Journal of Philosophy and Ethics in Health Care and Medicine, No.6, pp.62-78, August 2012

Scientific Contribution

On the Possibility of Explanatory Pluralism in Neuroethics

Yasushi ISHIDA
(Yokohama City University,
E-mail : yaishida@gmail.com)

Abstract :

It has been considered that when reduction holds, it precludes any explanatory significance of the properties of a reduced theory — when reduction is construed in a strict sense. We can see that this interpretation is not necessarily true, when we examine psycho-neural relation in terms of John Bickle's "revisionist reductionism." Bickle's view interprets reduction with a spectrum model in which reduction is considered to be a matter of degree; it is a continuum from smooth reduction (retaining a reduced theory), through revisionary reduction (revising a reduced theory) to complete displacement (eliminating a reduced theory). This view helps us see the psycho-neural relation in bi-directional explanatory dependence. Given the epistemological and methodological nature of Bickle's argument, we can apply this view to ethico-neural relation, and thereby obtain a pluralistic, multi-layered explanatory model of ethico-neural relation in the context of neuroethics, emphasizing the explanatory significance of ethical properties; that is, even when reduction is construed in the strict sense, ethical properties — properties of a reduced theory — play a significant explanatory role.

Keywords : neuroethics, ethics, neuroscience, reduction, explanatory role, psychology, ethico-neural dualism, Roskies

1. The problem and related issues

Significant theoretical progress is being made in the field of neuroscience. Given recent developments, it is quite natural for

neuroscientists to think that their science can explain all human behavior and psychological happenings – feelings, thinking and judgment included. They seem to think that they can explain, or will be able to explain in the future, even our moral life phenomena (moral feelings, moral judgments, moral reasoning, ethical behavior, moral responsibility, and so on) with their theories.[1] They would claim that neuroscientific theories are so explanatorily "powerful" that our moral life can be explained away ("completely explained" some would say) or understood in neuroscientific terms. For them, neuroscience is *the* explanatory "base" of our ethical or moral life; proper neuroscientific knowledge of the brain system will explain how and why one makes ethical or moral judgments, for example, your judgment that it is wrong to kill people, or my feeling free to smoke in public.

This conviction is backed by the notion of *reduction*.[2] Many neuroscientists maintain that when neuroscience has made enough progress, special sciences such as psychology will be *reduced* to neuroscience; that is, they will be explained completely in neuroscientific terms, even though it is still not completely possible. Some even say that the same idea applies to the field of morality and that neuroscience will eventually explain away what is going on in ethics. This claim of ethico-neural reduction will certainly face the objections of those who accept G. E. Moore's distinction between ethical and natural properties.[3] Moore's idea is that ethical properties (e.g., goodness) supervene on natural properties. Two objects differing in their natural properties could differ in their goodness. As Moore was a *dualist* about descriptive and normative properties, goodness is a nonnatural property despite (*or* because of) the supervenience relationship. In Moore's view, the supervenience relation was not ethico-neural reduction. Like Moore, ethical/natural property dualists would oppose the type of reduction. Two main influential reasons for the opposition are well-known *multiple-reliability* and *mental anomalousness* — the mainstream antireductionist arguments proposed in philosophy of mind. These arguments are so telling, I believe, that

it is quite a contentious issue how they fare with the neuroscientific claims for ethico-neural reduction.

In this paper, I will put forth a dualist view of ethico-neural relations that ethical properties are distinct from neural properties — dualist in a different sense from those claimed in terms of multiple-reliability and mental anomalousness. It is my contention that ethico-neural reduction holds only *to some extent*. I thus claim for a certain degree of independence of ethical properties while accepting neuroscience's reductive explanatory power. I will maintain that the properties of a "reduced" ethical theory can still play a significant role in the context of neuroethics, even when we construe reduction in the strict sense – in the sense that ethical properties can be explained, thus replaced, by neural properties. Also, in this paper, I take "neuroethics" in the sense of "neuroscience of ethics" – one of the two interpretations proposed by A. Roskies (Roskies 2002). It is a neuroscientific investigation of ethical issues, such as our (making) ethical judgments and applying ethical concepts or expressions – investigations including attempts to explain our moral phenomena in neuroscientific terms: for example, an attempt to explain one's feeling guilty after stealing through neuroscientific researches and considerations. It seeks neuroscientific foundation for our moral phenomena.

