful for the present work as well. *Armchair Traditionalism* mainly claims that empirical findings can only marginally contribute to our ethical questions. They can, for instance,

supply material for minor premises in ethical arguments—it is perhaps *a priori* true that creatures capable of pleasure and pain deserve moral consideration, but whether fetuses are sensate creatures is an empirical question.²⁸

Empirical data can, thus, tell us something about "why and how"29 people make moral judgments, but that is, according to Armchair Traditionalism, only partially relevant for a moral theory that concerns whether such judgments are true or truth-apt. Such knowledge of how morality works for human minds is relevant for a moral theory insofar as it aims at being psychologically feasible.³⁰ But again, the fact that a certain moral theory can be realized by human subjects says nothing about its being true or not. It is for this and for similar reasons that Armchair Traditionalism claims that «the role of empirical facts is marginal, not essential or fundamental to ethical inquiry».31

On the contrary, those who advocate *Ethical Empiricism* argue in favor of a much more relevant role for empirical findings. In particular, they claim that empirical data can *contribute to solving* ethical debates (*Moderate Ethical Empiricism*), or even that they are *fundamental* and *essential* for ethical inquiry (*Bold Ethical*

	Armchair Traditionalism	Bold Ethical Empiricism	Moderate Ethical Empiricism
Metaethics	Empirical facts are only relevant for causal explanations of particular moral judgments and the capacity to make moral judgments.	Questions about the nature of moral judgment or facts can be answered via empirical study.	Empirical results are an important source of evidence about the nature of moral judgment or facts.
Normative ethics	Empirical facts are only relevant for deriving judgments about particular cases from non-empirical principles and for practical recommendations.	Normative ethical questions are empirical questions.	Empirical results are an important source of evidence about non-derivative moral truths and/or the empirical presuppositions of normative theories.

Table 1. Kauppinen's distinctions - adapted from A. KAUPPINEN, *Ethics and Empirical Psychology*, cit., pp. 280-281.

Empiricism). They represent different forms of naturalism: Moderate Ethical Empiricism can be defined as a revisionist naturalistic account, Bold Ethical Empiricism as a reductionist one.³² Thus, Bold Ethical Empiricism claims that ethical issues can – and probably should – be reduced to empirical ones. In fact, according to a bold version of empirical ethics, normative ethical questions are nothing more than empirical ones and metaethical issues can be solved by empirical studies.

Things get more complicated when one takes into account Moderate Ethical Empiricism. Indeed, a moderate version of empirical ethics does not imply reductionism, but it still attributes a more robust role to empirical findings both in normative ethics and in metaethics compared to the marginal one allowed for by Armchair Traditionalism. Thus, as far as normative ethics is concerned, a moderate version of ethical empiricism would not simply grant the possibility that moral psychological facts are relevant for ascertaining that a certain normative theory is psychologically feasible, but also that they constitute a relevant source of evidence for accounting for «non-derivative moral truths and/or the empirical presuppositions of normative theories».33 Similarly, as far as metaethics is concerned, a Moderate Ethical Empiricism claims that empirical findings are a relevant source of evidence for understanding the very nature of moral judgments and of moral facts, instead of either being only marginally relevant for a causal account of moral judgments and of moral facts - as for Armchair Traditionalism - or being one and the same thing as moral judgments and moral facts, so that knowing the relevant empirical facts would exhaust what one could ever know about such judgments and such facts - as Bold Empirical Ethics would claim.

In this work, I do not intend to argue in favor of any among such "philosophical temperaments", but simply to acknowledge their difference and to show their respective plausibility as far as normative ethics and metaethics are concerned. A bold version of

ethical empiricism (or Bold Empirical Ethics) seems to be the least plausible option, because its reductionism could easily disregard relevant aspects of moral philosophical inquiry, focusing solely on its basic elements. As mentioned in section 3, different levels of explanation of one and the same phenomenon allow some elements to be relevant only in one of them. Similarly, reducing normative theories to facts about human moral psychology is to assume a substantive claim about the nature of moral facts, namely, that they are nothing more than human typical projections. Moral truth would, thus, be identical to typical human response. While such a form of moral anti-realism³⁴ could be argued for, the problem with Bold Empirical Ethics is that it just seems to take this for granted and to presuppose it. These considerations should be sufficient to see how a bold version of empirical ethics can be problematic.

