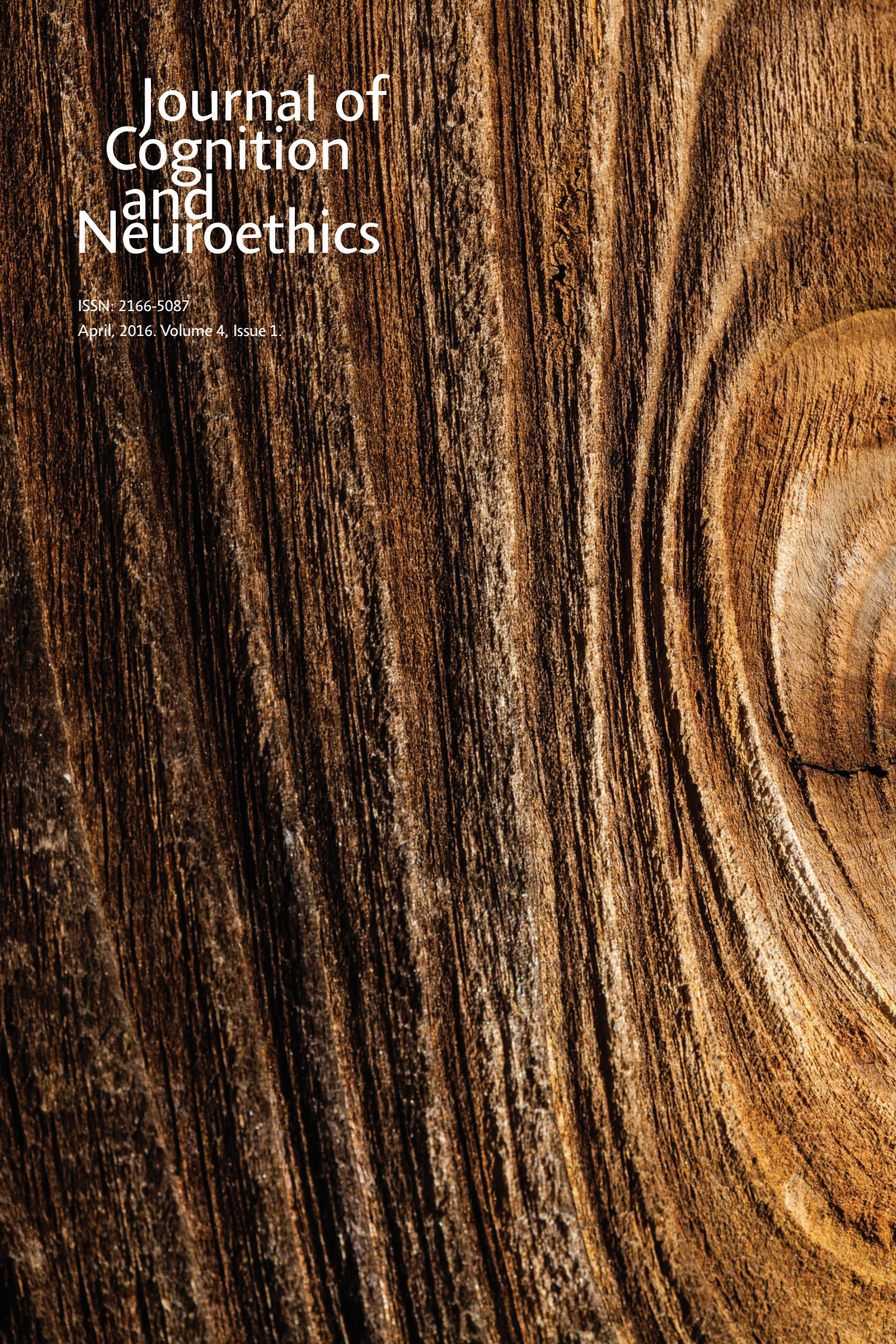
Journal of Cognition and Neuroethics

ISSN: 2166-5087
April, 2016. Volume 4, Issue 1.



Journal of Cognition and Neuroethics

Managing Editor
Jami L. Anderson

Production Editor
Zea Miller

Publication Details
Volume 4, Issue 1 was digitally published in April of 2016 from Flint, Michigan, under ISSN 2166-5087.

The *Journal of Cognition and Neuroethics* is produced by the Center for Cognition and Neuroethics. For more on CCN or this journal, please visit cognethic.org.

Center for Cognition and Neuroethics
University of Michigan-Flint
Philosophy Department
544 French Hall
303 East Kearsley Street
Flint, MI 48502-1950

Table of Contents

Journal of Cognition and Neuroethics

Consciousness, Neuroimaging and Personhood: Current and Future Neuroethical Challenges

James Beauregard
Rivier University

Macksood Aftab
Michigan State University
Central Michigan University

Aamna Sajid
Michigan State University

Biographies

James Beauregard Ph.D. is a Clinical Neuropsychologist and Lecturer at Rivier University, Nashua, NH. Macksood Aftab D.O. is a Neuroradiologist and Assistant Clinical Professor at Michigan State University College of Human Medicine and Central Michigan University College of Medicine.

Publication Details

Journal of Cognition and Neuroethics (ISSN: 2166-5087). April, 2016. Volume 4, Issue 1.

Citation

Beauregard, James, Macksood Aftab, and Amna Sajid. 2016. "Consciousness, Neuroimaging and Personhood: Current and Future Neuroethical Challenges." *Journal of Cognition and Neuroethics* 4 (1): 1–11.

Consciousness, Neuroimaging and Personhood: Current and Future Neuroethical Challenges

James Beauregard, Macksood Aftab, and Amna Sajid

Abstract
Neuroimaging has advanced our understanding of the biological bases of consciousness. At the same time, it is vital that these technologies be kept in proper perspective to avoid unsupportable claims and public misperceptions of its capacities and utility in health care and research. This presentation provides a philosophical anthropological context in which to examine current neuroimaging knowledge of consciousness and then examines the science of neuroimaging and the neuroethical considerations it raises.

Keywords
Bioethics, Juan Manuel Burgos, Consciousness, Neuroimaging, Muhammad Iqbal, Personalism

1. Introduction: Beginning in Context: Philosophical Anthropology

Perspective and realistic appraisal of technology are essential in neuroethics, particularly, maintaining an appropriate balance between the "neuro" and the "ethics" aspects of this discipline. Adrianna Gini et al have recently addressed these issues in an article titled, "Keeping the human: Neuroethics and the conciliation of dissident values in the 21st century," where they have written

> As neuroscience accumulates ever more factual information on brain operation, the normative problems raised by these findings become increasingly acute. In the past decade, as neuroscience has moved from peripheral sensory and motoric investigations to more central brain operation, ethical trends have shifted from an ethics concerned with the practice of neuroscience to interpretive aspects of human anthropology. No longer solely concerned with pharmacological enhancement, increasingly it reflects on the substance of our self interpretation.… The authors concern is that in our rush to change, we not overlook the prize already in our possession, the human mind in its manifold expression and oriented naturally to meaning and transcendence through beauty, truth and ethics. Keeping the human is more than a recommendation,

> it is a recognition that what is kept will be the patrimony that we bequeath to our future. (Gini et al. 2015)

Neuroimaging is a frontline medical and research technology where maintaining the proper balance between persons and technology has far-reaching implications. Neuroimaging is widely employed in the clinical care of individuals with disorders of consciousness, and has already demonstrated its utility in deepening our understanding of the *biological* bases of consciousness (Blumenfeld 2010; Kolb and Wishaw 2014). New technologies create new possibilities, which may not be immediately evident, and simultaneously raise new ethical questions.

In order to place neuroimaging technology and its ethical implications in context, we would like to approach the topic of neuroimaging ethics in three ways:

1. *Philosophical Anthropology* as a broader context to help guide an ethical framework for the use of neuroimaging in the study of consciousness and its disorders.

2. The necessity of an accurate understanding of the *nature and the limitations* of neuroimaging technology so that any ethical thinking in this area is grounded in accurate information.

3. As one looks across the literature on the ethics of neuroimaging, there is a topic that emerges again and again, one the stands at the border between science and science fiction. That topic is "mind reading," or the *detection of mental content*, which raises a host of bioethical issues.

First, we will touch on the field of philosophical anthropology from two different personalist perspectives, one European, one Islamic. Broadly speaking, Personalism is any philosophical system that takes the notion of person as its starting point, and as the key to understanding the major concerns of philosophy including ethics and anthropology (Buford). Personalist philosophy traces its roots to the ancient world, both East and West, to Hindu, Buddhist and Confucian thought, Judaism, Christianity and Islam, and the long tradition of philosophy bequeathed to us by Greece.

1.1 Person as Unity of Body, Mind, Spirit

A contemporary expression of this tradition is the Modern Ontological Personalism of Juan Manuel Burgos. Briefly, he envisions persons in an integrated and holistic fashion,

considering multiple *aspects* of persons including, in his terms, body, psyche and spirit (Burgos 2013; Burgos 2012). There are several aspects of his philosophy relevant to our considerations today. The first is that in order to comprehend human beings, it is necessary to think in categories specific to persons, rather than beginning with a vision of persons as animals–plus, i.e., "rational animals." For Burgos, reason, freedom and emotion have distinctly human/personal manifestations that are different from those in the animal world. When he writes of emotion, for example, he sees it as an original and often undervalued aspect of persons that manifests in the three levels of person he describes, our bodies, our mental life (psyche) and our highest human capacities, such as love (spirit, in his vocabulary) (Burgos 2013).

1.2 Embodiment and Action

For Burgos, the physical *aspect* of person alone is insufficient for a comprehensive understanding of persons. Action is also central. Persons are known through action, manifested through our bodies. We know other persons through embodied action, in other words, through personal activity. Interaction and reflection on personhood fully conceived has an ethical dimension, it can yield moral norms. It is here that we touch on the interaction of the personal body with medicine and medical technology, as we attempt to understand and seek the good of health, well-being and human flourishing.

2. Neuroimaging and Neuroethics

The recent advancement in neuroimaging has opened up new avenues to study consciousness and interrogate the mind-body problem. In the clinical setting it provides for the assessment of consciousness in ways previously unavailable. This technology, however, raises its own sets of ethical issues which relate directly to the use of the technology and its scientific basis. The clinical use of this advanced functional imaging carries significant philosophical implications.

2.1 Introduction to functional MRI

Functional MRI is a method of imaging which combines both structural and functional imaging of the brain. It allows for the localization brain activity to specific regions of the brain. It has allowed for the correlation of cognitive activity with specific areas of the brain parenchyma.

Functional MRI makes use of the detection of increased cerebral blood flow and oxygen uptake in areas of the brain which are metabolically active. The implication being

that areas which are accumulating and metabolizing oxygen are the areas which are being activated for specific sensory, motor or cognitive tasks. Functional MRI is thus a measure of the fuel uptake within active brain areas; it is not a direct measure of neuronal activity output. Nevertheless, this technology provides for the ability to assess for brain function in a way which was not possible before. A patient's motor response such as eye or limb motion is no longer required to assess for consciousness, instead it can potentially be directly assessed by assessing for areas of activation in the brain (Laureys et al. 2009).

2.2 Assessing Consciousness

Clinically the assessment of consciousness relies on the evaluation for arousal and awareness. Both must be present. The disorders of consciousness include coma, vegetative state, minimally conscious state and locked-in syndrome. In coma neither arousal or awareness are present. In a vegetative state patients may be arousable but lacks awareness. In a minimally conscious state patients demonstrate inconsistent and intermittent evidence of awareness. In locked-in syndrome patients are fully conscious but are unable to communicate with outside world due to disruption of brain communication pathways in the brainstem. Some locked-in syndrome patients are able to communicate with limited eye movement only. It is in these patients that fMRI has the most promise.

Awareness is, however, a subjective concept which is most immediately assessed on a first person basis. Only the individual is fully aware of their thoughts, mental status and cognitive ability. However, in the clinical setting when the patient is not fully functional and not able to fully communicate, the first person account is not available. In this setting an objective measure of consciousness is required in order to make clinical decisions regarding the patient's care and for prognostic purposes. This assessment requires some sort of response on the part of the patient based upon outside input. The idea is to assess for "intentional ability" on part of the patient. Is the patient able to understand and make decisions? If they can engage in "willed action" or demonstrate intentionality then awareness and consciousness would be established. This had traditionally been assessed on clinical grounds by a neurologic exam. More recently advanced neuroimaging has been used to assess for consciousness which does not require a motor/physical response on the part of the patient (Lloyd 2002; Bardin et al. 2011).

2.3 Philosophical Implications

Prior to proceeding to the use of functional MRI in consciousness, it is worth reflecting on the principles being applied in the determination of consciousness. Clinically

consciousness has been defined as the ability to engage in willed action. Those who are not in coma but lack consciousness are considered to be in a vegetative state and thus carry a very poor prognosis. This has implications for treatment and rehabilitation therapy which is considerably reduced for those unable to demonstrate intentionality. Thus, in practice intentionality is used as a measure of a person who is worth saving. This implies to some degree that what it is to be fully human is to be able to engage in willed action, the absence of which downgrades the resources placed into saving such a life. Willed action is then implicitly being used as the gold-standard for the highest state of consciousness and of personhood. Implying that what it is to be fully human is to engage in decisions and free will.

This echoes the thought of the 20th century Muslim philosopher Muhammad Iqbal who articulated a theory of the self which relies on the exercise of free will as the defining property of humanity. He argued that the mind and body come together in the exercise of free action and therefore it is misleading to create an artificial dichotomy between the physical and metaphysical elements of man (dualism). Neither define the human, rather what defines him/her is the ability to engage in free action and it is via such action that a man or woman is able to reach the full pinnacle of their human potential (Iqbal 2011).

2.4 fMRI and Consciousness

In the absence of functional neuroimaging, an assessment of a patient's ability to engage in wilful thought or action required a clinical neurologic exam which required the patient to demonstrate via physical action (such as finger or eye movement) that they understand, communicate and can make choices. Patients who lack appropriate control of their limbs and bodily function due to brain injury or stroke may not be able to communicate their thoughts in this way. Thus making it difficult to differentiate patients in a vegetative state from minimally conscious states and locked-in syndromes. With functional MRI brain states can be assessed directly by examining brain activation and opens up new possibilities in the study of consciousness.

