

Review of *Neuroethics and the Scientific Revision of Common Sense*

Paul Boswell

Université de Montréal

Gligorov, Nada. *Neuroethics and the Scientific Revision of Common Sense*.
Dordrecht: Springer, 2016. 169 pp. USD \$99.99 (hardcover), \$79.99 (ebook).

Research in neuroethics generally presupposes that investigation into the brain and the nature of cognition could usurp our current understanding of moral psychology. The concept of free will, for instance, is arguably the most endangered, with prominent authors arguing that the common sense notion of free will is libertarian (Greene & Cohen 2004) or requires our conscious willing to control our behavior (Wegner 2002). Both attributed theses have proved doubtful for well-known reasons.

Certainly there has been much to learn from the flourishing field of neuroethics. But in *Neuroethics and the Scientific Revision of Common Sense*, Nada Gligorov contends that the moral to draw is *not* that common sense psychological notions are somehow incompatible with scientific theorizing, for in her view such concepts are malleable. They can adapt to science's findings. In service of this general point, which she first articulates in Chapter 2, Gligorov investigates six separate debates in neuroethics: free will, the impact of cognitive enhancement on personal identity, the impact of memory management on personal identity, mental privacy, the supposed subjectivity of pain, and the definition of death (Chapters 3-8, respectively).

Below I'll first introduce these last six chapters. Then I'll make some general remarks about the book, and afterwards swing back to discuss Chapter 2 and the book's overall thesis.

Gligorov's main targets in Chapter 3 are those who would argue that the famous intention-predicting experiments by Libet et al. (1983) and Soon et al.

(2008) show the common sense notion of free will to be an illusion.¹ Against them Gligorov points to a number of unconscious processes, such as subliminal priming, that have a broad influence in shaping our behavior, and suggests that results from Murray & Nahmias (2014) indicate that the folk would consider these behaviors freely willed (pp. 42-5).

Chapters 4 and 5 concern the threat that new technologies may pose to the self. In Chapter 4 Gligorov argues that narrative identity, not numerical identity, is the proper locus of concern when discussing the risks of cognitive enhancement, and that even so the risks to narrative identity are similar to the risks of adopting a new discipline (yoga, say) and discovering you don't like how it's changed you. In service of this argument she outlines a very liberal conception of narrative identity (pp. 66-8) on which a person determines for herself which of any of her characteristics constitute her identity through simply endorsing them as part of her autobiographical self-conception.

Chapter 5 aims to rebut two related concerns about the permissibility of memory-modifying technologies such as propranolol, which may attenuate recall of traumatic events. The first, broached by Liao & Wasserman (2007) and Liao & Sandberg (2008), is that memory-altering technologies may cause us to lose identity-grounding knowledge of ourselves. Gligorov takes studies in the wake of Loftus & Palmer (1974) to show that memory is unreliable anyway, and as such provides a bad evidentiary basis for our sense of self (p. 84). The second, due to Erler (2011), is that memory modification could cause us to live inauthentically. To this Gligorov replies that the characteristics of one's true self are based in one's personal endorsement of them, and reckons that voluntary memory modification is likely to lead to *more* authentic living.²

Chapter 6 deals with brain imaging, especially fMRI, and its potential impact on the privacy of mental states. But in spite of what the reader may expect, Gligorov's main concern is whether this technology threatens traditional theses that attribute a special metaphysical or epistemic privacy to mental states; she

¹Note that Chapter 3 opens with "Free will can be defined as the ability to do otherwise" (p. 35). Most philosophers, however, take the ability to do otherwise to be neither necessary nor sufficient for free will, especially post-Frankfurt (1969).

²Sometimes Gligorov speaks of authenticity in terms of "being truthful" about a characteristic. Thus there are two virtues of truthfulness with respect to the self under discussion: being *honest* about oneself, and *staying true* to oneself.

concludes in the negative. Chapter 7 argues for a number of objectifying theses about pain, such as that first-personal pain reports are corrigible.