On the contrary, it is far more complicated to provide general reasons and intuitive arguments in favor of (or against) *Armchair Traditionalism* or *Moderate Ethical Empiricism*. The latter seems a viable option in order to take into account as much data as possible without either losing important sources of information or reducing normative issues to empirical ones. And yet, one has to admit that accepting such a thesis presupposes, at least, some form of revisionist methodological naturalism, which is far from being inarguable.³⁵

Since arguing in favor of (or against) Armchair Traditionalism or Moderate Ethical Empiricism seems far more complex, one possibility is to renounce assuming one single philosophical temperament for normative ethics and for metaethics and rather considering each of them separately to see which temperament suits each discipline best. In fact, these philosophical temperaments differ in the extent to which they allow empirical data to impact metaethical debates such as that between moral cognitivism or non-cognitivism, that concerning whether or not moral facts exist, or that between judgment internalism and externalism.

Taking this last debate as an example,

Adina Roskies considered patients with ventromedial prefrontal damage as counterexamples to substantive belief-internalism (or motive-internalism about belief), that is, to the view that the simple fact of holding certain moral beliefs and making certain moral judgments is sufficient to have a motive for acting in accordance with such beliefs and judgments.³⁶ If Roskies' argument is sound, then empirical data on ventromedial prefrontal damage is essential to show that holding a belief or passing a judgment is not sufficient for motivating an agent to act. If that is the case, it seems that Moderate Empirical Ethics is more suited to metaethical disputes, since empirical findings seem to do more than just marginally contribute to resolving them. In the case made by Roskies against beliefinternalism, data from this set of subjects is taken as a disproof, or a counterexample, of the thesis itself: hence, playing a crucial rather than a marginal role in our attempt to uncovmetaethical truths. Moreover, since metaethics is a descriptive enterprise rather than a properly normative one,³⁷ it seems less problematic to claim that empirical findings impact on it more than on normativity, for empirical findings and metaethics both lie in the domain of facts and being, whereas normativity lies in the domain of oughtness, or having to be.

On the contrary, Armchair Traditionalism seems more suitable for accounting for the relationship between empirical data and normative theories. The former can certainly provide information about how and why people reason in a certain way, how and why they pass certain judgments, and how and why they have certain moral beliefs; but they cannot by themselves solve normative debates.³⁸ This does not mean they can play no role, but simply that the priority should be given to philosophical reflection in the contemporary forms of conceptual analysis and argumentative reasoning. 59 This way of conceptualizing the contribution of empirical findings to normative issues is, as it may appear clear now, compatible with Moderate Ethical Empiricism just as much as with Armchair Traditionalism – if the latter is properly understood not as a theoretical framework denying any role to empirical data, but rather as a theory acknowledging they play only a marginal role. The difference between Armchair Traditionalism (so construed) and a Moderate Ethical Empiricism rests uniquely on a very small difference: both claim that empirical data per se cannot solve normative issues, but, while the former considers their contribution only marginal, the latter seems to acknowledge they play a more active role – even though it is extremely complicated to see to what extent it is more so.