As will be specified later, in this paper, a theory is considered to have its own linguistic, terminological sphere; all the theories discussed here — more specifically neuroscientific theory, psychological theory and ethical theory — are thus considered to present themselves as terminological units. Also I take the position that properties of a theory are expressed (or realized) by concepts or terms used in the theory. I will thus discuss the relation between the sets of vocabularies of theories (accordingly, intertheoretic reduction in discussing reduction) and give consideration to whether the vocabularies are accurate enough to describe phenomena (i.e., so-called "grainedness" of terms).

To present my view, I will first put forward the following three cardinal claims:

(X) The psycho-neural relation must be stronger than one-way dependence.

(Y) Psychological properties do not exercise downward causation on neural properties, but instead exercise a different type of influence on neural properties: explanatory influence.

(Z) Psychological properties and neural properties both have theoretical significance — in particular for the development of neuroscience.

(X) is necessary to evade full-blooded reductionism — reductionism that assumes one-way dependence with a reduced theory depending on a reducing theory; as a result of one-way dependence, the explanatory role of a reduced theory is completely assimilated into that of a reducing theory. I will avoid this version of reductionism by emphasizing a reduced theory's explanatory significance. I do not intend to set forth a full-blooded dualism, a dualism to the effect that ethical properties are causally efficacious, however. I thus need (Y). I do not commit to any argument about downward causation. My argument in this paper is neutral on causal efficacy of the mental. The dualist crux that I want to claim is captured by (Z), which is the main issue of this paper. As I will show below, the view based on these three claims was originally set forth in the context of the psycho-neural relation in a "revisionist view" of reduction presented by J. Bickle. But given the nature of the view, I will suggest that the argument applies to ethico-neural relation as well (read "psychological" as "ethical"); the theoretical significance of the properties of a reduced theory will be shown to obtain for the ethico-neural relation.

2. Ethico-neural relation and strict reduction

Recent developments in neuroscience have lead the neuroscientists to claim that it can explain our psychological phenomena, such as feeling thirsty and having desires, with their neuroscientific vocabulary. There are certain areas (like "thinking"

and "loving"), they admit, in which neuroscience can*not* give thorough and precise explanations yet, but as time goes by, they claim, their science will certainly reach that point. Granting this optimism, neuroscience is, or will be sooner or later, so well-theorized that it is, or will be, equipped with a complete theoretical language that is fine-grained and well-structured enough to explain any psychological happening. It is just matter of time, neuroscientists would say.

Some neuroscientists are convinced that they can go further. Given the assumed explanatory power of neuroscience, neuroscientific theories can explain away our morality; the idea is that the theories are so "powerful" that our moral life is explainable *solely* in neuroscientific terms without having to appeal to any other type of theories, such as intentional psychology or folk psychology, that is, without using any psychological vocabularies. To use the term "reduction," they may well say that ethical vocabularies are reduced to neural vocabularies. Morality is sufficiently explainable in the vocabulary of neuroscience. We do not need intentional psychology or folk psychology to do the job; we just need the exhaustive comprehension of the neuroscientific mechanism, nothing else.

This view will appear to be a threat to the proponents of the metaphysical independence of morality, for example, the imcompatibilists who contend that our moral life (for example, freedom – in the sense that one feels free in taking actions or one grasps or conceptualizes that feeling) is incompatible with determinism. Let me illustrate this by looking at an argument presented by A. Roskies, a neuroethicist.