Consciousness is assessed at three levels with functional imaging: 1) passive, 2) active, and 3) communicating. At the passive level brain response is assessed after a sensory input. So an image shown to the patient should elicit a response in the visual cortex. This provides a baseline for the functioning of the neuronal hardware. However, it is unable to disentangle automatic responses from voluntary conscious brain activation.

In the active experiments, patients are asked to engage in a particular thought to assess for command following via imagery tasks. They may be asked to imagine playing

tennis or imagine walking through their home. Their ability to selectively activate different brain areas provides evidence for voluntary modulation of brain function. Modulation of brain function is a strong indicator of a patient's ability to express a desire for willed action and free thought.

The highest level of consciousness is communication which is assessed via asking the patient to answer questions. In response to a question the patient is asked to engage imagery tasks. The questions have typically been ones with known answers such as place of birth, or number of siblings to assess for the accuracy of this method. The patient's ability to communicate then is considered clear evidence of consciousness and of the ability to engage in willed action.

Despite these advances there are some limitations to the functional MRI studies. Some patients have been upgraded from a vegetative state to a minimally conscious state by the use of fMRI immensely affecting their prognosis and treatment plan. However, relying solely on fMRI the majority of minimally conscious patients (MSC), as determined by clinical exam, could not be detected. Therefore, clinical exam remains the most sensitive study to assess for MSC even though it relies on neurologic exams. Furthermore, it has been noted that many healthy volunteers are able to perform physical tasks without being able to successfully perform the corresponding imagery tasks on fMRI. This further lowers the sensitivity of fMRI in detecting consciousness. Much of these limitations may relate to the limited about of data currently available and as the techniques for assessment improve the sensitivity of fMRI would likely improve as well.

2.5 The Hard Question

Does fMRI help answer the hard question of consciousness? i.e., how does neuronal activity generate a metaphysical thought or emotion? No. It simply provides correlation between areas of brain activation and the corresponding thought. Even so its clinical use has important clinical and philosophic implications. The most important of which is the reliance on free willed action as the gold standard test for consciousness. By extension this implies that what it is to be fully human is to be have the capacity to engage in intentional thought. This is an important advance, where a human is not being defined by the presence of a metaphysical entity such as a non-material soul or physical entity such as presence of certain body parts. Instead free will becomes the hallmark of humanity.

3. Current and Future Neuroethical Challenges for Neuroimaging

Neuroscience itself is changing. Going back to the article from which we quoted earlier, the authors note that

> A new, more integrated phase is about to begin. For 100 years neuroscience has labored to understand the constituency of the brain, it's functional units, their operations and how they interact to build up the brain. It's philosophy 'principale' was predicated on a part-determines-whole approach that informed the research directed to the manner in which the brain was built from the bottom up, an operational philosophy termed Neuroreductionism. Instead, the new neuroscience considers the operations of hundreds of thousands of neurons working in unison and the manner in which their concerted operation constrains output, a philosophy of systems and downwardly operative effects. Connectomes, the term for large – scale circuit structures are now variously explored. (Gini et al. 2015)

Our argument is that if we begin with a broader notion of the whole human person, these changes will have a sound context which allows for the creation of a comprehensive ethical vision of dealing with new technologies as they emerge.

At the outset, we mentioned mind reading – that may seem at first more appropriate to the science fiction conference that was recently held at the Center for Cognition and Neuroethics. Yet, in a rudimentary way, neuroscience has already begun to examine the content of thought both in the context of disorders of consciousness and in basic science research.

At a rudimentary level, neuroimaging studies have enabled the assessment of conscious activity in a medical/diagnostic context when disorders of consciousness occur, and when behavioral investigation alone may not give a fully accurate picture.

Functional MRI and PET studies already demonstrated the utility is a diagnostic and prognostic tool in assessing the Default Mode Network for patients in the acute stage of coma. EEG studies have helped distinguish between Vegetative State and Minimally Conscious State patients. And, neuroscience has established at least some rudimentary correlations between functional neuroimaging data and language content, raising the possibility that individuals with locked in syndrome might to be able to communicate more effectively with the outside world.

Consequently, we must ask a bioethical question: to what extent might neuroimaging in the near, or distant, future be able to accurately access an individual's

mental content, in other words, to "read" thought itself? Such a possibility raises a host of ethical questions about privacy, cognitive liberty, national security and the boundaries between public and private.

Privacy rights are a given in American law, but privacy is not absolute. Adina Roskies has recently raised the question of the nature of this right to privacy. If neuroimaging technology could assess the content of thought, what might this imply for medical use, employment, and prediction. Practically speaking, these issues arise in such areas as prediction of future neurologic illness, the possibility of accurate lie detection, predicting future dangerousness, criminal activity, and recidivism (Roskies 2015).

Presumably, individuals who, through neurological injury (e.g., Locked In Syndrome) are unable to communicate with others would want take advantage of such technology. If mental content could be accurately read, it would give such individuals the capacity to communicate far more quickly and effectively with others, maintaining relationships, giving informed consent, and making decisions about medical care.

This scenario assumes informed consent for the use of such technology in in individual who wants to establish communications with loved ones and health care providers. We should also ask, what if it should also become possible to determine the content of an individual's thought *against their will*? In this case, where would we draw the boundaries between public and private, and more fundamentally, how will the resulting moral and legal questions be framed?

Is there right to privacy in general, or some aspect of privacy, that is absolute, or are there conditions under which, for example, public safety might override individual privacy? There are situations in medicine where this is already the case, for example, the reporting of communicable diseases places the common good above individual medical confidentiality. In the field of mental health, privacy can be violated in cases of danger to self or others, as well as a need to contact relevant state agencies in suspected cases of abuse or neglect.

How would this public/private balance play out in the context of new technology? What happens if neuroimaging technology advances to the point where it could be employed as a national security tool? Imagine moving through security at the airport, which already includes physical search and one type of scan to detect weapons. If it were possible to assess the content of thought, should this now-routine airport scan also include neuroimaging to detect brain states indicating increased arousal or anxiety in a terrorist, or the content of his or her thought? And would this be done with or without an individual's consent?

September 11th in the United States and the recent terrorist attacks in Paris and San Bernadino CA take the matter a step further. Should such technology exist, what role might the neuroimaging access of mental content play in criminal investigations and terrorism investigations? Could it be used forcibly to extract information, and how would it be framed? Interrogation? Enhanced interrogation? Torture? One can envision the state's argument: "Public safety outweighs individual rights to mental privacy, and besides, it does less physical harm than waterboarding. It's a better way to get information from hostile and unwilling enemies who aren't US citizens and so not entitled to the protections of the Constitution and Bill of Rights." Who will make these decisions, and who will influence the debate?

4. Back to Context: Persons

In conclusion, we return to the notion of persons, of acting persons functioning in the world, embodied persons, as a category from which these considerations ought to be viewed. We have looked at neuroimaging as it plays a role in the study of consciousness, in disorders of consciousness, and as an interventional strategy for individuals in less than fully conscious states.

In terms of technological possibilities and neuroethical considerations, this is the tip of the iceberg, leading to deeper questions about human rights, privacy and cognitive liberty, and the relationship between the individual and the state. When we look at "neuroethics" these are questions the "neuro" aspect, the field of neuroscience on its own cannot answer. They can and must be dealt with in a broader context than the existence and potential uses of technology. Such questions can only be adequately approached from a sound philosophical anthropology and an ethical grounding that allows for the broader individual, social, medical, legal and security implications to be brought to the surface and addressed.

References

Bardin, J.C., Fins J.J., Katz D.I., et al. (2011) "Dissociations between behavioural and functional magnetic resonance imaging-based evaluations of cognitive function after brain injury." *Brain*, 134 (Part 3): 769-782.

Blumenfeld, Hal. 2010. *Neuroanatomy Through Clinical Cases*, Second Ed. Sunderland, MA: Sinauer Associates Inc.

Buford, Thomas O. (n.d.) "Personalism." In *Internet Encyclopedia of Philosophy,* http://www.iep.utm.edu/personal/.

Burgos, Juan Manuel. 2013. *Antropología: una guía para la existencia*. Madrid: Palabra.

Burgos, Juan Manuel. 2012. *Introdución al personalismo*. Madrid: Palabra.

Gini, A., D. Larivee, M. Farisco, and V.A. Sironi. 2015. "Keeping the human: Neuroethics and the conciliation of dissonant values in the 21st century." *Neuroscience and Neuroeconomics* 4: 1–10.

Iqbal, Muhammed. 2011. *The Reconstruction of Religious Thought in Islam*. Lahore: Institute of Islamic Culture.

Kolb, Bryan, and I.Q. Wishaw. 2014. *An Introduction to Brain and Behavior*, Fourth Ed. New York: Worth.

Laureys, S. M Boly, G Moonen, and P Maquet. 2009. *Coma*. Liege, Belgium: University of Liege Press.

Lloyd, D. 2002. "Functional MRI and the Study of Human Consciousness." *Journal of Cognitive Neuroscience* 14 (6): 818–831.

Roskies, Adina. 2015. "Mind Reading, Lie Detection and Privacy." In *Handbook of Neuroethics*, edited by J. Clauesn and N. Levy. Dordrecht: Springer.

Journal of Cognition and Neuroethics

The Last Modern Psychologist: Julian Jaynes' Search for Consciousness in the Natural World

Scott Greer
University of Prince Edward Island

Biography

Dr. Scott Greer is currently Associate Professor of Psychology at the University of Prince Edward Island in Charlottetown, PEI. He received his B.A. (Magna Cum Laude) degree from the University of Memphis, and his M.A. and Ph.D. degrees from York University in Toronto. Dr. Greer has published peer-reviewed articles in a variety of journals, including *Journal of the History of Behavioral Sciences*, *History of Psychology, Theory and Psychology*, and *Journal of Humanistic Psychology*. He also recently published a co-authored book, *A History of Psychology* (2015), through Bridgepoint Education. He has a broad range of research interests, including the social construction of self-measurement, as well as the life and work Sigmund Freud and Friedrich Nietzsche. His most recent work is on re-theorizing memory as based on metonymic relationships between the subject and space and time. Dr. Greer is also the coordinator for the Julian Jaynes Conference on Consciousness, which is held at the University of Prince Edward Island, and has served on the executive of the History and Philosophy of Psychology section of CPA for several years, including a term as section Chair and 3 years as Editor of the section's journal/newsletter, *The History and Philosophy of Psychology Bulletin*. Dr. Greer lives in Argyle Shore, PEI with his wife, daughter, Cocker Spaniel, a sandy beach, and the beautiful sights and sounds of the ocean.

Publication Details

Journal of Cognition and Neuroethics (ISSN: 2166-5087). April, 2016. Volume 4, Issue 1.

Citation

Greer, Scott. 2016. "The Last Modern Psychologist: Julian Jaynes' Search for Consciousness in the Natural World." *Journal of Cognition and Neuroethics* 4 (1): 13–25.

The Last Modern Psychologist: Julian Jaynes' Search for Consciousness in the Natural World

Scott Greer

Abstract

Julian Jaynes, late professor of psychology at Princeton, is best known for his controversial yet provocative book, *The Origin of Consciousness in the Breakdown of the Bicameral Mind*. Based on an unpublished manuscript, and other archival documents, this paper examines his unpublished work, the "History of Comparative Psychology," which represents a failed search for the origin of consciousness as a natural kind. Jaynes abandoned this work to begin the *Origin of Consciousness*, which represented a radical break in theorizing about the emergence and nature of consciousness. In Jaynes' mature theory, the "breakdown" of the non-conscious bicameral mind led to the process of internal narratization, existing through time in what he called a "mindspace." Jaynes' final definition of consciousness was that of "an analog 'I' narratizing in a mindspace," and as "based on metaphor, developed through language, and is an operator, not a thing." A number of profound implications follow from understanding of consciousness as socially constructed. Most dramatically, Jaynes brought the modernist conception of consciousness as a natural kind to a close and provided an alternative explanation; and eschewing centuries of reification, Jaynes he concluded that consciousness does not exist – at least not in the way it is often assumed, as a brain function. Consciousness, as phenomenal experience, can only be said to exist *intersubjectively*, pointed to a moral and ethical dimension that purely naturalistic investigations of consciousness are unable to address.

Keywords

Origin of Consciousness, Bicameral Mind, Julian Jaynes

> O, what a world of unseen visions and heard silences, this insubstantial country of the mind! What ineffable essences, these touchless rememberings and unshowable reveries! And the privacy of it all! A secret theater of speechless monologue and prevenient counsel, an invisible mansion of all moods, musings, and mysteries, an infinite resort of disappointments and discoveries. A whole kingdom where each of us reigns reclusively alone, questioning what we will, commanding what we can. A hidden hermitage where we may study out the troubled book of what we have done and yet may do. An introcosm that is more myself than anything I can find in a mirror. This consciousness that is myself of selves, that is everything, and yet nothing at all... (Jaynes 1976, 1)

Thus spake Jaynes! And thus, some 40 years ago, Julian Jaynes began his ingenious but highly controversial magnum opus, *The Origin of Consciousness in the Breakdown of the Bicameral Mind*. In this work, penned during the heyday of behaviorism, Jaynes offered some provocative new ideas on the nature of consciousness, and, as some of you may know or recall some rather startling conclusions as to its origins.

Three Interrelated Theories

Jaynes' ideas have been out of academic circulation for a while, so before proceeding to the focus for my paper, I would like to review the basic points of his theory, broken down by Jaynes into 3 main areas.