Joshua Greene – somehow replying to several objections his and others' studies have raised concerning how they seem to derive normative conclusions from the description of empirical facts⁴⁰ – advocates for some sort of *Moderate Empirical Ethics*:

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.⁴¹

The role Greene attributes to empirical data seems, in fact, more than marginal. In particular, Greene claims that deontological accounts are undermined by the data we have on moral cognition,⁴² even though the latter cannot fully dismiss them all alone.⁴³

As Kauppinen claims, ethical empiricism has four possible ways to try and bridge the gap between "is" and "ought", an attempt *Armchair Traditionalism* would have no interest in doing: via reductionism (*Bold Ethical Empiricism* would go this way), via aetiological debunking, 44 via ethical conservatism, 45 or via psychological unfeasibility. 46 Without entering into details about these strategies, I will very briefly consider the last three. As it happens in Greene, 47 aetiological debunking arguments

aim at showing that deontological or other non-consequentialist beliefs are explained by emotional reactions, and that «their aetiology renders them untrustworthy»⁴⁸ and unjustified. Thus, those who advocate for aetiological debunking arguments would claim that we should reject – or question – deontological or other non-consequentialist normative accounts since the beliefs they are based upon are untrustworthy and unjustified. Somehow similarly, ethical conservatism claims that empirical study can identify ethical commitments that have normative authority even though they result from «a-rational and a-reliable emotional processes».⁴⁹

For maintaining such kind of commitments and the current ethical outlook that derives from them, we would have only pragmatic and non-epistemic reasons, on the one hand, and reasons concerning the undesirable consequences that would derive from giving them up, on the other. That is so because, according to ethical conservatism, they are in truth unwarranted. Thus, empirical research would allow us to differentiate between warranted and unwarranted commitments. And, even though such a distinction would not necessarily imply depriving the latter of any kind of normative authority, they would certainly hold a different normative status: one that has to be granted only because of pragmatic and non-epistemic reasons and of the undesirable consequences that would otherwise be produced.

While reductionism, aetiological debunking, and ethical conservatism seem to imply a quite strong attempt to bridge the gap between description and prescription – one that is certainly not undisputed –, psychological unfeasibility is probably the least controversial way of taking empirical data seriously in normative ethics, and it is not a strategy that can only be ascribed to *Ethical Empiricism*. In fact, even *Armchair Traditionalism* can account for it – whereas it cannot account for the other three strategies. The main thesis of psychological unfeasibility consists in a series of constraints to the normative theories that

can be realized by human moral psychology. A theory that prescribes something completely impossible for human beings is psychologically unfeasible and, if "ought implies can"50 holds, then such a theory should be rejected. For instance, since psychological data have robustly shown that human beings' conduct is sensitive to situational features that are morally irrelevant,⁵¹ normative ethics should take such influences into account. If one prescribes a normative theory that completely disregards situational features, the account would be psychologically unfeasible for humans. This, however, says nothing about the specific way in which the normative theory should actually take such features into account (leaving open the possibility for several theoretical outlooks).

The novelty of the neuroscience of ethics consists, here, in moving this claim further: a normative theory that has an impact on human beings' lives needs to be feasible not only at a psychological and representational level, but also at a neural and implementational one. Data from the neuroscience of ethics show that moral judgment activates brain areas associated both with reasoning and with emotions. If that is the case, a normative theory claiming that moral judgments can and should only be founded upon one or the other source, is a neurally unfeasible theory. But again, this role of empirical data is totally compatible with it being marginal (as Armchair Traditionalism wants it to be): it just limits the theoretical possibilities, without telling us anything about the right or true normative moral theory.

Since it is so complicated to settle how marginal the contribution of empirical evidence, in general, and of the neuroscience of ethics, in particular, should be, there seem to be prudential reasons for being open to either *Armchair Traditionalism* or to *Moderate Empirical Ethics*, as far as normative issues are concerned. What is important is to avoid two opposite risks, that neither of these two perspectives actually faces while still reaching the goal of ending the isolation of conceptual analysis, namely reductionism of normative

accounts to empirical ones, on the one hand, and complete neglect of the relevance of empirical findings, on the other.

Conclusions

My aim has been to track down the methodological specificities of neuroethics. Firstly, I have tried to find out what traditional debates neuroethics resembles and to what extent it actually represents a novelty in ethical thinking. While the ethics of neuroscience has been defined as close to bioethics, the neuroscience of ethics seems akin to moral psychology. However, the latter explains human moral behavior and moral cognition at a representational level rather than at the implementational one to which the neuroscience of ethics aims.