In her paper, "Neuroscientific challenges to free will and responsibility," Roskies sets forth a compatibilist view on determinism vs. freedom (Roskies 2006) – a metaphysical position which asserts that determinism and freedom are compatible. She basically accepts the neuroscientific, deterministic framework, maintaining that the mechanistic or deterministic view that neuroscience presents has little or no bearing on the question of

the public's apprehension of the problem of whether we are free or morally responsible.[4] Strictly speaking, the position based on determinism is not considered to cohere with any metaphysical position that holds that one is free. Roughly, if one is always determined by some outside factors, one is not expected to hold responsibility for actions. Responsibility is an essence for freedom. Thus normally compatibilism will not hold. So in her paper, Roskies attempts to establish a compatibilist view by construing freedom and responsibility as our *perception* or *intuition* of freedom and responsibility. She cites experimental evidence in which subjects, who hold a deterministic world view, tend to express libertarian (i.e., indeterministic) intuition when given a scenario specifically depicting a rather concrete and emotionally-affecting episode of wrongdoings (for example, raping and killing a girl in an extremely cruel fashion). By so doing, Roskies puts much emphasis on the fact that we cannot ignore the influence of our emotion, our subjective state, when we consider our freedom and moral responsibility. She thereby points out:

> The actual psychological processes involved in everyday moral judgments of responsibility are likely to operate largely independently of theoretical views about determinism and mechanism (Roskies 2006, p.422).
>
> Preliminary results suggest that even if neuroscientific advances were to affect our theoretical views about human freedom, they are not likely to affect practical judgments of moral responsibility (Roskies, *Ibid.*).

Roskies' contention is that compatibilism should hold when a person *perceives* or *feels* that she can make judgments of freedom and responsibility for herself, even if the world is deterministic. In presenting her view, Roskies contrasts determinism not with freedom and responsibility — quite a philosophically tough contrast

— but instead with our *perception* or *awareness* of freedom and responsibility. This argument, which grounds the realization of freedom on our subjective evidence, does not deal with the traditional difficulty that the problem of freedom has been facing, that is, the metaphysical conflict between determinism and freedom. If we accept that one can only talk of one's feeling or thinking that one is free in discussing the conflict, a person who has been under some mind control can illegitimately consider herself to be free. Obviously, this is odd. To put it in terms of the philosophy of mind, the dualistic opposition of "the first-person report" and "the third-person observation" — the opposition that has been considered to be a real challenge or has often been highlighted with the term "explanatory gap" — remains unresolved in Roskies' argument; she tries to present a solution for the problem simply by omitting one of the binary opposition.

Interestingly, while Roskies discusses her concept of freedom in a subjective manner, she cannot ignore the deterministic power of neuroscience. In the last part of her argument, Roskies acknowledges that when deterministic neuroscience has made enough progress, the neurological descriptions that are taken to be "causes of behavior" (i.e., neuroscientific base) will bypass the description of our mental states or human agency, allowing no causal role for them (Roskies, *Ibid.*). Certainly, this reductionist understanding presents a threat to freedom and responsibility. Roskies clearly admits that reductionism on which neuroscience is based "bypasses" our mental states and "precludes causal mentalistic descriptions," making mentalistic descriptions inefficacious, i.e., mentalistic vocabulary useless in explaining our behavior. But we will find that Roskies' fear of reductionism will turn out to be groundless when we investigate the notion of reduction more closely. I will argue that this investigation even supports the idea that ethical vocabularies can retain their explanatory significance when reduction takes place.

Roughly, reduction or scientific reduction is primarily understood to be a way to "integrate phenomena (facts, entities)

into a theory" or "explain phenomena (facts, entities) in terms of a theory," or "assimilate a theory into another theory." Here, it is assumed that a theory has its own space of theoretical language, and in case there are two theories under discussion the translatability between them is also at issue. Scientific reduction can thus be considered to be *intertheoretic* reduction. The reductive relation between psychology and neuroscience implies that each theory has its own theoretical or conceptual space, psychological space and neuroscientific space — the latter "assimilating in" or "explaining" the former. What does this mean?