1) The first and certainly most controversial aspect of his theory of consciousness is that of *the bicameral mind*. Jaynes argued that until around 1200 B.C. humans did not have consciousness as we understand it: they were unable to introspect, reminisce, make plans, be deceptive, or engage in any reflexive deliberation. When faced with important or meaningful decisions, they heard voices that they took to be gods, which directed their behaviour. Jaynes (1976) proposed that these "admonitory" and "executive voices" emanated from the right side of the brain and were communicated to the left side, the 'human' side, as an external voice. The right side, being more creative and better at solving more complex and long-term problems, appeared to the person as a voice of authority that understood the world in a larger, more abstract, even god-like way. As you might imagine, a society full of hallucinating people is hardly stable. And as the numbers of people living together and needing to be coordinated grew, more stress was placed on these people and their 'gods.' According to Jaynes, although the bicameral civilization worked under conditions of consensus and strict hierarchy, it was a fragile arrangement; one that had worked for the 'hunter-gather,' but was too inflexible in a context with greater and more numerous and varied social connections. With increasing internal and external stressors (e.g., earthquakes, volcanic eruptions, invasions), this bicameral mentality gave way to what was simply a more efficient use of our brain—namely, consciousness. Most notably, the "breakdown" of this bicameral mind was precipitated, in addition to these other factors, by an evolution in language (including the spread of writing), and the use of metaphor (and symbolic thought) more specifically. Through this evolution of language and language use, the breakdown of the bicameral mind also lead to a further symbolic process of internal narratization. Jaynes (1976) described this as the linguistic assimilation of the voices of the gods into a single sense of self, existing through time in what he called a "mindspace" (more on that later). So, God isn't dead

after all, just silent! It is important to note that it is this linguistic shift rather than any biological or neurological change that resulted in consciousness as we now experience it. According to Jaynes (1976; 1986a), there is no substantial difference between our brains today and those of bicameral people 3,500 years ago.

2) The bicameral mind theory is rooted in neurological differences (primarily with regard to speech) between the right and left hemispheres. One of his key insights into the origin of consciousness came in 1967, when Jaynes realized that if evolution had confined speech areas to the left side of the brain, what was the corresponding right side for, since most important brain functions are bilateral? These differences are still there, he noted, and can be witnessed today in cases of schizophrenia, through electrical stimulation to the right side of the brain, or in certain aspects of childhood, such as in having imaginary friends (Jaynes 1976; Keen 1977). Sperry's split-brain research came out a short time later, and then, Jaynes recalled, "I knew I had something big" (Time 1977, 52).

3) Extending from these first two points, Jaynes argued that the origin of consciousness rests *not* in evolution through natural selection, or some biological adaptation, but consciousness is a product of *culture and language*, of a cultural evolution in the use of writing and language. Our mentality—whether bicameral or conscious—is thus more a function of social context, language, and forms of communication than a hard-wired neurologically-based system. Understanding consciousness, therefore, has more to do with understanding our society rather than our brain, our language practices rather than neurotransmitters, and our cultural history as opposed to our genetic endowment. Of course, biological factors clearly play an important role; for example, the evolution of communication and language in humans is something genetically and biologically grounded, but consciousness itself is something that emerged from that, from the use of language more specifically. Consciousness is thus a kind of social practice. Put another way, while consciousness may be partly *enabled by* the brain (i.e., the brain is a necessary condition), it will not and simply cannot be *found in* the brain (i.e., it is not a sufficient condition).

Jaynes (Keen 1977; Rhodes 1978) stated that his theory of consciousness (#3) does *not necessarily* commit one to his bicameral hypothesis on its origins (#1), or his neurological theory on the structure of the bicameral mind (#2). Jaynes' bicameral mind theory is rooted in neurological assumptions, while his argument for the development of consciousness is not dependent on a bicameral mind (which was not so much a mind as a dual brain). Ontologically, Jaynes argued, consciousness itself is outside the parameters of genetics and natural selection, and is on an entirely different order than the brain.

In telling this story, Jaynes wove together, in an almost polymathic fashion, a narrative that draws on aspects of philosophy, psychology, history, neurology, anthropology, archaeology, religion, and linguistics. Jaynes' search for the origin–or origin**s**–of consciousness was also a highly personal quest, resulting in a theory that makes a fascinating blend of science, literature, history, and crypto-biography (cf. Kuijsten 2006).

Having reached the status of something of a cult figure today, Jaynes' theory was and still is extremely controversial, and has been the subject of an intense and wide-ranging debate, both inside and outside of academia. The wide reception of his book, and its many reviews, nearly constitute a sub-literature, and vary tremendously from "one of the books of the century" (William Harrington, in *Columbus Dispatch*) to Mike Holderness, in the *New Scientist*, who remarked, "It has been a while since a philosophical book made me laugh out loud."

In any case, it could not be said that Jaynes' ideas have been irrelevant, nor the point of his questioning moot. Jaynes' book and ideas came at a time when psychology was rather loath to discuss the topic of consciousness, which had been essentially *Verboten* since the days of Watson. However, Jaynes predicted that consciousness would return to Psychological discourse, and he of course was correct. Since the 1980s, developments in computer science, cognitive psychology, and neuroscience – supported by the development of technologies such as PET and MRI scans, have brought consciousness to the fore once again – although it is hardly the same sense of "consciousness" the field had once known in the theories of William James and Edward Titchener.

So, what does Jaynes believe consciousness is?

The Search for and Development of a Theory of Consciousness

Jaynes traced the start of his quest for understanding consciousness to a vivid memory from the early age of six: while raking leaves in his yard, he was suddenly struck by the idea that the 'yellow' he saw in the forsythia bush before him may not be the same 'yellow' that others see; and, moreover, how would one ever know what someone else saw? Jaynes recalled, "As a child, I was fascinated by the inner world I alone could see, and I wondered what was the difference between seeing inwardly and outwardly" (Rhodes 1978, 62).

Jaynes started his search in earnest as an undergraduate majoring in philosophy and literature, attending the University of Virginia his first year, Harvard his second, and McGill his third and fourth. Jaynes graduated from McGill in 1941. He had studied philosophy

with the hope that he might understand this "interior space we call consciousness," but later considered this a "false start": "...after going through Kant's Critique of Pure Reason and various epistemologies, I felt that we had to be out in the world gathering data to get anywhere" (Keen 1977, 60). With this in mind, Jaynes continued with graduate work at Yale's Institute of Human Relations in 1946, studying human physiology and animal behaviour. Jaynes looked for his 'data' by examining the relationship between the brain and behaviour, looking for the physiological and biological bases of the mind. His research was clearly connected to the theory of evolution, and, from that, the idea that consciousness must have evolved–its origins should be traceable back through history and through our links with other animals. Jaynes then began a systematic search for consciousness by studying how animals learned. He started with plants and moved on to single-celled organisms, neither of which appeared to learn. Jaynes recalled, "I began running paramecia and protozoa through little T-mazes, all in the blissfully absurd notion that I was researching consciousness" (Keen 1977, 60). He then studied simple animals, such as flatworms, and then on to fish, reptiles, and cats, which obviously could learn--but Jaynes was feeling restless, and he was beginning to wonder if he was in any way coming closer to finding consciousness.

The Turn Toward Culture and Language

During his work in animal behaviour studies and ethology in the 1960s, Jaynes had begun to compile some journal publications with his mentor Frank Beach. He had also started composing what would have been a book-length manuscript tentatively titled, "The History of Comparative Psychology." I have been fortunate enough to obtain this 100+ page manuscript as part of a larger Jaynes archive (at UPEI). Here, Jaynes presented an historical review of the study of animals, and what it told us about our close evolutionary relationship with the rest of the animal kingdom. We can see how this work, along with his laboratory investigations and ethological studies, were a precursor to, and was actually abandoned for, his *Origins of Consciousness* in 1976.

So, frustrated by his own failure after many years to uncover even a glint of consciousness, Jaynes determined by the end of the 1960s that the search for consciousness as a natural kind (object) and a product of evolution was a "dead end" (Jaynes 1986). He slowly began to realize, "...the problem of consciousness had stumped so many people because it wasn't in evolution, it was in human culture" (Hilts 1981, 87).

Jaynes (1986) elsewhere elaborated:

> This error, I think, comes from John Locke and empiricism: the mind is a space where we have free ideas somehow floating around and that is consciousness. And when we perceive things in contiguity or contrast or some of the other so-called laws of association, their corresponding ideas stick together. Therefore, if you can show learning in an animal, you are showing the association of ideas, which means consciousness. This is muddy thinking. (129)

However, Jaynes realized some progress had been made. Through his experimental research with animals, Jaynes had systematically and deductively come to understand what consciousness was not: it was not all of mentality or perception, it does not copy experience, nor is it necessary for learning (in a complete reversal of his initial assumption) – in fact, consciousness can interfere with learning (sometimes called "self-consciousness")—and it is not even necessary for thinking or reasoning, which the Wurzberg School demonstrated over 100 years ago.

So, Jaynes began a new line research, with a new bold set of assumptions. He looked for evidence of consciousness throughout world history and culture; searching ancient literature and art, and any kind of archeological evidence that might indicate the presence or absence of consciousness. The most direct kind of evidence seemed to be from language, and so he looked for concepts or actions that would denote consciousness. We can see, for instance, that Plato and Aristotle were conscious, although they do not have a well-defined concept of consciousness per se. He continued on through the Homeric Greeks, tracing consciousness back, back until it disappeared between *The Illiad* and *The Odyssey*. This would then place the origin of consciousness for the Greeks between 1200 and 1000 B.C. For Jaynes, these two works seemed to bracket the emergence of conscious-type thinking (*or at least concepts tantamount to consciousness*). In *The Illiad,* the Greeks and the Trojans are depicted, more or less, as "puppets" of the gods, who are much more salient in determining the course of human action than in *The Odyssey.* In this work, crafty Odysseus is capable of, for instance, acts of deception, something that requires consciousness (Jaynes 1976).

Jaynes found similar evidence of consciousness emerging in the writings of the near and Middle East, such as in the Bible and the Upanishads, and there was a remarkable degree of consistency around the dates, all centring around 1200 to 1000 B.C. (Jaynes 1976; 1986). Again, Jaynes argued that until this time humans were basically "zombies;" they able to talk, reason, solve problems – all of the same things we do without drawing on consciousness per se. While this may sound patently absurd, Jaynes argued that we

have lived and evolved for millennia without consciousness, and so it would not have been necessary for many basic human behaviours. His earlier research on animal behaviour and the history of comparative psychology had demonstrated this. It only seems essential now, since consciousness is the awareness of our actions. Perhaps the behaviorists were right, but for the wrong reasons. Jaynes pointed out how often consciousness is simply the awareness of what we have done or said, reflected back to us – it is actually not the all-encompassing causal factor we often assume it to be.

However, what I believe is the most decisive, and perhaps radical, point in Jaynes' theory is that he also argued that consciousness is not simply the brain–in fact, it "does not have a location," and elsewhere he stated that "the location of consciousness is arbitrary" (Jaynes 1976; 1986; Harvard tape). I believe many people when asked to point to their 'mind' or to their consciousness would point to their head. Or one might argue, as current models of cognitive neuroscience and cognitive psychology suggest, the brain is really the mind, or at least a properly functioning brain is, among other things, a conscious mind. According to these perspectives, consciousness can be located in the workings of the brain, and a scientific understanding of consciousness involves understanding its underlying neural connections, processes, and structures.

For Jaynes, this is a very common and most unfortunate mistake in that it reifies consciousness into a thing, and misses the essential aspect of its origin: *consciousness developed through the process of generating and fitting metaphors to objects and events* (Jaynes 1986; New Hampshire tape). Consciousness is thus a kind of mental activity, socially and biologically enabled. The 'space' of consciousness is not a physical space; it is what Jaynes called a "mindspace," which he defined as a functional space that exists in the same way as mathematics. We would not argue (I hope) that 2+2=4 can best be understood as something residing in our brain; naturally, the ability to use this information involves the brain, but mathematics itself does not somehow reside in the brain. Similarly, consciousness clearly involves the brain, and like mathematical formula, grammatical structures (i.e., our syntax and semantics) are the tools of conscious thought. However, this can in no way be taken to mean that consciousness itself is rooted in the brain (just as most certainly mathematics is not either).

A further example given by Jaynes (1986) is that of riding a bicycle: we all use our brains in riding a bike, but we do not ride bicycles in our head, nor would anyone consider the location of 'bicycle riding' to be in our heads. Consciousness is a thus functional concept that is expressed and 'found' in our use of language and metaphor. Although Jaynes was not opposed to the metaphor of cyberspace to characterize consciousness as a functioning representational system, he was wary of taking computer or technological

analogies too far, calling them "unnecessary," "inaccurate," and is a path that "can lead us astray" (Jaynes, anon. interview transcript.).

With this in mind (so to speak), Jaynes offered two general but slightly different definitions of consciousness. The first comes from *The Origins of Consciousness*, and defines it as "an analogue 'I' narratizing in a mindspace" (Jaynes 1986; New Hampshire tape). We have also heard Jaynes' contention that consciousness developed through the process of generating and applying metaphors to objects and events, and that this occurs in what he called a "mindspace." This can be expressed in three interlocking points:

> 1) The operations of consciousness are based on metaphors, often visual in nature: e.g., "she's very bright."
>
> 2) The relationship to these metaphors is based on a sense of "I"; this I exists and moves about in mindspace, where it can engage in any number of activities, actually possible or not.
>
> 3) This activity occurs in time and is put into a temporal sequence which Jaynes (1976) called "narratization." The modes of conscious narratization can be verbal, perceptual, bodily, or musical.

Jaynes' other, later definition of consciousness repeats the main features of the first, but is describing the *origins* of consciousness rather than its structural features: consciousness is "based on metaphor, developed through language, and is an operator, not a thing" (Jaynes, Harvard tape). As noted earlier, both definitions highlight Jaynes' belief that consciousness and its origins are tied to language and cultural practices, and consciousness is not, in itself, a biological system.

Jaynes (APA talk 1969; 1976) believed that the origin and spread of consciousness was much too recent and much too fast to be accounted for by the (usually) quite slow process of evolution by natural selection. Jaynes (1976) used the example of children's 'imaginary friends,' where we see a vestige of bicamerality before consciousness has fully emerged. As the child becomes socialized to not only the meanings of language, but its metaphorical and representational features, the child learns what it means to be conscious. In essence, we build up a metaphorical "analogy" of the real world through the acquisition of language and the enculturalization of meaning. Once we learn this lexicon of metaphors, the analogue I is able to move about in this mindspace, which is a representation of the external world, and make decisions and choices.

Conclusions and Consequences

In closing, I would like to pose some conclusions about Jaynes' ideas, and some consequences for the investigation of the type of consciousness he described.