Marr's distinction and the elaboration he advances clarifies to what extent moral psychology and the neuroscience of ethics can be related to each other, as well as the extent to which they differ. While the former illuminates the representations and the conscious mechanisms involved in moral reasoning and in moral agency; the latter explains the neural and even the hormonal functioning of the human brain. As Marr points out, these two levels are related though independent: as I have recalled above, it is at least theoretically possible to imagine other species or other entities that would be able to realize the same representations humans' display when reasoning and acting morally; and yet this does not necessarily entail that they would need to have the same exact brain. Furthermore, as we learn about neuroplasticity even after brain injuries in adults,⁵² the multiple realizability thesis seems to increase in likelihood.⁵³ In fact, if our brains were found to be plastic even as far as the so-called moral brain is concerned, then the disjunction between the representational and the implementational level, between moral psychology and the neuroscience of ethics, would not refer only to hypothetical and unreal cases, but also to real ones.

Hence, by properly distinguishing between

different explanatory levels, one avoids the risk of believing that moral psychology can be reduced to the neuroscience of ethics or the other way around.

Having shown that opting for a reduction would mean facing the risk of disregarding important explanatory elements, I have, thus, moved to the consideration of whether there is any prospect for a form of naturalist reductionism in the study of moral cognition. For this reason. I have considered the divide between facts and values and the extent to which empirical findings might be conceived of as bearing consequences for our normative or metaethical accounts, thus bridging such a gap. On this, the debate is extremely rich and I could just gesture at a few relevant issues. Certainly, the huge amount of research in neuroethics, moral cognition, developmental psychology, social psychology, and the like that have been conducted in the last two decades,54 have prompted a reaction by philosophers challenging them to refine their conception of the relationship between empirical data and philosophical insight. The debate, thus, is not new, but it has been revitalized and stimulated by recent developments in other disciplines.

Since I aimed at providing a portrait of the debate concerning the methods of neuroethics, I have tried to avoid going into substantial debates. Clearly, once a general framework has been chosen or agreed on, then more detailed and specific debates, concerning both the methods and the substance, can be raised. For instance, it is easy to claim that the methods of an fMRI study on moral judgment, as well as its implications for normative theories, can differ from the methods and the implications of, say, a developmental longitudinal and behavioral study on altruistic behavior towards strangers in need. Thus, one needs to bear in mind that things can get much more specific as far as both the implications for normative theories and the methods are concerned. And yet, the general framework and the general issues that I have considered here seem to hold throughout these more specific debates.⁵⁵

Thus, even considering more specific topics and issues, one would always have to face some version of the three philosophical temperaments considered above. Prudential reasons would, then, recommend openness to the possibility that data coming from the neuroscience of ethics would play some role in either metaethics or normative ethics. Avoiding the two major risks, namely reductionism and neglect, there is plenty of room for debate concerning the extent of the contribution such data might offer. What seems implausible is to think that empirical evidence would have an ultimate role or no role at all. However, which exact position along the continuum between these extremes they might occupy is open to discussion and likely to depend on the specific issue at hand. Moderate Ethical Empiricism and Armchair Traditionalism (understood as advocating a marginal role and not as denying any role at all to empirical findings) are both plausible ways of accounting for the relationship between empirical data and normative ethics.

One thing this paper aimed at adding to the debate on this relationship and, in particular, on the thesis that a normative theory ought to be psychologically feasible is a consequence of an appropriate distinction between the psychological and the neural levels (adapted from Marr's one): a normative theory should not only be psychologically feasible, but it should also be neurally possible for human brains. Hence, moral psychology would illuminate the constraints we have at a representational level; while the neuroscience of ethics would enlighten those that refer to the specific implementation humans have. As for the relation between such data and metaethics, I have argued that a moderate empirically informed metaethics presently appears the most promising option.