The most influential "classic work" on *intertheoretic* reduction in the philosophy of science was presented in Chapter 11 of Ernest Nagel's *The Structure of Science* (Nagel 1961). For Nagel, reduction means intertheoretic reduction; it is logical deduction (derivation) of the statements of a reduced theory, T_R, from those of a reducing theory, T_B. In many cases, the T_R contains terms that do not occur within the descriptive vocabulary of T_B (e.g., when it is said that equilibrium thermodynamics is *reduced* to statistical mechanics and the kinetic/corpuscular theory of matter, the former contains, for instance, "pressure" and "temperature," which do not occur in the latter). To derive such T_R from T_B, both of which may not share the same terms, Nagel insists that the premises of the derivation require corresponding principles called "bridge principles" that connect terms across the theories (e.g., in the above example, "temperature in a gas is mean kinetic energy of molecular constituents"). We can hence represent Nagel's account as follows:

(R) T_B and BP logically entail T_R,

where BP refers to "bridge principles." (R) means that when reduction holds between T_B and T_R, T_R is deduced from T_B through BP. T_R is not logically independent from T_B and BP. But does T_R become explanatorily redundant, when reduction holds?

This classic account of reduction implies quite a crucial point on

the notion. Even though T_R may include terms that T_B does not have, bridge principles connect — translate in another word — the two different sets of vocabularies, namely T_R and T_B. (Remember the terms "pressure" and "temperature" in the above case T_R.) This means that the fact that a reducing theory has certain vocabularies that a reduced theory does not have, or vice versa, does not imply reduction is impossible; reduction is nevertheless possible thanks to the bridge principles which connect the two sets of vocabularies. The difference between the sets of vocabularies of T_R and T_B does not block reduction. Conversely, it *is* possible for the vocabulary of T_R to be explanatorily useful even when reduction holds.

Nevertheless, it is still true that T_R and T_B are different in precision in explaining a phenomenon; normally, T_B does a better job in analyzing it because T_B is equipped with more detailed and accurate vocabularies. To borrow a much-used term from the philosophy of science, T_R and T_B have different terminological "grainedness"; the vocabulary of T_B is more fine-grained. Here, we can understand intertheoretic reduction in terms of the grainedness of a theory's vocabulary. When intertheoretic reduction holds, there is *asymmetry of terminological grainedness* between a reduced theory and a reducing theory. As a reducing theory normally does a better job in explaining phenomena, the vocabulary of the theory is more fine-grained than that of a reduced theory. This notion of asymmetry of terminological grainedness between the two theories will play a crucial role below in applying J. Bickle's argument. As we will see later, Bickle contends that psychology is reduced to neuroscience only to a certain degree and that psychological vocabularies can exercise a certain explanatory influence on neuroscience — based on the asymmetry of terminological grainedness between a reduced theory and a reducing theory. If Bickle's argument holds solely on that basis, we can apply the argument to other relations where the same asymmetry of terminological grainedness holds between a reduced theory and a reducing theory.

It has been pointed out that this classic reductionism involves

some difficulties. When we look back at the history of science, we witness that classic reductionism failed to capture what happened in actual cases of reduction. A most typical case is one in which the reduced theory (T_R) is false. Remember the phlogiston theory of combustion. The theory was evidently false; "phlogiston" was non-existent and "dephlogistification" was theoretically impossible. The theory was later "reduced" to oxygen chemistry. Strictly speaking, however, it is impossible to derive, in the Nagelian sense, a false T_R from a true T_B, because there are no bridge principles in such cases — there can be no correct bridge principles between a T_B and a false theory. The (R) above thus does not hold in this case. Another example is Galilean physics vs. Newtonian physics. One may say that the former was reduced to the latter in the course of history. But Galilean physics did not describe how objects behave near the surface on earth as precisely as Newtonian physics does. As the descriptions in Galilean physics are just an approximation of those in Newtonian physics, no correct bridge principles hold between the two. Then, the (R) does not hold in this case, either. These cases pose a serious problem on Nagel's account of reduction.

What can we do, then? To solve this problem, some attempts to retain Nagel's spirit have been proposed. These proposed solutions, I take, have some plausibility, but will lead to the idea that a reduced theory is not completely assimilated or integrated into a reducing theory even after reduction takes place, both theories being still needed in some sense — seemingly a conclusion contradictory to the spirit of full-blooded reductionism, i.e., reductionism that assumes the complete explanatory assimilation of a reduced theory into a reducing theory. In the following section, I will see how this understanding obtains.