First and foremost, Jaynes represents an endpoint of the modernist (ie., Cartesian) conception of consciousness (and mind in general) as a metaphysical thing contained in the body, and as the homunculus that causes conduct. He is also one of the last grand theorizers of mind that attempts to embrace the big picture questions about mind, conduct, and human nature and how it all fits together – itself a hallmark of modernist thought. Research on consciousness for the past 30 years or so, since the advent of imaging technologies, has been much more specific and empirically driven; one might even say pragmatic. Furthermore, Jaynes' connection of consciousness to language and socio-cultural praxis constitutes a clear foreshadowing of many social constructionist (i.e., postmodern) arguments on this point (e.g., Gergen 1985). By transforming questions about consciousness from something to be investigated in the laboratory to the unfolding of consciousness in the realm of cultural evolution and social praxis, he portrayed consciousness as a kind of social construction (as opposed to a natural kind). This, of course, also raises the related question about whether the lab, and our current reductionist neuroscientific discourse, is the most suitable context for revealing all that consciousness entails. Jaynes came from an experimental tradition where consciousness was in the head, a brain even, and only after a long and exhaustive systematic search for consciousness under these assumptions was the naturalistic discourse on consciousness questioned. Jaynes not only questioned these very basic assumptions about the status of consciousness, but he also eschewed the reification of the mind (or soul) to explain consciousness, thus rejecting the ontological dualist tradition and the subject/object dichotomy that have plagued modern, naturalistic accounts of mind and conduct.

The title "last modern psychologist" is thus obviously more symbolic than literal, since consciousness research within a neuroscientific discourse is clearly modernist. Rather, it is Jaynes' failed search for consciousness as part of the natural world, his realization of boundaries and limitations, and the type of alternative explanations he offered in response, that leads me to think of Jaynes as "the last modern psychologist;" or perhaps, he is "the first" last modern psychologist.

Secondly, that said, the modernist-inspired experimental construction and pursuit of consciousness through laboratory methods have clearly flourished since Jaynes. However, I wonder if the tremendous interest in a neuroscientific understanding consciousness is really about asking the "big questions"? Most reductionist explanations do not seem capable of solving the so-called "hard problem" of consciousness, and similar questions

of qualia and embodiment. I wonder if the vast majority of this recent research, and the tremendous amount of funding it has brought, has more to do with how to create a technology that can interface with a particular understanding or problem of consciousness. There are hundreds, maybe thousands, of experimental studies of consciousness and reams of data, but whither theory?? As noted above, gone are the grand theorizers, most researchers are framing their questions in terms of practical considerations concerning publication and effective application (e.g., methodological feasibility, ethical requirements and restraints, budgetary limitations, etc.). I further wonder if, in pursuing this thoroughly modern mind, we are not in danger "of creating a technology, not a science" (to paraphrase Titchener's comment to Watson about his Behaviorism). Again, as the last modern psychologist, Jaynes represented that tradition of theorizing questions of human nature and value. Now, like the concept of the "individual" and the "self," consciousness has become a commodity among researchers and industry. Neuroscience research means lucrative research grants, where the methods of investigation drive the questions, and the applications to the consumer steer the funding. It is no wonder that in capitalist liberal democratic cultures, such a technologically advanced (yet theoretically impoverished and passively mechanistic) construction of mind predominates.

Third, consciousness (qua phenomenal experience) in Jaynes' view can only be said to exist *intersubjectively* (i.e., within a community, or perhaps a field, of language use and meaning). Take 'gender' as a parallel example: its existence is predicated on the simple fact that we began talking about it. It arose in the 19th century out of an evolution in our discourse about sex, sexuality, and other ideas and questions about the development of our private interior. It is of course possible to talk about the biological enabling factors of gender, but, like consciousness and other intersubjective notions, their meanings only come fully into view when seen actualized in a particular context. Ideas such as gender or consciousness soon have little, or a vastly truncated, meaning in a reductionist discourse. When the reification of psychological concepts and metaphors is taken as operationalization, then psychological theory, to quote Jaynes' on this point, becomes "bad poetry masquerading as science."

Last, and perhaps most significantly in the long run, if consciousness is understood to be generated (in part) by language and metaphor, operating as a socially constructed phenomena, then there are obviously moral and ethical dimensions to explore. However, these dimensions are invisible (or rather neutered) in the asocial, purely experimental formulation of consciousness. Jaynes was well aware of this. In the recordings of some of his later lectures, Jaynes discussed how one of the "consequences of consciousness" was an

increased sense of interdependence among conscious individuals, now that the gods had died. Without the voices in their heads, that a priori, authoritarian voice giving people a sense of good and evil, right and wrong, morality became much more complex—now truly a matter of human construction and deliberation. In this light, consciousness might be argued to be something that is just as fundamentally ethical as it is neural or social in nature.

References

Gergen, K. 1985. "The social constructionist movement in modern psychology." *American Psychologist* 40: 266–275.

Hilts, P. 1981. "Odd man out." *Omni* (January): 68–88.

Jaynes, J. (Undated). "The history of comparative psychology." Unpublished manuscript.

Jaynes. J. (Interview, undated). "Unedited and abbreviated Julian Jaynes interview." New Media Associates.

Jaynes, J. 1969. "American Psychological Association conference audio tape." *UPEI Julian Jaynes Collection*.

Jaynes, J. 1976. *The origin of consciousness in the breakdown of the bicameral mind*. Toronto: Houghton Mifflin.

Jaynes, J. 1983. *University of New Hampshire* audio tape (4-28-83). *UPEI Julian Jaynes Collection*.

Jaynes, J. 1986. "Consciousness and the voices of the mind." *Canadian Psychology* 27: 128–139.

Jaynes, J. 1986a. "How old is consciousness?" In *Exploring the concept of mind*, edited by R. Caplan. Iowa City: University of Iowa Press.

Jaynes, J. 1988. "Harvard audio tape (12-3-88)." *UPEI Julian Jaynes Collection*.

Keen, S. 1977. "Julian Jaynes: Portrait of the psychologist as maverick theorizer." *Psychology Today* 11: 66–77.

Rhodes, R. 1978. "Alone in the country of the mind: When did humans begin thinking?" *Quest* (January/February): 71–78.

Time. 1977. "The lost voices of the gods." *Time Magazine* (March 14): 51–53.

Journal of Cognition and Neuroethics

The Epoche and the The Intentional Stance

David Haack
The New School

Biography

An element that ties all of my work together is a deep pluralistic impulse. Before beginning the study of philosophy at the New School for Social Research, the tradition in philosophy I was most aligned with was pragmatism. I am currently completing my master's thesis in philosophy of mind and phenomenology. In addition to my interest in epistemology and ontology, I have a secondary interest in ethical theories from antiquity. I find myself both interested in modern philosophy as a science (its intersection with the cognitive sciences) and antiquity's interest in ethics as how to live a life. In addition to my interest in ethics within antiquity, I have a specific affection for John Dewey's theories of democracy and education.

Publication Details

Journal of Cognition and Neuroethics (ISSN: 2166-5087). April, 2016. Volume 4, Issue 1.

Citation

Haack, David. 2016. "The Epoche and the The Intentional Stance." *Journal of Cognition and Neuroethics* 4 (1): 27–44.

The Epoche and the The Intentional Stance

David Haack

Abstract
In this paper I explore the issue of intentionality by looking at the thought of Daniel Dennett and Edmund Husserl. I argue that despite the differences between Dennett's 'hetero-phenomenology' and Husserl's phenomenology, the two ways of viewing intentional content, and therefore consciousness, more broadly are not incompatible. I claim that we can view consciousness in a way that incorporates both the phenomenological and hetero-phenomenological methods. I begin by outlining Husserl's phenomenology before moving on to a description of Dennett's hetero-phenomenology. Next, I bring the difference in their thought into sharper contrast by exploring a criticism made by Hubert Dreyfus and Sean D. Kelly who put forward the claim that Dennett's hetero-phenomenology over-generates belief content and under-generates intentional content. I argue that this is an unfair criticism because Dennett's goal is to give a simple description of conscious states. Dennett is following Occam's razor exclusively in order to make claims about consciousness that can be backed up by the kind of verification characteristic of the natural sciences. For Dennett, under-generating intentional content is a strength. Conversely, I point out that there are many descriptions of intentional states that Husserl can account for which Dennett cannot. Lastly, in a final section I explore what a combination of the phenomenological method might look like if intertwined with a hetero-phenomenological method.

Since Dreyfus and Kelly's critique centers around belief attribution, I explore the following question: is intentionality possible without holding a belief about the intentional object? Dreyfus and Kelly claim that we can be intentional towards something without an 'I believe' being attached to what we are intentional towards. Related to this is how much of what has been considered consciousness by phenomenlogists really comprises consciousness. Husserl sketches out a fuller consciousness than Dennett, and one that is achieved as an object of study through the epoche or phenomenological bracketing. The epoche shifts the view to 'pure consciousness' and away from the natural world. Dennett's hetero-phenomenology tries to achieve a study of consciousness through a third-person study of a subject's rational belief. For this reason naturalism, when it comes to the study of consciousness, is also a subject of investigation within my paper. I maintain in the final section that we can move back and forth between attempting a naturalistic view and conversely performing the epoche and exploring the wider territory this makes available to us.

Keywords
Intentional Stance, Hetero-Phenomenology, Phenomenology, Eidetic, Cogito, Ego Splitting, Phenomenological-Hetero-phenomenological Harmony

In this paper I take a look at the thought of Edmund Husserl; particular attention is paid to his work *Ideas: For a Pure Phenomenology*. Compared to this work is Daniel Dennett's "True Believers the Intentional Strategy and why it Works" as well as his essay "Whose on First: Hetero Phenomenology Explained." To engage with these two different ways of viewing consciousness, I will take issue with an argument made by Hubert

Dreyfus and Sean D Kelly in their 2007 paper entitled "Hetero-phenomenology: Heavy Handed Sleight-of-Hand." In this paper, Dreyfus and Kelly argue that Dennett's Hetero-phenomenology over-generates beliefs and under-generates intentional phenomena. By this, they mean Dennett does not take note of how consciousness can be directed without having a belief attached to this direction. To account for this, only a phenomenological, and not a hetero-phenomenological, set of concepts will do. I maintain, however, that while there are certain ways to understand consciousness that only phenomenological views can account for, there are other reasons a hetero-phenomenological view can be helpful.

To present this, I will draw from Husserl's *Ideas*. I will show what hetero-phenomenology as a system has no way of accounting for. Many of these notions are discussed at the very founding of phenomenology. These include ways of being intentional towards an object without having a belief about that object and the outer rim of a perception that we are focused on and its role in the way we interpret what we are focused on. In addition to this the method does not have direct accesses to different layers of reflection on reflections and/or fantasies of fantasies or memories of memories, The way fantasy plays a role in interpretation of an object and for intentionalities within these multi-layers of perception (higher and lower).

I will argue that just because Husserl can cover ground that explains parts of consciousness Dennett cannot, this does not mean that Dreyfus is correct about Dennett. It means instead that Dennett from a solely naturalistic perspective has found a way to have a natural science of certain aspects of consciousness. I will insist that the argument that he overpopulates and under-populates the conscious realm is unfair, showing instead that Dennett's view is a helpful tool in understanding the intentional content of human consciousness. My essay will consist of a defense of Husserl and phenomenology and a description of what only phenomenology tells us about consciousness, as well as how Dennett cannot explain these insights, and conversely a defense of Dennett. The last section then will be a look at how these two views of consciousness can live together in harmony, one hand washing the other.

I will begin with a description of the epoche and the phenomenological reduction, flow of consciousness and a very general view of intentionality before a description of the thought of Daniel Dennett. Following this Dreyfus and Kelly's argument will be laid out, and then a defense of Dennett against this argument, then a return to Husserl and what he gives specifically in contrast to Dennett, before a final section of phenomenological and hetero-phenomenological harmony.

If we are to follow Edmund Husserl into the place that led him to what would become phenomenology, we must first agree with him that eidetic universals while not having an address in space and time, do have a truth to them. By eidetic Husserl means something's generality or as he often writes, it's eidos. He uses the terms eidos or how it is eidetic to separate his thought from Kant or as he remarks, "The need to keep the supremely important Kantian concept of the idea purely separate from the general concept of the (formal or material) essence also moved me to alter the terminology. Thus I employ, as a foreign word, the terminologically little used eidos and, as a German word, essence" (Husserl 2014, 7). If we are going to take phenomenology in Husserl's sense seriously we must let ourselves believe that there can be a science that deals with the generality of objects. As he tells us "not, as a science of facts, but instead as a science of essences (as an 'eidetic science'), a science that aims exclusively at securing 'knowledge of essences' and no 'facts' at all"(Husserl 2014, 5). How does Husserl achieve this eidetic science of essences? With the epoche or phenomenological bracketing, this move on Husserl's part is along with intentionality the most central feature of phenomenology and what allows for a separate study of consciousness.

The goal of the epoche is to get to consciousness as such or the term Husserl prefers, 'pure consciousness.' To do this Husserl wants to bracket a certain view of the natural world. Therefore, consciousness embedded in the reduction focuses on consciousness only in its "sui generis" way of being. He writes that this is the "insight that consciousness in itself has a being of its own that is not affected in its own absolute essence by the phenomenological suspension. It accordingly remains as a 'phenomenological residuum,'"(Husserl 2014, 58). Husserl is then out to study consciousness in its uniqueness, but this is not a consciousness that is separate, though a strictly empirical world view is bracketed, it is not one cut off from the world and its objecthood, it in fact takes the world and the objects in it as its point of departure. For this reason Husserl will need to describe for us the unique way we encounter objects as our perceptions. We get a good example of this on page 60 of *Ideas* with a description of the object of a paper that is under a dim lighting, "This seeing and touching of the paper in perception, as a complete concrete experience of the paper lying here, and to be sure, of the paper given exactly with these qualities, appearing precisely with this relative lack of clarity, in this imperfect determinacy, in this orientation to me-is a cogito, an experience of consciousness" (Husserl 2014, 60). Here we are given a description of a paper but not purely as an empirical object, its chemical makeup does not change depending on the lighting, but a paper as it is conceived by our consciousness depends on the lighting. This paper is conceived specifically the way we conceive a paper in dim lighting. We also notice

here that just because the epoche has taken place it does not mean that phenomenology is not based on objects in the world, it in fact starts with these objects. This is Husserl's insight into the study of consciousness; we take the objects of perception, bracket the naturalistic conception of the world and what then comes into view is the phenomenal realm of the objects we are intentional towards.