Notes

¹ Cf. W. SAFIRE, *The But-What-If Factor*, in: «The New York Times», May 16, 2002; A. ROSKIES, *Neuroethics for the New Millenium*, in: «Neu-

ron», vol. XXXV, n. 1, 2002, pp. 21-23; S.J. MARCUS (ed.), Neuroethics: Mapping the Field, Dana Press, New York 2002; J. ILLES, Neuroethics in a New Era of Neuroimaging, in: «American Journal of Neuroradiology», vol. XXIV, n. 9, 2003, pp. 1739-1741; J. ILLES, Neuroethics: Defining the Issues in Theory, Practice, and Policy, Oxford University Press, Oxford / New York 2006.

- ² For a perspicuous survey, see J. CLAUSEN, N. LEVY (eds.), *Handbook of Neuroethics*, Springer, Dordrecht 2015, pp. 9-29.
- ³A. ROSKIES, *Neuroethics*, in: E.N. ZALTA (ed.), *The Stanford Encyclopedia of Philosophy*, Spring Edition 2016, URL: https://plato.stanford.edu/archives/spr2016/entries/neuroethics/ [retrived: November 16th, 2018].
- ⁴ Cf. D. Hume, A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects (1739), critical edition by D.F. NORTON, M.J. NORTON, Clarendon Press, New York 2007, p. 302.
- ⁵ Cf., e.g. N. LEVY, Neuroethics. Challenges for the 21st Century, Cambridge University Press, Cambridge/New York 2007; M. REICHLIN, Neuroethics, in: D. FASSIN (ed.), A Companion to Moral Anthropology, Wiley-Blackwell, Malden 2012, pp. 595-610.
- ⁶ A. ROSKIES, Neuroethics for the New Millenium, cit., p. 21.
- ⁷ Ibidem.
- ⁸ N. LEVY, Neuroethics. Challenges for the 21st Century, cit., p. 1.
- ⁹ Cf. J.D. GREENE, Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)Science Matters for Ethics, in: «Ethics», vol. CXXIV, n. 4, 2014, pp. 695-726; J.D. GREENE, Why Are VMPFC Patients More Utilitarian? A Dual-Process Theory of Moral Judgment Explains, in: «Trends in Cognitive Sciences», vol. XI, n. 8, 2007, pp. 322-323; J.D. GREENE, Cognitive Neuroscience and the Structure of the Moral Mind, in: P. CARRUTHERS, S. LAU-RENCE, S.P. STICH (eds.), The Innate Mind: Structure and Contents, Oxford University Press, Oxford/New York 2005, pp. 338-352; J.D. GREENE, F.A. CUSHMAN, L.E. STEWART, K. LOWENBERG, L.E. NYSTROM, J.D. COHEN, Pushing Moral Buttons: The Interaction Between Personal Force and Intention in Moral Judgment, in: «Cognition», vol. CXI, n. 3, 2009, pp. 364-371; J.D. GREENE, L.E. NYSTROM, A.D. ENGELL, J.M. DARLEY, J.D. COHEN, The Neural Bases of Cognitive Conflict and Control in Moral Judgment, in: «Neuron»,

vol. XLIV, n. 2, 2004, pp. 389-400; J.D. GREENE, R.B. SOMMERVILLE, L.E. NYSTROM, J.M. DARLEY, J.D. COHEN, *An fMRI Investigation of Emotional Engagement in Moral Judgment*, in: «Science», vol. CCXCIII, n. 5537, 2001, pp. 2105-2108.

Off. J. Doris, S. Stich, J. Phillips, L. Walmsley, Moral Psychology: Empirical Approaches, in: E.N. Zalta (ed.), The Stanford Encyclopedia of Philosophy, Winter Edition 2017, URL: https://plato.stanford.edu/archives/win2017/entries/moral-psychemp/ [retrived: November 16th, 2018]; J. Doris, Moral Psychology Research Group (eds.), The Moral Psychology Handbook, Oxford University Press, Oxford/New York 2010.