3. Bickle's spectrum account of reduction

Let us take a look at some of the proposed solutions. Schaffner modifies the above (R) and presents the following (R*) (Schaffner 1967). Given the "irregular" cases (in the Nagelian sense) in the

history of science, what must be deduced from a base reducing theory (T_B) is not the original T_R but a logically *corrected* version T_R*, thus:

(R*) T_B and bridge principles logically entail T_R*.

It is not T_R but T_R* that is deduced. There are some noteworthy points on T_R*. First, Schaffner specifies, the terms and vocabularies of T_R*, compared to those of T_R that lack exact reference, have to provide more accurate empirical explanations and predictions than T_R. In the case of reduction of Galilean physics to Newtonian physics, a more correct version, Galilean physics*(a corrected version of Galilean physics), would give more accurate empirical explanations or predictions. Second, T_R* must be as explanatorily successful as T_R in its domain of inquiry. A more correct Galilean physics* must give correct explanations or predictions about the facts about which the original Galilean physics could give explanations or predictions. As a third point, Schaffner adds that the correct structure T_R* must express a large "positive analogy" (Schaffner 1967, p.144) to T_R, remaining expressible in the vocabulary of T_R; that is, T_R* must be translatable to T_R. But as Bickle criticizes (Bickle 1998, pp. 25f.), when the T_R* is evidently false (e.g., the phlogiston theory of combustion), translation is hardly possible between the two theories. It is hence doubtful whether we can hold this third point.

This controversial account presented by Schaffner was modified by Hooker, though in quite an abstract fashion (Hooker 1981). In Hooker's account, intertheoretic reduction involves deduction, but the conclusion of the derivation is not the T_R, nor a corrected version T_R* of T_R. What is derived instead is an *analog structure* I_B of T_R; I_B is within the vocabulary and conceptual framework of the reducing theory T_B, and is designed to mimic the syntactic structure of T_R. I_B must match the domain of application of T_R. As I_B is specified by the vocabulary and resources of T_B, there is no need for the "bridge principles" or "corresponding rules" to produce the required derivation. This technical arrangement eliminates the well-known vexing problem of specifying the logical

and ontological status of bridge principles. Further, by dropping Schaffner's third point (the translatability issue between T_R and T_R*) Hooker seems to cope with the case in which T_R is false. But Hooker's further argument remains unclear. Of Hooker's proposal, Paul Churchland presents an interesting and enlightening interpretation. According to Churchland:

> The point of a reduction, according to this view, is to show that the new or more comprehensive theory contains explanatory and predictive resources that parallel, *to a relevant degree of exactness*, the explanatory and predictive resources of the reduced theory (Churchland 1985, p.11) (italics added).

Churchland implies that Hooker's argument suggests that when a reducing theory takes over a reduced theory, the latter's explanatory and predictive power not be lost. But the question is how "relevant" the "relevant degree of exactness" is.

Bickle gives us a way to put together all these arguments on reductive relation (Bickle 1996, 1998). In Bickle's view, the discussions by Hooker and others have paved the way to the following understanding: we can judge what reduction is being called on according to how distant T_R is from its ideally corrected version (i.e., Schaffner's T_R*). Here is presented a spectrum model of reduction in which reduction is considered to be the matter of degree; it is a continuum from smooth reduction (retaining a reduced theory), through revisionary reduction (modifying or revising a reduced theory) to complete displacement (eliminating a reduced theory). In this *revisionist* reductionism, the treatment of a reduced theory depends on the correctness of the theory. Bickle places in this spectrum, as instances, "the wave theory of light" vs. "Maxwell's electromagnetic theory," "Kepler's theory" vs. "Newtonian mechanics," and "classical equilibrium thermodynamics" vs. "statistical mechanics" (Bickle 1996, pp. 65-6).