It is important to think of this bracketing not as a one-time move for Husserl it has multiple layers. Husserl is a thinker of the layer, consciousness once this bracketing takes place comes into view as something like a very large cake. So bracketing happens throughout *Ideas* multiple times as he writes on page 58 of this work, "This operation will break down methodologically into various steps of 'suspension,' 'bracketing,' and so our method will assume the character of a step-by-step reduction"(Husserl 2014, 58).

Once a naturalistic view is bracketed, when we go to look at consciousness we do not see something that is still. Consciousness for Husserl is in a flow, what does he mean by this? How is consciousness a flow? We get a very clear description of this in "Philosophy as Rigorous Science," he writes that the psychical "appears as itself through itself, as an absolute flow, as a now and already 'fading away,' clearly recognizable as constantly sinking back into 'having been.'" (Husserl 1965, 43). Husserl goes on to describe that experience can be recalled in recollection and we can experience those recollections themselves, as well as recollections of those recollections and on and on. This is very important to Husserl, he writes in fact "In this connection, and this alone, can the a prior psychical, in so far as it is identical to such "repetitions" be "experienced" and identified as being" (Husserl 1965, 43). He goes on to write this creates the unity of consciousness that exists within the epoche outside of the naturalistic world of space and time, he calls this a "monadic unity of consciousness" (Husserl 1965, 43). Once a naturalistic view is bracketed we see what goes on within consciousness, what goes on within consciousness is a world of flowing connections, a stream of consciousness. Husserl continues describing this flow even more vividly "Looking back over the flow of phenomena in an imminent view, we go from phenomena to phenomena (each a unity grasped in the flow and even in the flowing) and never to anything but phenomena" (Husserl 1965, 43). An imminent view for Husserl is a view that sees our purified consciousness, as opposed to a naturalistic view. He refers to our phenomenal content, as unities because our perception of phenomena consists of many layers, these many layers are what are grasped and then become part of this flow, finally he reminds us that nothing enters into this flow but phenomena. It is a plausible view of what consciousness would look like once a naturalistic conception is bracketed. Once experiences of the phenomena enter

into a flow in their multi-faceted unity, a unity that does not consist of the paper in a naturalistic sense but the way we experienced it in the dim light, for example, this flows with other monadic unities one into the other fading in and fading out, backgrounds of one impression flowing with foregrounds of the next (as one possible kind of flow) and on and on.

Though Husserl studies consciousness in its uniqueness, it is not cut off from the world and its objecthood. We relate to the world in its objecthood by being intentional towards the objects. Intentionality is fundamental for Husserl because it is what relates a being to another being. Consciousness therefore is intentional; it is a being of encounter. Whenever we think we are thinking about something, Husserl tells us that this is without exception, or as he puts it "each currently actual cogito is to be consciousness of something" (Husserl 2014, 62). This has the added connotation that what we are currently conscious of is a direction towards something. It will be important for the argument later that this directedness is not necessarily a belief about what we are directed towards. Once we explain Husserl's concept of doxis we will see that this relation between belief and non-belief will become more complex.

How does something such as our intentionality function once we have performed the epoche? Husserl explains this for us in *Ideas* when he writes, "the modified cogitation is equally consciousness, and consciousness of the same thing as the corresponding unmodified consciousness is. Hence the universal essential property of consciousness remains preserved in modification" (Husserl 2014, 62–63). After the epoche we still have within consciousness our directedness towards the world, in fact consciousness to a very high degree is this very directing. Husserl lays out for us later within *Ideas* "It is intentionality that characterizes consciousness in the precise sense of the term and justifies designating the entire stream of experience at the same time as a stream of consciousness and as the unity of one consciousness" (Husserl 2014, 161). Without intentionality, then the amorphous term consciousness would be unclear for Husserl, we get a 'precise' object of study from the introduction of this term. It is what lets experiences as a stream parallel the stream of consciousness and lets this stream of consciousness be a unified stream. Without intentionality we have no way to grasp the thing we call consciousness at all. Intentionality is not only an essential part of consciousness for Husserl, but it also helps distinguish what is specific about experiencing. He states, "The sphere of experiences in general is essentially distinguished by virtue of the fact that they all in one way or another have some share in intentionality, even if we cannot say of every experience in the same sense that it has intentionality as, for example we can say of every experience that comes into focus as an object of possible reflection that it is temporal" (Husserl 2014

161). He elaborates this further, writing that although the most fundamental element of what we experience is that it relates to the world, the sphere of our consciousness of experience is that it is almost always intentional as well.

For Daniel Dennett also, intentionality is a central concept, as he introduced it early on in his writing, most notably in his essay "True Believers The Intentional Strategy and Why it Works." As the title of this essay alludes, belief will be central to the way Dennett will attempt to study consciousness. While for Husserl consciousness is grasped by the epoche, for Dennett third person belief attribution is the way he achieves a study of consciousness. As Dennett writes in a more recent essay, "Who's on First: Heterophenomenology Explained," "You *reserve judgment* about whether the subject's beliefs, as expressed in their communication, are true, or even well-grounded, but then you treat them as *constitutive* of the subject's subjectivity. (As far as I can see, this is the third-person parallel to Husserl's notion of bracketing or epoché..."(Dennett 2003, 22). So for Dennett consciousness will be captured by third-person empirical belief attribution. This requires, however, some preliminary understanding of the sorts of beliefs one has. If we are to accept Dennett's schema we must first believe in a common set of rational principles. While a common set of interests among agents may sound limiting, Dennett gives a compelling argument for it in his "True Believers," that there are in fact some ways agents act that can be predicted quite accurately by belief attribution. It is difficult to argue against certain features of our behavior based on our physical composition or on our design. If someone leaves food we enjoy in a room after we have not eaten for seven days we will most likely eat that food, we could make a further prediction that if it is food we are not particularly fond of we are more likely to eat it in that situation. While this fact is based on our physical constitution as the kinds of things that need to eat to live, we can also understand some of what is going on in our heads at this moment by assuming we believe that eating this food is a good idea. This would be an example of performing the kind of rational belief attribution Dennett wants us to adopt, as he remarks, "first you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose"(Dennett 1987a, 17).

The example I have just proposed is simple but Dennett wants to persuade us that it works for situations that are more complex. Dennett even maintains that for situations that seem rather uncanny, parts of these situations can be broken down into this sort of belief attribution as he writes, "Suppose the US Secretary of State were to announce he was a paid agent of the KGB. What an unparalleled event! How unpredictable its

consequences! Yet in fact, we can predict dozens of not terribly interesting but perfectly salient consequences, and consequences of consequences. The President would confer with the rest of the Cabinet, which would support his decision to relive the Secretary of State of his duties pending the results of various investigations, psychiatric and political, and all this would be reported at news conferences to people who would write stories about it that would be commented on by editors" (Dennett 1987a, 25). This may seem a rather boring view of human behavior, when even the uncanny can be broken up into a bunch of predictable acts. It is not the act Dennett wants to draw our attention to, however, it is that through belief attribution we can perfectly predict the way people would behave in such a situation. Through this third person stance of belief attribution we come into contact with the believer's consciousness, a far more interesting prospect.

In order to show that this stance has the same kind of scientific legitimacy as other studies of the physical world, Dennett shows us how it is one that can be taken after two other stances have been applied, the physical stance, and the design stance. Dennett describes the physical stance when he writes "if you want to predict the behavior of a system, determine its physical constitution (perhaps all the way down to the microphysical level) and the physical nature of the impingements upon it, and use your knowledge of the laws of physics to predict the output for any input" (Dennett 1987a, 16). This is the sort of predictability philosophers have had strong attraction to since Rene Descartes, these are the rules of prediction within physics. Dennett goes on to say that physics itself sometimes falls short of what we want to predict within the physical world. He offers another 'stance,' this one he calls the 'design stance' remarking, "the design stance, where one ignores the actual (possibly messy) details of the physical constitution of any object, and, on the assumption that it has a certain design, predicts that it will behave as it is designed to behave" (Dennett 1987a, 16–17). Dennett gives the example of a computer remarking we do not know how computers run (most of us at least) but we know how to interface with them and can predict much of the way computers will run based on this information. Then if the design stance still cannot predict the behavior of what we are studying there is a chance we are studying something like us, something that makes choices based on rational interests. It is evident from these examples that what Dennett wants from the intentional stance is a theory as firm as the physical stance. He puts it in a succession after these other two stances in order to show there is firm scientific base for adopting this intentional stance. This is Dennett's attempt, almost as novel as Husserl's to reach a scientific theory of consciousness.

We may ask, what about the problem of belief attribution to those who act irrationally, who knowingly do things against their own interest? Dennett has an answer to this although he maintains "the [perverse] claim remains: all there is to being a true believer is being a system whose behavior is reliably predictable via the intentional strategy"(Dennett 1987a, 29). He readily acknowledges, "No one is perfectly rational, perfectly un-forgetful, all-observant, or invulnerable to fatigue, malfunction, or design imperfection. Since this is the case Dennett observes that we need a particular explanation for non-rational behavior. His answer is this: "the attribution of bizaree and detrimental desires thus requires, like the attribution of false beliefs, special stories"(Dennett 1987a, 20) This is Dennett's answer on how to avoid this issue: sure there will be some things human beings do that we cannot predict with belief attribution but it will still be a story shot through with steps up until the one moment we cannot account for with acts we can attribute belief attribution to. Dennett maintains that since "One is not supposed to need an ulterior motive for desiring comfort or pleasure or the prolongation of one's existence," (Dennett 1987a, 20) this kind of attribution is possible. So for Dennett our false beliefs require special stories, and these stories consist for the most part of true beliefs.

In his more recent essay "Who's On First? Hetero-phenomenology Explained," Dennett defends his intentional stance in light of many of the counter arguments that have been introduced in the twenty years between the two essays. This work is a development of his intentional stance, which is the main element of his hetero-phenomenology (meaning phenomenology of another not one's self). Dennett wants to more clearly define why this is desirable by writing that "if you have conscious experiences you don't believe you have –those extra conscious experiences are just as inaccessible to you as to the external observers" (Dennett 2003, 3). He is maintaining here that without beliefs we cannot make sense of the intentional content. Belief-hood then is another requisite for Dennett, in order for something to be considered a naturalistic object of study. In contrast to the inclinations of Husserl in phenomenology, Dennett maintains that "You are not authoritative about what is happening in you, but only what seems to be happening in you"(Dennett 2003, 4).

Recently, Herbert Dreyfus and Sean D. Kelly have engaged with Daniel Dennett. Both Dreyfus and Kelly hold phenomenological views, which were influenced by Husserl and many of the phenomenologists who followed in his footsteps (though none followed Husserl to the extent he would have liked to see). They maintain that Dennett falls prey to a similar dysfunction, although avoiding a phenomenon they see as problematic in later Husserl: Ego splitting. Ego splitting, Husserl in the *Cartesian Meditations,* consists

of "the phenomenological Ego establishing himself as disinterested onlooker, above the naively interested Ego"(qtd. in Dreyfus and Kelly 2007, 34, *Cartesian Meditations*). Dreyfus and Kelly take issue with this notion. If we followed this later Husserl, Dreyfus and Kelly claim, we would distort the experience we were interpreting; it is like "dancing while observing where one is placing one's own feet"(Dreyfus and Kelly 2007, 46). Husserl's later view that we can split our Ego in order to study it, is in this view naïve. They maintain that Dennett avoids this 'transforming through reflection': "The subject studied by the hetero-phenomenologists does not have to reflect in order to report on his experience, so the hetero-phenomenologists can legitimately take utterances of his subjects to be unreflective reports on all and only the content of their experience" (Dreyfus and Kelly 2007, 47). While the Hetero-phenomenologist's third person stance avoids this ego-splitting distortion, it unfortunately falls to an equally destructive distortion of its own.

According to Dreyfus and Kelly, Dennett falls into this trap by attributing an 'I believe' to everything we are intentional towards. How can one be intentional towards something they have no belief about? The notion that we can be intentional toward something we have no belief about was first articulated by Husserl. We get a description of this in *Ideas:* "consciousness in general is so fashioned that it is of twofold type: prototype and shadow, positional consciousness and neutral consciousness. The one is characterized by the fact that its doxic potentiality leads to doxic acts that actually posit something; the other by the fact that it permits only shadow images of such acts" (Husserl 2014, 225). Belief is quite important for Husserl as well as Dennett, it is intentionality's most basic form, Husserl refers to this as originary doxic. There are cases that are modifications on this originanary doxis which is what the above quote refers to. While some of our intentional experience actually posits something that is doxic or as we might say holds a belief, we can also hold an 'I don't believe.' Others are not as straight forward in holding something clear that would correspond with an 'I believe,' but they are neutral in terms of the belief or disbelief.

In order to describe this notion of being intentional towards something without having a belief about it, while also describing another notion, ego-submersion, absent in Husserl, Dreyfus is fond of Sartre's example of when someone is running towards a street car, he remarks when running towards a street car we are directed towards this object but there is no I, so there can be no 'I believe,' only a directedness towards our intentional object. If it is the case that Dennett cannot account for intentionality without a belief tied to it, if we are to follow Dreyfus and Kelly in this claim, it is a major blow to his conception of consciousness. As we have seen earlier, in order to have Hetero-

phenomenology we need the intentional stance, and the intentional stance relies on belief attribution. This is a deep structural problem with this account of consciousness. As they write, "instead of simply recording the subjects utterance 'getting closer' the hetero-phenomenologist writes down for example 'the subject believes he is getting closer" (Dreyfus and Kelly 2007, 47). Therefore, utterances by a subject get tied to beliefs, these utterances may have no belief tied to them at all. To utter something is not the same as being able to attribute a belief to it, whereas unfortunately Dennett's method would lead someone to the false conclusion that they do. According to Dreyfus and Kelly, "If the hetero-phenomenologist takes his notes to be his data, as Dennett insists, the hetero-phenomenologist is not just conveniently attributing the assertion, 'getting closer,' to the subject; he claims the subject is expressing a believe he actually holds" (Dreyfus and Kelly 2007, 48). This results in the hetero-phenomenologist treating "the subject's reports as if they were the result of reflection"(Dreyfus and Kelly 2007, 48). Dennett's system then is one always putting too many "I believes" in our conscious experience. In thinking he could overcome the singular aspect of phenomenology, he instead ended up with an inaccurate picture of consciousness, one over-crowded with beliefs in places where there are none.