¹¹ Cf. D. MARR, Vision. A Computational Investigation into the Human Representation and Processing of Visual Information, Freeman, New York 1982, pp. 24-25; D.M. KAPLAN, Explanation and Levels in Cognitive Neuroscience, in: J. CLAUSEN, N. LEVY (eds.), Handbook of Neuroethics, cit., pp. 9-29.

¹² D. MARR, *Vision*, cit., p. 23.

¹⁶ On how things like minds can relate to things like brains, cf. K. AIZAWA, C. GILLETT, *Realization, Reduction, and Emergence: How Things Like Minds Relate to Things Like Brains*, in: J. CLAUSEN, N. LEVY (eds.), *Handbook of Neuroethics*, cit., pp. 49-62.

¹⁷ S. SCHLEIM, *Moral Cognition: Introduction*, cit., p. 98-99.

¹⁸ *Ivi*, p. 99.

19 Cf. M. Christen, M. Alfano, Outlining the Field – A Research Program for Empirically Informed Ethics, in: M. Christen, C. van Schaik, J. Fischer, M. Huppenbauer, C. Tanner (eds.), Empirically Informed Ethics: Morality between Facts and Norms, Springer, Heidelberg/New York/Dordrecht/London 2015, pp. 3-27, here pp. 3-4. See also S. Matusall, M. Christen, I. Kaufmann, The Emergence of Social Neuroscience as an Academic Discipline, in: J. Decety, J. Cacioppo (eds.), The Oxford Handbook of Social Neuroscienc, Oxford University Press, Oxford 2011, pp. 9-27.

²⁰ Cf. A. KAUPPINEN, Ethics and Empirical Psychology – Critical Remarks to Empirically Informed Ethics, in: M. CHRISTEN, C. VAN SCHAIK, J. FISCHER, M. HUPPENBAUER, C. TANNER (eds.), Empirically Informed Ethics, cit., pp. 279-305, here p. 280. See also R. JOYCE, What Neuroscience Can (and Cannot) Contribute to Metaethics, in: W. SINNOTT-

ARMSTRONG (ed.), Moral Psychology, vol. III: The Neuroscience of Morality: Emotion, Brain Disorders, and Development, MIT Press, Cambridge (MA) 2008, pp. 371-394, here 271.

²¹ Cf. D. HUME, A Treatise of Human Nature, cit., p. 302.

²² Cf. H. PUTNAM, The Collapse of the Fact/Value Dichotomy and Other Essays, Harvard University Press, Cambridge (MA) 2002.

²³ Cf. G.E. MOORE, *Principia Ethica* (1903), edited by T. BALDWIN, Cambridge University Press, Cambridge 1993; P. CLIPSHAM, *Does Empirical Moral Psychology Rest on a Mistake?*, in: «Philosophical Studies», vol. CLXX, n. 2, 2013, pp. 215-233.

²⁴ W. SINNOTT-ARMSTRONG, *Moral Intuitionism Meets Empirical Psychology*, in: T. HORGAN, M. TIMMONS (eds.), *Metaethics After Moore*, Clarendon Press, Oxford 2006, p. 339-365, here p. 339 italics added.

²⁵ Cf. S.M. Brown Jr., *Does Ought Imply Can?*, in: «Ethics», vol. LX, n. 4, 1950, pp. 275-284; B. CELANO, *Dialettica della giustificazione pratica. Saggio sulla legge di Hume*, Giappichelli, Torino 1994.

²⁶ Cf. R. JOYCE, What Neuroscience Can (and Cannot) Contribute to Metaethics, cit., p. 371.

²⁷ A. KAUPPINEN, *Ethics and Empirical Psychology*, cit., p. 279: «it is important not to caricature these positions. Moral philosophers have never claimed that empirical facts play no role in ethics».

²⁸ A. KAUPPINEN, *Ethics and Empirical Psychology*, cit., pp. 280.

²⁹ A. KAUPPINEN, *Ethics and Empirical Psychology*, cit., pp. 280.