By applying this understanding to the psychology- neuroscience relation, Bickle attempts to clarify the theoretical status of

psychology (Bickle 1998, pp. 200f.). Given the inaccurateness in psychology's concepts and expressions (so-called "coarse-grainedness"), psycho-neural reduction should be located somewhere in the middle of the spectrum. This specifies, Bickle points out, the following characteristics: (1) psychology explains cognitive acts/functions (albeit in a coarse-grained fashion), (2) psychological characterizations of phenomena receive corrections or revisions from neuroscience that is more accurate on the phenomena, (3) in practice, the psychology-neuroscience relation is exemplified by three conditions: *approximation*, *fragmentation*, and *co-evolutionary development*. I explain these conditions below.

Normally, psychological concepts are more coarse-grained than neuroscientific concepts; neuroscience is equipped with more detailed and accurate concepts and expressions than those in psychology. While psychological theories provide *roughly* correct characterizations of cognitive or mental phenomena, they thus *approximate* in a coarse-grained fashion what happens in the actual neural processes involved in the phenomena. The precise descriptions of the way these cognitive or mental processes are implemented in the neural system are beyond the scope and power of psychological theories. For example, a state characterized in psychology as "pleasure" is explained in more fine-grained neural terms referring to "prefrontal cortex," "limbic system," etc. For fine-grained characterizations of cognitive or mental phenomena, neuroscientific descriptions are absolutely necessary.

Second, coarse-grained psychology cannot be accurately and precisely differentiated or articulated on psychological or cognitive processes. Each theoretical posit in psychology *fragments* into many actual distinct neural processes. Coarse-grained psychological characterizations or concepts may subdivide into more precise neural characterizations. For example, a simple psychological notion "memory consolidation" has turned out to involve many neural processes such as "postsynaptic potentiation," "second messengers," "retrograde transmission." Reductive mapping of psychological characterizations onto their underlying

neural processes thus amounts to *fragmentation* into more fine-grained neural categories. On the other hand, in such cases, psychological categories can thereby become refined or well-structured. In other words, fragmentation leads to *bottom-up refinement or correction* of psychological characterizations and concepts. Through reduction, psychology becomes revised.

Approximation and fragmentation clearly indicate the influence of neural concepts or properties onto psychological concepts or properties. But there is also influence the other way around between the two. A look at theory-forming in history tells us that psychology, due to its *simple* characterizations, directs or guides neuroscience especially at the early stage of its investigations. The investigation of "memory consolidation" in neuroscience was only possible because the phenomenon was first specified in psychology, leading eventually to the discovery of the above-mentioned neural processes. This illustrates the possibility that coarse-grained psychological characterizations induce fine-grained neuroscientific characterizations — *psychology's influence on neuroscience*. Seen in actual contexts of theory-forming, approximation and fragmentation are coupled with mutual feedback and development, which Bickle called *co-evolutionary development*. To accept the idea of co-evolutionary development is to accept the idea of a bi-directional dependency relation between psychology and neuroscience, which is stronger than one-way dependence. We can thus hold the above (X):

(X) The psycho-neural relation must be stronger than one-way dependence.

Notice that Bickle's argument is logically based on the difference of the grainedness of vocabulary between a reduced theory and a reducing theory — what I called above the asymmetry of terminological grainedness. Approximation, fragmentation, and co-evolution hold even if neural properties do not have *ontological* influence on psychological properties or vice versa; the relation at issue is of *explanation*, that is, of just *epistemological* nature. We do not have to specify the ontological status of the influence.

Though Bickle points out that there are influences from psychology to neuroscience, they are epistemological or methodological ones. Hence, unlike many proponents of "downward" influence from "higher properties," we do *not* commit downward causation here. This idea is stated by (Y):

> (Y) Psychological properties do not exercise downward causation on neural properties, but instead exercise a different type of influence on neural properties: explanatory influence.

The relationship is explanatory and methodological. As is notably pointed out by the term "co-evolution," psychology retains its significance even after reduction has obtained; reduction does *not* make a reduced theory redundant. The above (Z) states this pluralistic structure of explanation in psycho-neural investigation:

> (Z) Psychological properties and neural properties both have theoretical significance — in particular for the development of neuroscience.