Chapter 8, on brain death, largely departs from the work of James Bernat (Bernat et al. 1981; Bernat 2006), and a part of it has already appeared in this journal (Gligorov 2016). Gligorov argues that biological death, the death of the whole organism, is whole-brain death. Bernat's view is that death is the loss of an organism's critical functions, among which he suggests are consciousness, control of metabolic processes, and control of homeostatic systems (Bernat 2006, p. 36). Gligorov generalizes these to bodily integration, psychophysical integration, and psychological integration, and holds that all three functions must be lost for the human organism to die.

My impression is that Gligorov's arguments on cognitive enhancement, memory modification, and brain death make helpful contributions to the respective literatures on those subjects, and ought to be read by specialists in those areas. But below I'll mention some areas of concern.

The dialectic of the book can be difficult to follow. For an example, recall that in Ch. 5 Gligorov argues that memory cannot ground narrative identity because, in her view, the psychological evidence indicates it is not a dependable source of self-knowledge. However, it is not obvious how this argument coheres with Gligorov's views in Chapter 4. There she argues narrative identity is grounded only in the person's endorsement of traits, and she emphasizes that the scope for criticism of one's narrative identity on the basis of inaccuracy is very limited, adding that what we think of ourselves is often "not based on facts, but on interpretation" (p. 67). On this view it's not clear why we should expect any evidentiary basis for narrative identity or sense of self; considerations of accuracy would appear to have little rational bearing on forming narrative identities. Indeed one might think that poor self-knowledge makes it easier to maintain one's chosen narrative identity. In this light, the focus on the alleged unreliability of memory in Chapter 5 seems misplaced, for on Gligorov's theory we should expect narrative identity to be preserved through memory modification *whether or not* memory is a reliable source of self-knowledge.

Readers of the book will encounter more exposition of scientific and philosophical views than they might expect. As an example, by my reckoning 75% (15 of 20 pages) of Chapter 6 is introduction, recapitulation, or preliminary discus-

sion. Moreover, the level of detail at which this background is presented does not always seem to be demanded by Gligorov's arguments, especially when they are focused on the recent neuroethics literature. Does the specificity theory of pain need yet another refutation (§7.2)? Why the short history of dualism from Descartes to Chalmers in §6.3?

Of course, because neuroethics combines neuroscience and philosophy, it necessarily addresses itself to readers some of whom will have competence in one but not both disciplines, and so additional background is often desirable. Background might also help a work serve as a general introduction to the subjects it treats. But both of these goals are compromised in this work by a number of misreadings and questionable interpretations. To take just three examples: Gligorov characterizes Fodor (1974) as holding, along with Paul Churchland, that "thoughts and beliefs do not exist in the brain" (p. 109). That statement is ambiguous; is the claim that psychological kinds like *belief* are not neurological kinds, or that psychological events are not identical to any of our neurological events? Fodor can be attributed the former and not the latter interpretation, but the comparison with the eliminativist Churchland may incline the reader towards the latter. Also, Gligorov's two examples of pain asymbolia are stress-induced analgesia and lobotomy as a treatment for pain (p. 120), conditions quite distinct from asymbolia, a cortico-limbic disconnection syndrome likely brought on by insular damage.³

And most significantly for her overall argument, a Lewisian platitude (for the definition of a theoretical term) concerning psychology is not merely something such that "many of us know it and use it to explain and predict aspects of human behavior" (p. 26). It specifies a causal role that is common knowledge (everyone knows it, everyone knows that everyone knows it, etc.) and needs to have an air of analyticity about it, for the platitudes together are meant to establish the *meaning* of a term.⁴ (Example: "Depression tends to make people socially withdrawn.") Gligorov takes such platitudes to constitute folk psychology. But does it sound *analytic* that runner's high is explained by neurotransmitters in the brain, as Gligorov suggests is part of current folk psychology (p. 23)? Of course

³The modern *locus classicus* on pain asymbolia is Berthier et al. (1988).