³⁰ Cf. also O. FLANAGAN, *Varieties of Moral Personality*, Harvard University Press, Cambridge (MA) 1991, p. 32; K.A. APPIAH, *Experiments in Ethics*, Harvard University Press, Cambridge (MA) 2008, pp. 22-23 and 46.

³¹ A. KAUPPINEN, *Ethics and Empirical Psychology*, cit., pp. 280.

³² Haack distinguishes between five different conceptualizations of naturalism in epistemology. I take such disambiguation – or a similar one – to be needed also in the field of moral epistemology. See S. HAACK, *Naturalism Disambiguated*, in: S. HAACK, *Evidence and Inquiry. Towards Reconstruction in Epistemology*, Blackwell, Oxford/Cambridge(MA) 1993, pp. 118-138, here 118-120.

³³ A. KAUPPINEN, *Ethics and Empirical Psychology*, cit., pp. 281.

³⁴ Cf. R. JOYCE, Moral Anti-Realism, in: E.N. ZALTA (ed.), The Stanford Encyclopedia of Philosophy, Win-

¹³ Ivi, p. 25.

¹⁴ Ibidem.

¹⁵ Ibidem.

ter Edition 2016, URL: https://plato.stanford.edu/archives/win2016/entries/moral-anti-realism/[retrived: November 16th, 2018].

35 Cf. D. Papineau Naturalism, in: E.N. Zalta (ed.), The Stanford Encyclopedia of Philosophy, Winter Edition 2016, URL: https://plato.stanford.edu/archives/win2016/entries/naturalism/ [retrived: November 16th, 2018]; J.J. Prinz, Naturalizing Metaethics, in: T. Metzinger, J. M. Windt (eds), Open MIND, vol. XXX, MIND Group, Frankfurt a.M. 2015 - doi: 10.15502/9783958570511; J. Lenman, Moral Naturalism, in: E.N. Zalta (ed.), The Stanford Encyclopedia of Philosophy, Spring Edition 2014, URL: http://plato.stanford.edu/archives/spr2014/entries/naturalism-moral/ [retrived: November 16th, 2018]; M. De Caro, D. Macarthur, Naturalism and Normativity, Columbia University Press, New York 2010.

³⁶ Cf. A. ROSKIES, Are Ethical Judgments Intrinsically Motivational? Lessons from "Acquired Sociopathy", in: «Philosophical Psychology», vol. XVI, n. 1, 2003, pp. pp. 51-66; A. ROSKIES, Patients with Ventromedial Frontal Damage Have Moral Beliefs, in: «Philosophical Psychology», vol. XIX, n. 5, 2006, pp. 617-627.

³⁷ Cf. R. JOYCE, What Neuroscience Can (and Cannot) Contribute to Metaethics, cit., p. 371.

³⁸ Cf., e.g., S. BERKER, *The Normative Insignificance of Neuroscience*, in: «Philosophy and Public Affairs», vol. XXXVII, n. 4, 2009, pp. 293-329.

³⁹ For an interesting reconstruction of how our contemporary attempt to bridge the gap between psychology and philosophy is a return to their original indistinctiveness rather than an actually novel enterprise, see K.A. APPIAH, *Experiments in Ethics*, cit., pp. 5-32.

⁴⁰ Cf., e.g. J.D. GREENE, Why Are VMPFC Patients More Utilitarian?, cit.; J.D. GREENE, Cognitive Neuroscience and the Structure of the Moral Mind, cit.; J.D. GREENE, F.A. CUSHMAN, L.E. STEWART, K. LOWENBERG, L.E. NYSTROM, J.D. COHEN, Pushing Moral Buttons, cit.; J.D. GREENE, L.E. NYSTROM, A.D. ENGELL, J.M. DARLEY, J.D. COHEN, The Neural Bases of Cognitive Conflict and Control in Moral Judgment, cit.; J.D. GREENE, R.B. SOMMERVILLE, L.E. NYSTROM, J.M. DARLEY, J.D. COHEN, An fMRI Investigation of Emotional Engagement in Moral Judgment, cit.; J.D. GREENE, J. HAIDT, How (and Where) Does Moral Judgment Work?, in: «Trends in Cognitive Sciences», vol. VI, n. 12, 2002, pp. 517-523.