In addition to over-generating beliefs, Dreyfus and Kelly's other claim is that he under-generates intentional content. The reason for this is part and parcel with the reason Dennett is accused of over-generating beliefs. If his hetero-phenomenology can only account for beliefs, it may potentially both over-generate how much of our intentional content is beliefs, while at the same time not accounting for content that is outside of this framework of belief. One of the mental phenomena this way of viewing things cannot account for, according to Dreyfus and Kelly, is the way we experience products in the environment. As they write, "insofar as the hetero-phenomenologist fails to capture the subject's way of experiencing objects or properties, he excludes a vast array of intentional content" (Dreyfus and Kelly 2007, 50). They further clarify by explaining that there are some experiences of objects and proprieties that are not identical with beliefs one has about experiencing these objects and proprieties. To believe we are having an experience does not suffice to explain experiencing as such. As they write, "if the subject were to believe he was having the experience instead of merely having it, the intentional content of the experience would be different" (Dreyfus and Kelly 2007, 51). According to Dreyfus and Kelly Dennett does not allow the for the existence of qualitative experience, or what is sometimes called qualia they write, he argues "against qualia on the grounds that they are completely inaccessible to us except through the beliefs we have about them" (Dreyfus and Kelly 2007, 51). This puts beliefs and not experiences to the forefront of

Dennett's theory of mind. Dreyfus and Kelly hammer in this point further by returning to the theme they brought out with the streetcar example, "Affordances draw activity out of us only in those circumstances in which we are not paying attention to the activity they solicit. As we have seen already, this is how they are not like beliefs" (Dreyfus and Kelly 2007, 52). In the account they hold that objects and activities have normative qualities that cause us to react to them and have intention towards them without holding belief about them, the example they use is a large painting making us step backward. The fact that Dennett over-generates beliefs then is linked to the fact he under-generates intentional experience. Dreyfus and Kelly make this explicit: "If this account of the normative aspects of phenomenology is right, then we have isolated a whole range of intentional phenomena that the hetero-phenomenologist method in principal excludes" (Dreyfus and Kelly, 54).

Dreyfus and Kelly's argument does not meet Dennett on his own ground. While it is true that from a certain perspective he does indeed over-generate beliefs and under-generate intentional phenomena, this claim ignores what he is trying to achieve. Dennett wants to find a method for describing phenomenal content that stays within a naturalistic framework. To say he under-generates intentional phenomena is just to translate what he sees as a virtue into a weakness. Dennett wants to find a way to naturalistically verify intentional content; if he can only capture certain intentional content through his method, this for Dennett is the best we can do at this particular junction in history. His method is set up along the lines of Occam's Razor, he is looking for simplistic intentional content because he believes it will yield the most verifiable results. The claim that Dennett over-generates beliefs is then a more serious claim than that he under-generates intentional content. For from his point of view if his hetero-phenomenological method captures less intentional content, there is a problem with the content that is not captured, not with his method.

We can see an example of this in Dennett's "*Setting Off on the Right Foot*," which introduces his collection of essays *The Intentional Stance*. Dennett introduces criticism made by Thomas Nagel, who he uses as his anti-realist foil in order to show his perspective on topics such as qualia. Dennett quotes Nagel as writing "the attempt to give a complete account of the world in objective terms detached from these perspectives inevitably leads to false reductions or to outright denial that certain patently real phenomena exist at all" (qtd. in Dennett 1987b, 5, *Nagel The View from Nowhere*) Dennett's reply to this quote of Nagel's could double as an answer for Dreyfus and Kelly: "My intuitions about what 'cannot be adequately understood' and what is 'patently real' do not match Nagel's. Our tastes are very different. Nagel, for instance, is oppressed by the desire to

develop an evolutionary explanation of the human intellect (78–82); I am exhilarated by the prospect. My sense that philosophy is allied with, and indeed continuous with, the physical sciences grounds both my modesty about philosophical method and my optimism about philosophical progress" (Dennett 1987b, 5). Dennett does not wince at Nagel's critique. He does not deny he is a reductivist. He maintains, however, contrary to Nagel's 'taste' this is a strength.

Dennett even entertains the idea that the "orthodoxy of his scientific starting point might [even] be due to social and political factors" (Dennett 1987b, 7), although for the most part brushing this aside. He next mentions a critique by Nagel which states that Dennett is quick to try and ground phenomena that require more openness, or 'confusion' than Dennett is willing to give them. Dennett responds with yet another statement that can help us put Dreyfus and Kelly's engagement with him into greater perspective. Dennett states, "My tactical hunch, however, is that even if this is so, the best way to come to understand the situation is by starting here and letting whatever revolutions are in the offing foment from within. I propose to see, then, just what the mind looks like from the third-person, materialistic perspective of contemporary science" (Dennett 1987b, 7). Dennett then readily admits there may be phenomena his method cannot account for, he just has a 'hunch' or is favorable towards the kinds of results that will come out of his third person method. If there are intentional phenomena that cannot be grounded in this way it does not bare on his method much at all, nor do the motives behind adopting the method effect his choice to adopt it. It is only that the results come from this third person method which he attributes as being a part of the empirical sciences. In Dennett's thoughts within *Hetero-Phenomenology Explained,* he writes if we have conscious experiences we do not believe we have, they are inaccessible to ourselves as well as others. We have seen that the under-generation of intentional content is something that Dennett's system produces by design and so attacking his hetero-phenomenology because of this, though pointing towards the existence of a different way to view consciousness, critiques his system from within different rules of engagement. It is not that Dennett's system does not account for this phenomena, it is that he is unwilling.

What about the claim that Dennett over-generates beliefs? This becomes a bit less of a problem if we understand that the two claims that he both under-generates intentional phenomena and over-generates beliefs are linked. It is not that Dreyfus and Kelly are claiming he creates new intentional phenomena that do not exist. They are rather claiming that he puts a 'the subject believes' in front of actually existing intentional phenomena that are not necessarily beliefs. This claim, unlike that he under-generates

intentional content cannot be explained away. However, we can just present Dennett's point mentioned above, that these are in fact beliefs. So these under-generated bits of content are all 'I believes' for Dennett, but within his hetero-phenomenological view they make sense. If Dennett was more open to qualitative experience, he could avoid this later problem while still keeping the framework of his method intact.

Thus far we have been introduced to Edmund Husserl and Daniel Dennett's philosophy and to critics of Dennett's thought and responded to them in defense of Dennett. On the converse of this defense, there is a very large ground of intentional phenomena Dennett does not account for. Although Dreyfus and Kelly do not attack Dennett on his own grounds, such an attack helps us to see the ground he does not cover. Another way of saying this would be to remark they are correct in saying there is much intentional phenomena he cannot account for whether or not his system is designed specifically to only account for certain intentional phenomena. We can argue that it is worth keeping this vast array of phenomena skipped over by Dennett, while also maintaining some results captured Hetero-phenomenologically give us a different kind of verifiability. To see all of the phenomena Dennett cannot account for we return to Edmund Husserl's ideas. We began by showing certain aspects of Husserl's thought, the most fundamental of which include the epoche or phenomenological reduction, the flow of consciousness as well as intentionality. We will now view in more depth some of Husserl's other conceptions, with a new eye towards the fact that a hetero-phenomenologist method cannot account for them.

A purely empirical method of understanding an object will never be able to directly account for the role fantasy plays in our interpretation of an object. For instance if you see the front of a coffee cup I am holding, you can already have in your head an image of what the back of the coffee cup looks like, even if you have never seen the back of this particular coffee cup. In this way our perceptual history plays a role in interpenetrating of the object. What is more you may have a specific emotional connection to the brand of coffee I am drinking, this affection effects the way your consciousness is intentional towards the cup. In *Ideas* we get the more Husserlin example of geometry, "the geometer operates incomparably more in fantasy than in perception of a figure of model" (Husserl 2014, 121). Husserl wants us to understand that mathematics deal in ideals, for example a perfect circle, which cannot be comprehended without the use of fantasy. He goes on further to tell us, "Even where something is "contemplated" by looking at the figure, the newly initiated processes of thinking have as their sensory underpinning the processes of fantasy, the results of which secure the new lines on the figure" (Husserl 2014, 126). If we are to engage with the idea of a shape, even if we are basing it on a shape in the world,

we have our fantasy of what the figure looks like as we draw the new line onto the figure. What one goes about when engaged in geometric thinking cannot be a way of thinking that refers only to empirical examples; this is also the case for phenomenology. Imagination then and our history with objects in the world color our relationship with new objects. As Husserl remarks, "Extraordinary profit is to be drawn from the offering of history and, in even richer measure, from what art and, in particular, literature have to offer" (Husserl 2014, 127). History and literature help to bring up imaginative functions towards the world around us and therefore add dimension to objects that otherwise seem only explicable empirically. There is no way for a third-person perspective to account for the way fantasy could effect intentional content in a manner where we can isolate how they are doing the effecting.

If we look at an object we are not just seeing the object we are intentional towards, but instead we are picking up other things in our view other than what we are intentional towards. Husserl's phenomenology can account for this, while a third-person perspective to what we are conscious of, one that needs an 'I believe,' cannot. How could we believe something that enters our consciousness with out us knowing it? It seems clear, although Dennett denies the verifiability of this claim, that things enter our mind without our knowledge. For Husserl noema is the "actual components of intentional experiences" (Husserl 2014, 173) and noasis are there "intentional correlates"(Husserl 2014, 173). Husserl has a parallelism between an object side of intentional experience and a purely intentional side. Within this object side of our intentionality, there is a core we perceive. As he recounts, "Within the noema in its entirety, we have to sort out essentially diverse layers that group around a central 'core,' around the pure 'objective sense'" (Husserl 2014, 181).

Finally, Husserl's phenomenology can account for fantasies of fantasies and memories of memories and reflections of reflections something a third-person intentional stance could not. How does Husserl account for the fact that we can have a fantasy of a fantasy? He gives us the example of turning towards a picture in a gallery that is only in our memory, writing that we can be "Turned towards the 'picture' (not the depicted), we apprehend nothing actual as an object but instead just a picture, a fictum. The "'apprehension' is the actual process of turning towards the object, but it is not 'actual' apprehended is 'as though it were the case,' the positing is not a current positing but instead a positing modified in the 'as though it were the case' fashion," (Husserl 2014, 220). We see from this example then we could be not actually in a gallery but in a memory of the gallery. In this memory, it is possible that what we remember is not the same, but we can still move around within our conscious remembering. This is how

we can have a memory of a memory. It is impossible to imagine getting third-person verification on something such as a memory of a memory, as there is no way this could be reported to a third party with any kind of accuracy which captures this phenomena.

What would a combination of phenomenology and hetero-phenomenology look like? It might be maintained that putting them together is a Sisyphean task. If we understand both phenomenology and hetero-phenomenology as shifts in view point in order to capture consciousness, we could potentially understand the possibility of a phenomenology/hetero-phenomenology as the possibility of looking at one object of study, consciousness, from different viewpoints. One viewpoint, phenomenology, is the viewpoint that involves the epoche and phenomenological bracketing. For this viewpoint we capture a vast array of subjects. If we want a method that has the most dynamism in capturing the content of intentional experience we should perform the epoche and take up the phenomenological method. If we are interested in a naturalistic definition of conscious experience but one that is limited, we can switch to the intentional stance. These two could even be used for the same phenomena. For example, if we start with the hetero-phenomenological view that 'we believe we saw a painting in the Whitney Tuesday at 9pm last week' and report this to the hetero-phenomenologist, we get a certain understanding of our conscious experience by using this method. We can also infer from the hetero-phenomenological other results from 'I deduce x.' We know that we do not believe that we were anywhere else at this time. Once we hit this point, whether or not we have exhausted our possibilities within the hetero-phenomenological view, we can thank the party who helped us in this third-person belief attribution and take these results to then perform the epoche. We now explore the same experience from a phenomenological perspective and realize that this painting at the Whitney and the whole interpretation of it as an empirical object was influenced by a person wearing a red shirt standing next to the painting we were not focused on but who entered our consciousness. Also we where influenced by the view of a panting we saw at 8:45 previously. This experience through our flow of consciousness got caught up with the experience of looking at the painting we saw at 9 and we can then understand the way our conscious experience followed from our experience with the painting we saw at 8:45 to my memory of the painting at 9 (this is only one of countless possibilities).

We can also use the two different views for different phenomena we think they are better suited to. The hetero-phenomenological intentional stance may do us well for deciding certain data about the reason why someone behaves a certain way based on belief attribution that lets us assume certain things about their conscious intentionality about this experience. If someone eats at 6 o'clock every day we can assume they do

this because "they believe they are hungry" at this time or "they believe it is a good idea to have a regular schedule." If we want to explore how objects manifest themselves in memory or fantasy and what role intentional experience has towards these objects we can perform the epoche and try to look at consciousness flow between these intentional experiences and fantasy elements or memories of these experiences. We may employ this to explore other phenomena more fully as well since hetero-phenomenology does not have a way to account for nomatic cores of phenomena, and the way edges of phenomena factor into us being intentional towards these phenomena. We would employ these phenomenological methods in a more general fashion and less specifically to situations that do not necessarily deal with places where we can find norms and standards of rational belief attribution. This could be anything since there is no phenomena after the epoche is performed that cannot be looked at.

We can then have both a phenomenological and hetero-phenomenological view. Although this way of viewing consciousness has not been developed much herein it is the subject that warrants investigation more fully.

References

Dennett, Daniel. 1987a. "Setting Off on the Right Foot." In *The Intentional Stance*. Cambridge: MIT Press.

Dennett, Daniel. 1987b. "True Believers: The Intentional Strategy and Why it Works." In *The Intentional Stance*. Cambridge: MIT Press.