⁴Lewis (1972, p. 256). Note also that Gligorov often calls Lewis's O-terms "observational terms", though Lewis himself forbids this characterization (*ibid.*, p. 250).

it is *possible* for a concept in folk psychology to be definable in scientific terms, but not all explanatory, predictive folk commonplaces about folk mental states are meaning-determining Lewisian platitudes.

This takes me to Gligorov's overall thesis about common sense concepts. To the extent that they put pressure on common sense psychological concepts, most of Gligorov's interlocutors throughout the book do so piecemeal. They might argue that we do not have free will, but they will generally leave belief untouched. However, Gligorov's direct target in Chapter 2 is the much stronger thesis that *all* common sense psychological concepts are incompatible with science, as Churchland's (1989) eliminative materialism entails.

Gligorov's strategy to defeat this skepticism begins with a Sellarsian conceptual anti-atomism: one cannot possess only one concept, for the possession of any concept requires the acceptance of a background theory expressed with additional concepts (Sellars 1956). Change the background theory, and roughly speaking, the meaning of the concept changes. Gligorov views common sense psychological concepts like free will as concepts with changeable background theories. New results in science can produce local discrepancies with common sense psychology (p. 166), but the need to use the folk psychological states to explain and predict behavior ensures that these same concepts will, over time and across cultural contexts, become associated with background theories that accommodate the results.⁵

This is a very attractive idea, but Gligorov's way of cashing it out faces a central difficulty. Gligorov further endorses David Lewis's philosophical functionalism, which holds roughly that terms for psychological states are defined implicitly by the (aforementioned) platitudes that concern those states.⁶ According to this theory the term "melancholy" as used by humorist psychologists denotes the unique thing caused by an excess of black bile and which causes one to be socially withdrawn; what's more, the term's whole meaning is given by that role.⁷ Since black bile does not exist, it follows from Lewis's theory that the melan-

⁵Thanks to Nada Gligorov for a clarification on this point.

⁶See Lewis (1970, 1972). Another issue, which I lack the space to address, is whether extending Lewis's causal analysis of mental terms can be extended to normatively-inflected notions like free will, as Gligorov seems to think it can (p. 28). Philosophers across the board have been skeptical of causal analyses of moral concepts, including Frank Jackson (1998).

⁷Or something close to this role; I am simplifying for the purposes of the example.

choly such theorists spoke of does not exist, for the role assigned to “melancholy” is unoccupied. Eliminativists about humorist melancholy are right. Indeed the problem is worse: if any of the terms the humorist theory introduces are denotationless, the postulate corresponding to the humorist theory is false (see Lewis 1970, p. 80). The fact that we would later use the word “melancholy” as a synonym for depression, which does exist, does not address the problem. We cannot make humorism true by changing the meanings of words.

On Lewis’s theory, it would appear that allowing the meanings of folk psychological terms or concepts to change when one of the platitudes of folk psychology is shown to be false is no way to save folk psychology or folk concepts, any more than altering the meaning of “melancholy” is a way to save humorist psychology. Put another way, it seems that Gligorov’s theory requires thinkers to maintain the *same* theoretical concept even as the meaning-determining platitudes associated with it change significantly. It is not clear how Lewis’s theory can permit that.

Lewis does worry about what to do if only near-realizations of platitudes are true. He writes that the meanings of terms could be set by a disjunction of closely related causal roles (Lewis 1972, p. 256), so that a few peripheral falsehoods will not matter. Since Gligorov thinks that eliminativism would be justified for a mental state only if the part of folk theory that defines it is radically false, perhaps this is the kind of maneuver she has in mind. The main problem with this approach is that eliminativists are inclined to think that folk mental theories are indeed massively or fundamentally wrong, so interpreting Gligorov in this way will not address their concerns. Lewis (1970, p. 94) briefly floats the suggestion that the meaning of theoretical terms should change as the theory is revised, but he is ambivalent about it, and only countenances slight changes — not the radical changes eliminativists think are required. So it seems that fully articulating the kind of solution Gligorov has in mind would require going significantly beyond Lewis.