⁴¹ J.D. GREENE, Beyond Point-and-Shoot Morality, cit., p. 711.

⁴² J.D. GREENE, The Secret Joke of Kant's Soul, in: W. SINNOTT-ARMSTRONG (ed.), Moral Psychology, vol. III, cit., pp. 35-79; J.D. GREENE, Beyond Point-and-Shoot Morality, cit.

⁴³ For a critique of this aspect of Greene's account, see R. DEAN, *Does Neuroscience Undermine Deontological Theory?*, in: «Neuroethics», vol. III, n. 1, 2010, pp. 43-60.

⁴⁴ Cf. J.D. Greene, The Secret Joke of Kant's Soul, cit. ⁴⁵ Cf. S. NICHOLS, M. TIMMONS, T. LOPEZ, Using Experiments in Ethics – Ethical Conservatism and the Psychology of Moral Luck, in: M. CHRISTEN, C. VAN SCHAIK, J. FISCHER, M. HUPPENBAUER, C. TANNER (eds.), Empirically Informed Ethics, cit., pp. 159-176.

⁴⁶ A. KAUPPINEN, Ethics and Empirical Psychology, cit., pp. 290-304.

⁴⁷ Cf. J.D. Greene, *The Secret Joke of Kant's Soul*, cit. ⁴⁸ A. Kauppinen, *Ethics and Empirical Psychology*, cit., pp. 292.

⁴⁹ S. NICHOLS, M. TIMMONS, T. LOPEZ, *Using Experiments in Ethics*, cit., p. 160.

50 Cf. I. Kant, Critique of Pure Reason (1781/1787), edited by N. Kemp Smith, Palgrave MacMillan, Basingstoke 1929, A548/B576; I. Kant, Religion within the Boundaries of Mere Reason (1793), edited by A.W. WOOD, G. DI GIOVANNI, Cambridge University Press, Cambridge 1998, VI, p. 50. See also J. Kekes, "Ought Implies Can" and Two Kinds of Morality, in: «The Philosophical Quarterly», vol. XXXIV, n. 137, 1984, pp. 459-467; R. Stern, Does "Ought" Imply "Can"? And Did Kant Think It Does?, in: «Utilitas», vol. XVI, n. 1, 2004, pp. 42-61; K.A. Appiah, Experiments in Ethics, cit.; F. Howard-Snyder, Ought Implies Can, in: The International Encyclopedia of Ethics, Wiley-Blackwell 2013 - doi: 10.1002/9781444367072.wbiee402.

⁵¹ K.A. APPIAH, Experiments in Ethics, cit., pp. 33-72. Cf. Y.R. SOPHIE SU, A. VEERAVAGU, G. GRANT, Neuroplasticity after Traumatic Brain Injury, in: D. LASKOWITZ, G. GRANT (eds.), Translational Research in Traumatic Brain Injury, CRC Press, Boca Raton/London/New York 2016, pp. 163-178.

⁵³ For a review of the debate on multiple realizability, see J. BICKLE, *Multiple Realizability*, in: E. N. ZALTA (ed.), *The Stanford Encyclopedia of Philosophy*, Winter Edition 2016, URL: https://plato.stanford.edu/entries/multiple-realizability/[November 23rd, 2018].

⁵⁴ Cf. M. CHRISTEN, M. ALFANO, *Outlining the Field*, cit., pp. 3-4.

55 For a more detailed analysis of these debates,

see J. CLAUSEN, N. LEVY (eds.), *Handbook of Neuroethics*, Springer, Dordrecht 2015; M. CHRISTEN, C. VAN SCHAIK, J. FISCHER, M. HUPPENBAUER, C.

TANNER (eds.), *Empirically Informed Ethics*, cit.; M. CHRISTEN, M. ALFANO, *Outlining the Field*, cit., pp. 16-25.