Dennett, Daniel. 2003. "Who's *On* First: Hetero-Phenomenology Explained." *Journal* of *Consciousness Studies* 10 (9–10): 19–30.

Dreyfus, Hurbert, and Sean Kelly. 2007. "Hetero-phenomenology: Heavy-Handed Sleight-of-Hand." *Phenomenology and the Cognitive Sciences* 6 (1): 45–55.

Husserl, Edmund. 2014. *Ideas*: *for a Pure Phenomenology and Phenomenological Philosophy*. Trans. Daniel O. Dahlstrom. Indianapolis, Hacekett.

Husserl, Edmund. 1965. "Philosophy as Rigorous Science." In *Phenomenology and the Crisis of Philosophy*. Trans. Quentin Lauer. New York: Harper and Row.

Journal of Cognition and Neuroethics

What You Don't Know *Can* Hurt You: Situationism, Conscious Awareness, and Control

Marcela Herdova
Florida State University

Biography
Marcela Herdova is Visiting Professor at Florida State University. She previously worked as Research Associate on the "Self-Control and the Person: A Multi-Disciplinary Account" project at King's College London and as Postdoctoral Research Fellow in Free Will and Self-Control at Florida State University. Her research interests include action theory, free will, moral psychology, consciousness and applied ethics.

Acknowledgements
I would like to thank Stephen Kearns for his extremely helpful feedback on various drafts of this paper.

Publication Details
Journal of Cognition and Neuroethics (ISSN: 2166-5087). April, 2016. Volume 4, Issue 1.

Citation
Herdova, Marcela. 2016. "What You Don't Know *Can* Hurt You: Situationism, Conscious Awareness, and Control." *Journal of Cognition and Neuroethics* 4 (1): 45–71.

What You Don't Know *Can* Hurt You: Situationism, Conscious Awareness, and Control

Marcela Herdova

Abstract

The thesis of situationism says that situational factors can exert a significant influence on how we act, often without us being consciously aware that we are so influenced. In this paper, I examine how situational factors, or, more specifically, our *lack of conscious awareness* of their *influence* on our behavior, affect different measures of *control*. I further examine how our control is affected by the fact that situational factors also seem to *prevent* us from becoming *consciously aware* of our *reasons* for action. I argue that such lack of conscious awareness *decreases* the degree of control that agents have. However, I propose that while being influenced by situational factors in such ways may impair and diminish one's control, it (typically) does not eradicate one's control. I further argue that being influenced by situational factors, in the way set out above, also *decreases* one's degree of moral responsibility.

Keywords

Situationism, Conscious Awareness, Control, Moral Responsibility, Bystander Experiments

1. Introduction

The thesis of situationism says that situational factors can have a significant influence on how we act, *often* without us being *consciously aware* that we are so influenced. Some have discussed how being affected by situational factors impacts our (having) character and virtues (e.g., Doris 2002, Miller 2013). Others have focused on what situationism tells us about autonomy (Nahmias 2007), freedom (Nelkin 2005), moral responsibility (Vargas 2013), and how situationism relates to moral luck (Herdova & Kearns 2015).

In this paper, I examine how situational factors, or, more specifically, our *lack of conscious awareness* of their *influence* on our behavior, affect our *control*. I further examine how our control is affected by the fact that situational factors also seem to *prevent* us from becoming *consciously aware* of our *reasons* for action. (I refer here to *normative* reasons—those reasons which *justify* actions). I argue that such lack of conscious awareness *decreases* the degree of control that agents have. However, I propose that while being influenced by situational factors in such ways may impair and diminish one's control, it (typically) does not *eradicate* one's control.

In the concluding section of the paper, I consider how my arguments about situationism and control affect considerations about moral responsibility. I propose that being influenced by situational factors, in the way set out above, also *decreases* one's degree of moral responsibility (in virtue of decreasing one's degree of control). This is not to say that that being influenced by situational factors exonerates agents *altogether*—situationist agents (i.e., agents influenced by situational factors) may still be held responsible (and, in some cases, blameworthy) for their actions.

Before I proceed to support my main theses, some clarifications are in order. Below I set out some assumptions about the nature of conscious awareness, control and moral responsibility.

1.1 Conscious Awareness

Because much of the discussion below concerns cases in which agents lack conscious awareness of various things, it is a good idea to start with a note on how I shall understand the idea of *conscious awareness*. My aim is to make remarks about its nature that are relatively uncontroversial, and that do not commit me to any specific theory of conscious awareness.[1]

How, then, should we think of conscious awareness? Though perhaps not *universal* or *essential* features of conscious awareness, I take it the following generally hold if S is consciously aware of X:

> S can *reflect* on X (S is able to form states that are *about* X).
>
> S can *report* the existence or obtaining of X.
>
> X can easily and readily serve as the basis for S's non-automatic overt behavior, reasoning, inference, and other related *personal* (i.e., not unconscious) processes.
>
> S's being aware of X has a distinctive phenomenal feel—*there's something it's like to be* for S to be aware of X.[2]

1. Given the purpose of this paper, committing to one (controversial) theory over others would be to unnecessarily alienate those who hold different theories. I wish to remain neutral between such theories (e.g., between the different higher-order theories of consciousness, access theories of consciousness, phenomenal theories of consciousness, etc.).

2. All of the above features correspond to different theories of conscious awareness, according to which being

Though, as I say, it may be true that some of these features may be absent in genuine cases of conscious awareness (e.g., we can conceive of a case when S cannot report that X because S is coerced into silence), in the vast majority of normal cases, all such features will be present. Conversely, in cases of unconscious awareness, the above features are (almost always) *absent*. So when one is, for example, consciously aware of a certain situational factor, one will be then, typically, able to reflect and report on it. Being consciously aware of this situational factor will also allow one to make use of this factor in non-automatic processes; for instance, one can formulate plans to utilize this factor or plan to avoid its influence, etc. Further, being aware of this situational factor will usually have a certain phenomenal feel—there will be something what it's like for the agent to be consciously aware of that situational factor.

1.2 Conscious Awareness, Control and Responsibility

How are conscious awareness and control related? Plausibly, an agent's conscious awareness of the relevant things can often *enhance* her control of her behavior. In a nutshell, if an agent is consciously aware of X, she can much more easily and straightforwardly formulate plans that incorporate X. Thus, for example, if an agent is consciously aware of a physical obstacle O to her performing an action A, she can plan her behavior in such a way that she avoids or overcomes O in executing her intention to A. Her conscious awareness of O helps the agent exercise greater control in translating her plans into action.

On the other hand, if the agent is *unaware* of O altogether, she cannot formulate plans that incorporate O. If she is *aware* of O, but not *consciously* aware of O, she will either not be able to formulate such plans at all, or not be able to do so with the ease and flexibility that she can when consciously aware of O. One can further expect that plans formed on the basis of unconscious awareness might lack the required complexity and detail, making them less effective. This is because, if an agent is merely *unconsciously* aware of O, O is not ready to serve as the basis of S's non-automatic personal behavior, such as reporting, conscious reasoning, and so forth, all of which equip agents with more multifaceted or sophisticated means of control over their behavior.

In essence, conscious awareness often increases an agent's control. Such awareness enhances the agent's control over her putting her plans into action, which I have

consciously aware of something simply *amounts* to (one of) those features. One might hold, for example, that being consciously aware of X *just is* X being available for reports, etc. I do not wish, for the reasons set out above, to commit to any such strong claims.

illustrated by the example of an agent's conscious awareness of a *physical obstacle* to her action. There are other things, however, of which one might have (or lack) conscious awareness, and other ways in which one's control might be increased (or decreased) as a result.

Of particular interest to us, given the topic of this paper, is the idea that conscious awareness of (a) certain relevant causal influences on one's actions, and (b) some of one's reasons for action can enhance one's control over one's behavior. I shall argue for the related claim that *lacking* conscious awareness of (a) or (b) can *decrease* our control over our behavior (and can do so in more than one way). Certainly, the claim that a lack of conscious awareness of and due to situational factors can decrease control has been considered before. Mele and Shepherd, for instance, entertain the hypothesis that:

> … people have very little control over their behavior … [behaviour] is largely driven by the situations in which people find themselves and the effects these situations have on automatic behavior-producing processes. (2013, 68)[3]

In this paper, I investigate in depth the ways in which our lack of conscious awareness of the influence of situational factors, and of the reasons which these factors obscure from us, can decrease the control we exercise over our behavior.

One important reason to explore this topic is the fact that control is connected to other significant notions—most obviously to *moral responsibility*. Moral responsibility is typically thought to *require* control—an agent is responsible for her action only if she exercises sufficient control over it. One worry is, then, that, by decreasing an agent's control, the agent's lack of conscious awareness both of and due to the influence of situational factors *also* decreases the agent's moral responsibility.[4] This worry comes in two varieties. First, we might worry that an agent's control is reduced to such an extent that she is entirely exculpated—that she bears no moral responsibility for her actions at all. Second, we might worry only that the agent is *less* responsible than she otherwise would have been, but is nonetheless responsible. In the moral responsibility section, I shall defend the latter claim.

It is worth noting that the claim that moral responsibility requires control is not uncontroversial. So-called non-volitionists reject this requirement, and, instead, insist

3. Mele and Shepherd do not, in the end, endorse this thesis.

4. This further presupposes that both control and responsibility come in degrees: i.e., we can have less or we can have more of either.

on other requirements that do not focus on control (for instance, Angela Smith [2008] proposes a *rational relations* view, according to which "To say that an agent is morally responsible for something ... is to say that that thing reflects her rational judgment in a way that makes it appropriate, in principle, to ask her to defend or justify it" [369]). In this paper I shall assume that non-volitionism is false, and that control is indeed central to moral responsibility.[5]

1.3 Outline

The paper proceeds as follows. In Section 2, I provide evidence from situationist experiments to the effect that agents lack conscious awareness of certain significant phenomena. In particular, I argue, in Section 2.1, that agents are often unaware of the *influence* of situational factors on their behavior. In Section 2.2, I establish that situational factors often make agents unaware of their *reasons* for action. I do not mean to suggest that agents subjected to powerful situational factors are *always* consciously unaware of the influence of these factors, or of their reasons for action. Indeed, in Section 2.3, I provide evidence that situational factors can affect us adversely even when we *are* consciously aware both of how they influence us and of our reasons for action. My point is simply that on *many* occasions, we lack such conscious awareness.

In Section 3, I show how this lack of conscious awareness can affect various measures of control. A measure of control is, roughly, something such that, if one *has* it, one's overall amount control is higher than if one lacks it. I argue, in Section 3.1, that lacking conscious awareness of our reasons adversely affects our *ability* to act on reasons. In Section 3.2, I propose that lacking conscious awareness of the influence of situational factors on our behavior adversely affects our *ability to directly combat* such influence. In Section 3.3, I argue that our *reasons-responsiveness* is decreased by our lacking either of these kinds of conscious awareness. In Section 3.4, I suggest that the effectiveness with which we translate our values into action is also decreased by our lacking either of these kinds of conscious awareness.

In Section 4, I conclude by addressing the implications of the above arguments for moral responsibility. In essence, I argue that moral responsibility is somewhat diminished

5. Given this assumption, one possible reaction to some of the results I adduce (that agents are less responsible than we might think) may be to reject the idea that control is so central to responsibility. However, there are plausibly ways in which lack of conscious awareness of and due to the influence of situational factors may threaten moral responsibility other than by affecting one's control. It is, however, beyond the scope of my paper to entertain this hypothesis.

in those agents who lack conscious awareness of the influence of situational factors, or of their reasons. Though reduced, however, responsibility is not eliminated.

2. Situationist Experiments and Lack of Conscious Awareness

As mentioned above, there are two ways in which we might lack conscious awareness when subject to certain situational factors. First, agents are often unaware of the ways in which these factors may or do influence their behavior. Second, situational factors may hinder agents from becoming consciously aware of their normative reasons for action. In this section, I examine both of these ways in more depth, and provide evidence that many situationist agents indeed lack such types of conscious awareness.

2.1 Lack of Conscious Awareness of the Influence of Situational Factors

Oftentimes we *are* indeed consciously aware of the different situational factors in our environment and how such situational factors affect our actions. For instance, I may want to cross the street but there is a red light for pedestrians, so I patiently wait for it to turn green. Once it does, I start walking across. Even though I may not explicitly think, in that very moment, about the fact that I started to cross the road because the light turned green, I will, most likely, be able to explain why I did so when I did (and point to the light turning green) *if* prompted to give an explanation (and report on the light being green, etc.). I am thus consciously aware of the green (and the red) pedestrian light and its impact on my actions. Examples similar to these are quite usual and abundant. However, various situationist experiments show that we in fact often lack conscious awareness of the (sometimes rather subtle) influence that situational factors have on our behavior. In the words of Matthew Lieberman:

> All of the most classic studies in the early days of social psychology demonstrated that situations can exert a powerful force over the actions of individuals…people are largely unaware of the influence of situations on behavior, whether it is their own or someone else's behavior. (2005, 746)

Take, for example, the *bystander experiments*, which show that the number of people one is accompanied by often makes a difference with regards to whether one offers assistance in an emergency situation. According to the so called *bystander effect*, the likelihood of helping in an emergency situation *inversely correlates* with the number of people present in that situation. In other words, the bystander experiments show that *the more* people

present in an emergency setting, *the less likely* it is that any of the individuals present will intervene. In an experiment conducted by Latané and Darley (1968), subjects witnessed smoke filling up a room. Out of those subjects who witnessed the smoke *on their own*, most of the subjects—18 out of 24—intervened in light of this (apparent) emergency. However, the number of intervening subjects was significantly smaller in a condition where the subjects were *accompanied* by two passive experimental confederates. In this condition, only one out of 10 experimental subjects intervened.

Similar results were observed by Darley and Latané (1968) in another bystander experiment concerning a medical emergency. In this experiment, the subjects overheard an (apparent) epileptic attack. Out of those who thought they were *alone* to witness this attack, 85% intervened in the specified timeframe (125s). In a condition where *four* other people also overheard the attack, only 31% of the subjects intervened in the said timeframe. Given the structure of the experiments, with the experimental conditions differing only in the number of people present, it is plausible to assume that whether the subjects intervened largely depended on their being accompanied or not.