4. Two-way dependence between the ethical and the neural

This whole argument, which derives from Bickle's revisionist reductionism, is simply *epistemological* and *methodological* just as Bickle's argument is; it is based on the difference of the grainedness of vocabularies of a reduced theory and a reducing theory — on the asymmetry of terminological grainedness in my term. As I have pointed out above, this argument does not involve any *ontological* commitment. It follows from this that this argument can apply to the relation between any type of properties and neural properties as long as the relation at issue is of equally explanatory nature.

I have assumed that the relation between neuroscience and ethics is considered in terms of their terminological grainedness. If neuroscientists contend that their theory has tremendous "explanatory power," as we have seen at the beginning of this paper, and is designed to explain any moral phenomena when reduction

holds (in particular, reduction in the strict sense holds), implying that neuroscientific theory is much more fine-grained, the relation can be discussed in terms of the asymmetry of terminological grainedness between neuroscience and ethics. The above characteristics (1), (2), and (3) including the three conditions (approximation, fragmentation, and co-evolutionary development) apply to the discussion of ethico-neural relations as well. The line of argument allows us to make claims similar to (X) to (Z) in the context of ethico-neural relations, establishing a pluralistic, multi-layered explanatory view in the field of neuroethics. In the view, even when reduction is construed in the strict sense, ethical properties play a significant explanatory role. This entails that ethical characterizations can lead ethico-neural discussions, i.e., discussions in neuroethics, and contribute to neuroscientific investigations of ethical behavior. In other words, if this argument, which involves "co-evolutionary development," is correct, the development of neuroscience does not just allow for ethics but also requires it. Neuroscientific investigations of ethical behavior must be guided by ethical consideration.

My argument presented in this paper provides a theorization of the two-way dependence between ethical properties and neuroscientific properties — including the underestimated influence of ethical properties on neural properties. Contrary to what the neuroscientists who have faith in reduction in the strict sense believe, it shows the possibility of ethics leading neuroscientific studies.

Notes

1 Following the general tradition in Anglo-Saxon philosophy, I will not distinguish "ethical" from "moral" and "ethics" from "morality" in this paper, unless my argument really requires me to do so.

2 As will be shown later, reduction in the sense discussed in this paper is primarily reduction in science and *intertheoretic* reduction in the Nagelian sense, as I will explain later. More importantly, it's *explanatory* reduction. As a result of *explanatory* reduction, *ontological* reduction may take place, but it is secondary.

3 To formulate his idea I use the oft-quoted term “supervenience,” though Moore himself did not use the term. I owe this formulation to Davidson (1973).
4 Precisely speaking, Roskies distinguishes “mechanism” and “determinism.”

Reference

Bickle, J. 1996. “New Wave Psychophysical Reductionism and the Methodological Caveats,” *Philosophy and Phenomenological Research* 56 (1):57-78.
———— 1998. *Psychoneural Reduction: The New Wave*, Cambridge, MA: MIT Press.
Churchland, P. M. 1985. “Reduction, Qualia, and the Direct Introspection of Brain States,” *Journal of Philosophy*, 82: 8-28.
Davidson, D. 1973. “The Material Mind”, in Davidson 1980.
———— 1980. *Essays on Actions and Events*, Oxford: Clarendon Press
Hooker, C. A. 1981. “Toward a General Theory of Reduction. Part I: Historical and Scientific Setting. Part II: Identity in Reduction. Part III: Cross-categorical Reduction,” *Dialogue* 20: 38-59, 201-36, 496-529.
Nagel, E. 1961. *The Structure of Science*, New York: Harcourt, Brace, and World.
Roskies, A. 2002. “Neuroethics for the New Millennium,” *Neuron*; vol. 35, 1: 21-3.
———— 2006. “Neuroscientific challenges to free will and responsibility,” *Trends in Cognitive Sciences*; Vol.10, No.9: 419-23.
Schaffner, K. F. 1967. “Approach to Reduction,” *Philosophy of Science* 34:137-47.