References

- Bernat, J. L., C. M. Culver & B. Gert. 1981. “On the definition and criterion of death.” *Annals of Internal Medicine*, vol. 94 (3): 389–394.

Review of Gligorov: *Neuroethics and the Scientific ...*

- Bernat, James L. 2006. "The Whole-Brain Concept of Death Remains Optimum Public Policy." *The Journal of Law, Medicine & Ethics*, vol. 34 (1): 35–43.
- Berthier, Marchcelo, Sergio Starkstein & Ramon Leiguarda. 1988. "Asymbolia for pain: A sensory-limbic disconnection syndrome." *Annals of Neurology*, vol. 24 (1): 41–49.
- Churchland, Paul M. 1989. *A Neurocomputational Perspective: The Nature of the Mind and the Structure of Science*. MIT Press, Cambridge, MA.
- Erler, Alexandre. 2011. "Does Memory Modification Threaten Our Authenticity?" *Neuroethics*, vol. 4 (3): 235–249.
- Fodor, J. A. 1974. "Special sciences (or: The disunity of science as a working hypothesis)." *Synthese*, vol. 28 (2): 97–115.
- Frankfurt, Harry G. 1969. "Alternate Possibilities and Moral Responsibility." *The Journal of Philosophy*, vol. 66 (23): 829–839.
- Gligorov, Nada. 2016. "A Defense of Brain Death." *Neuroethics*, vol. 9 (2): 119–127.
- Greene, Joshua & Jonathan Cohen. 2004. "For the law, neuroscience changes nothing and everything." *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 359 (1451): 1775–1785.
- Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford University Press, New York, NY.
- Lewis, David. 1970. "How to Define Theoretical Terms." *Journal of Philosophy*, vol. 67 (13): 427–446. Reprinted in Lewis, David. 1983. *Philosophical Papers*, vol. 1. Oxford University Press, New York, NY, pp. 78–95. References to this edition.
- . 1972. "Psychophysical and Theoretical Identifications." *Australasian Journal of Philosophy*, vol. 50 (December): 249–58.
- Liao, S. Matthew & Anders Sandberg. 2008. "The Normativity of Memory Modification." *Neuroethics*, vol. 1 (2): 85–99.
- Liao, S. Matthew & David T. Wasserman. 2007. "Neuroethical Concerns About Moderating Traumatic Memories." *American Journal of Bioethics*, vol. 7 (9): 38–40.
- Libet, B., C. A. Gleason, E. W. Wright & D. K. Pearl. 1983. "Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act." *Brain: A Journal of Neurology*, vol. 106 (Pt 3): 623–642.

Review of Gligorov: *Neuroethics and the Scientific ...*

- Loftus, Elizabeth F. & John C. Palmer. 1974. "Reconstruction of Automobile Destruction: An Example of the Interaction Between Language and Memory." *Journal of Verbal Learning and Verbal Behavior*, vol. 13: 585–589.
- Murray, Dylan & Eddy Nahmias. 2014. "Explaining Away Incompatibilist Intuitions." *Philosophy and Phenomenological Research*, vol. 88 (2): 434–467.
- Sellars, Wilfrid S. 1956. "Empiricism and the Philosophy of Mind." *Minnesota Studies in the Philosophy of Science*, vol. 1: 253–329.
- Soon, Chun Siong, Marcel Brass, Hans-Jochen Heinze & John-Dylan Haynes. 2008. "Unconscious determinants of free decisions in the human brain." *Nature Neuroscience*, vol. 11 (5): 543–545.
- Wegner, Daniel M. 2002. *The Illusion of Conscious Will*. MIT Press, Cambridge, MA.