Now, it is very likely that most (if not all) subjects in the above experiments were consciously aware of the salient situational factor (being accompanied/number of people present). However, the post-experiment debriefing interviews suggest that at least some of the subjects *lacked* conscious awareness of the *influence* of the relevant situational factor. With regards to the smoke experiment, Latané and Darley note that the majority of the experimental subjects claimed not to have paid any significant attention to the reactions of the people who accompanied them:

> Despite the obvious and powerful inhibiting effect of other bystanders, subjects almost invariably claimed that they had paid little or no attention to the reactions of the other people in the room. (1968, 220)

If that is indeed the case, it is implausible to conclude that the experimental subjects were consciously aware of how the presence of other people affected them, since this would require that they paid enough attention to those people and their reactions in the first place. There is, of course, a possibility that at least some of the subjects were indeed consciously aware of such an influence, but they did not want to disclose this fact to the experimenters (perhaps they were embarrassed about their reaction or, more precisely, lack thereof). Latané and Darley thus conclude that:

> Although the presence of other people actually had a strong and pervasive effect on the subjects' reactions, they were either unaware of this, *or* unwilling to admit it. (1968, 220) [italics added]

One certainly ought to be cautious about taking any such post-experiment interviews at face value. However, despite the fact that some experimental subjects might have been dishonest about what they took notice of (and thus about what influenced their actions), it is highly unlikely that *all* of the experimental subjects were lying in this manner.

The experimenters observed similar debriefing responses in the medical emergency experiment. Darley and Latané explain that they:

> asked all subjects whether the presence or absence of other bystanders had entered their minds during the time that they were hearing the fit. Subjects [accompanied by other people] … reported that they were aware that other people were present, but they felt that this made no difference to their own behavior. (1968, 381)

Again, while one may be somewhat (and rightly) concerned about the reliability of these subjective reports (and intentional or unintentional confabulation), it is implausible that all of the experimental subjects were dishonest about the perceived situational influences (or lack thereof) on their behavior. The post-experiment interviews in the bystander experiments thus provide evidence to the effect that at least *some* subjects in these experiments lacked conscious awareness of being influenced by the relevant situational factors.

Other situationist experiments also support the thesis that agents often lack conscious awareness of the influence of situational factors. Consider, for instance, a study by Bateson et al. (2006) in which the experimenters tracked the amount of 'honesty box' contributions for refreshments, in relation to the type of picture presented on the instruction sheet placed above the honesty box. People contributed to the honesty box, on average, 2.76 times more in those weeks when the information sheet had a picture of a pair of eyes, in comparison to when it had a picture of flowers. Given the results, it seems that being exposed to the images of eyes had significant influence on whether people paid for the refreshments or not.

Were the experimental subjects consciously aware of the fact that the images on the instruction sheet had this kind of impact on their behavior? Due to the lack of post-experiment interviews in this case, it may seem more difficult to establish what the subjects were consciously aware of, at the time they had the opportunity to contribute to

the honesty box. However, given the findings on how sensitive our perceptual system is to different social cues such as faces (see, e.g., Emery 2000; Haxby et al. 2000),[6] Bateson et al. (2006) entertain the hypothesis that the subjects were *not* consciously aware of how the images impacted them:

> it is therefore possible that the images exerted an automatic and unconscious effect on the participants' perception that they were being watched. (2006, 413)

When discussing how such situational cues may enhance cooperative behavior—by inducing a feeling of "being observed", and, subsequently, triggering "reputational concerns"—the experimenters further build on the thesis that the aforementioned situational cues affect agents on an unconscious level (with agents lacking conscious awareness of this influence):

> If even very weak, subconscious cues, such as the photocopied eyes used in this experiment can strongly enhance cooperation, it is quite possible that the cooperativeness observed in other studies results from the presence in the experimental environment of subtle cues evoking the psychology of being observed. The power of these subconscious cues may be sufficient to override the explicit instructions of the experiment to the effect that behaviour is anonymous. (2006, 413)

There are other experiments in this paradigm, involving even more subtle face/eye-based situational cues, which demonstrate that people often lack conscious awareness of the impact such cues have on their behavior. In the dictator game experiment, Rigdon et al. (2009) tracked the amount of contributions in relation to the arrangement of three dots on a sheet, which the subjects used for noting down their contributions. The experimenters found that, on average, male players whose sheet of paper contained three dots arranged in the shape of a face contributed $3.00; while those in the neutral dots condition contributed $1.41. Given that the experimental conditions were relevantly similar except for the arrangement of the dots on the contribution sheet, it seems that the shape of the dot arrangement largely contributed to the amount of one's donations.

The above experiment (and other similar experiments in this paradigm) shows that even extremely subtle cues in the form of a face or a pair of eyes can have a rather strong impact on what people do (in this case, how much money [or whether] they contribute

6. These references are taken from Bateson et al. 2006.

in the dictator game). More importantly, within the context of the current debate, it is highly unlikely that the experimental subjects were consciously aware that they were being so influenced. Rigdon and colleagues agree with this diagnosis when explaining the mechanism through which such cues likely influence the agents' behavior:

> Processing the stimulus ultimately activates the fusiform face area of the brain, making the environment seem—at a pre-conscious level, perhaps accessible to the decision-making process but not to introspection... (2009, 363).

Aside from appealing to the workings of the human perceptual system, there are at least two other points which reinforce the conclusion that many subjects in the above experiments lack conscious awareness of the influence of the situational cues.[7] In the first instance, in many of the experimental situations (and *similar* situations outside the experimental setting), being consciously aware of the influence of situational cues on one's behavior requires that one *knows* that the relevant situational cues *can* indeed have such an influence (or, in some cases, one needs to have knowledge about the mechanisms in virtue of which these cues might influence one's behavior). However, most people do not know the relevant research, and are not likely to be familiar with the pertinent facts: people do not typically know how seeing faces or eyes (or subtle cues in the shape of faces or eyes) might affect them. Similarly, not many people are educated about the bystander effect and the potential influence of the presence of other people on their behavior. This applies to many other documented effects of situational cues. Some of these show that a mood boost, resulting from, for example, the agent being subject to pleasant fragrances (Baron 1997), or the agent finding a small amount of money (Isen and Levin 1972) is often conducive to her helping others. Again, this is not something that the general public is (well) educated about. It is thus unlikely that people are, typically, consciously aware of the effect that the different situational cues have on their behavior because they lack knowledge they could be potentially so influenced.

Secondly, many people are likely to find being influenced by such arbitrary situational factors as *undesirable*—typically, we value our decisions and our actions being based on reasons and other relevant facts. For instance, it is valuable if our decision to intervene in a medical emergency is based on the fact *that* there is someone who needs medical attention, *that* we are able to provide the relevant kind of assistance, *that* helping

7. A good case can be made that, sometimes, subjects are not even consciously aware of the situational cues *themselves*. I do not, however, need to establish this point for my purposes here.

someone in need is the moral or virtuous thing to do, etc. On the contrary, it is somewhat troubling if our decision to help were to be largely based on situational factors like the ones outlined above (for example, how many other people one is accompanied by, whether there are posters with faces in our immediate surroundings, etc.). Such decisions or actions would seem to lack the appropriate motivation. Now, if it is the case that many people would find the influence of such arbitrary and normatively irrelevant situational factors undesirable, it seems reasonable to expect that—if they were, at the same time, appropriately aware of such (potential) influence—they would attempt to combat it. However, the situationist data suggest that people often do succumb to such influences (and, at the same time, the subjects often do not appear to try to combat those either). Then, given the perceived undesirability of this type of influence, it thus makes it unlikely that people are consciously aware of it. This is further supported by the observation that people who are educated about the influence of situational factors, such as the bystander effect, are less likely to be adversely affected by it (for a more extended discussion on this see, for example, Mele and Shepherd 2013).

2.2 Situational Cues and Lack of Conscious Awareness of Reasons

Aside from agents lacking conscious awareness of the situational influence, it may be that, in some cases, the situational factors prevent agents from becoming consciously aware of the relevant normative reasons. First of all, agents may be unaware, due to their being influenced by situational factors, that a certain fact which is a reason obtains. Second of all, agents may be unaware, due to the situational influence, that a reason is a *sufficient* or a reason for action (i.e., the kind of reason that determines what one ought to do). Let me expand on and illustrate these points with different situationist experiments.

What does it mean to say that a subject may not be consciously aware that a certain fact, which is a reason to act, obtains? Simply that there is a fact or a state of affairs which also is a reason for the subject to act in a certain way, and the subject is not consciously aware of this fact/state of affairs. Take, for instance, the bystander smoke experiment. The relevant fact, which is also the subject's reason to act, is that there is a potentially dangerous situation occurring (there is smoke filling up a room). That there is such a potentially dangerous situation is a reason for the subject to do something about it—to alert the authorities, to try to locate the source of the smoke (or whatever else one may do in such circumstances to avert the potential danger). To lack conscious awareness of this fact amounts to failing to consciously become aware that one is facing

a potentially dangerous situation. This is, indeed, what seems to happen in the smoke bystander experiment. When the subjects were asked in the debriefing interviews if they encountered any difficulties while in the waiting room, most subjects did mention the smoke. However, when further prompted to explain what happened, Latané and Darley state that:

> Subjects who had not reported the smoke … *uniformly* said that they had rejected the idea that it was a fire. Instead, they hit upon an astonishing variety of alternative explanations, all sharing the common characteristic of interpreting the smoke as a non-dangerous event. (1968, 219) [italics added]

According to the experimenters, *all* of the subjects who failed to report the smoke interpreted the situation in a similar fashion: as something not dangerous. This means that they failed to consciously recognize or consciously become aware that there was a potentially dangerous event occurring which needed to be reported. Of course, as explained in the previous section, one may be concerned about the reliability of these debriefing reports. However, it is unlikely that all such reports, or even a large proportion, were unreliable.

Another good example to illustrate a lack of conscious awareness of reasons is the *Good Samaritan* experiment, conducted by Darley and Batson (1973), in which seminary students were asked to give a talk in a nearby building. Making their way to the lecture hall, the seminarians came across a person in apparent need of medical help. Some students were told they were running late. Only 10% of the students in this group offered assistance. On the other hand, out of those in a low-hurry condition (who were told they had enough time), 63% of the subjects helped. The students did also differ, aside from how much time they had, in the content of their lecture: some were going to talk on the parable of the Good Samaritan, and some on job prospects. However, while the hurry factor did make a significant difference with regards to whether they offered assistance or not, their lecture content did not.

In the post-experiment interviews, all subjects mentioned the victim—on reflection—as possibly needing help. However, Darley and Batson suggest that some of the participants seem not to have worked this out when they were near the victim, either (i) failing to interpret the situation in a timely fashion as that of someone requiring help, or (ii) being delayed in their empathetic reaction. According to the experimenters, it would be inaccurate to claim, about at least some of the subjects, that they:

> realized the victim's possible distress, and then chose to ignore it; instead, because of the time pressures, they did not perceive the scene in the alley as an occasion for an ethical decision. (1973, 108)

This suggests that at least some participants failed to interpret the relevant reason as a fact (that someone needed help) and/or that they failed to recognize the fact as something that gives them a reason to help (in those cases where their empathetic reaction might have been delayed).

This is, however, not true for all of the subjects in the Good Samaritan experiment. For some, it is more accurate to say, according to the experimenters, that they *decided* not to help. This choice was presumably a result of a conflict between stopping to help and fulfilling the duty to carry out the experiment. In these cases, then, it seems more fitting to say that subjects recognized the relevant fact as a reason to act, but decided to act in line with a conflicting reason. This may suggest that these subjects were unaware of the *strength* of the reason they had to help (and the comparative *weakness* of the reason they had to get to the talk on time)—they were not aware that their reason to help was *sufficient*. Both of these sets of judgments and attendant behaviors (failing to interpret a reason as a fact and failing to recognize that a reason is sufficient) may be ascribed to the influence of the relevant situational factor (being in a hurry).

It should be noted that the above remarks about a lack of conscious awareness do not apply solely to the subjects in the situationist experiments. Given the structure of the experiments, it is reasonable to assume that the experimental results generalize to the population at large. After all, the experimental subjects were assigned their experimental conditions randomly, and the subjects were not chosen for the experiments on the basis of their susceptibility to situational factors. That is, the data above (and other data from the situationist literature) strongly suggest that all of us are very often significantly affected by the presence of various situational factors. In other words, what we may do (or refrain from doing) in different scenarios largely depends on the presence of arbitrary situational factors, and, what is more, we often *lack conscious awareness of this dependence*.

2.3 Situationism and the Presence of Conscious Awareness

It needs to be noted, however, that in some situationist experiments, the subjects do seem to be consciously aware of the influence of situational factors, and, at the same time, these situational factors do not seem to prevent people from becoming consciously aware of their normative reasons for action. Consider, for instance, the obedience

experiments conducted by Stanley Milgram (1963, 1974). The experiments focused on studying subjects' behavior under the influence of authority. Subjects, who believed that they were taking part in a learning experiment, were asked, by a figure of authority, to deliver apparent electric shocks to "learners", upon the learners providing wrong or no answers to the relevant questions. Since the subjects were strongly encouraged (by the authority figure) to keep delivering the shocks despite the learners' apparent discomfort (which, in some experimental conditions, was rather graphically displayed), it is likely that the subjects were consciously aware of the authority's influence on their decision to keep going on with the experiment, and to keep delivering what appeared to be increasingly higher and higher shocks.[8] (This assumes, of course, that the subjects did not have other reasons to stick with the experiment, such as that they would enjoy causing pain to the learners).

It is extremely plausible that many of the subjects in these experiments were not just consciously aware that (a) the shocks apparently caused someone extreme pain (given the nature of the auditory and/or visual feedback they received), but also that (b) this fact is a reason to stop pulling the levers, and (c) this reason is *sufficient*. Despite this, these subjects acted in line with the requests of the confederate. Milgram notes that the experimental procedure created "extreme levels" of nervous tension in the subjects, many of which:

> showed signs of nervousness in the experimental situation, and especially upon administering the more powerful shocks. In a large number of cases the degree of tension reached extremes that are rarely seen in sociopsychological laboratory studies. Subjects were observed to sweat, tremble, stutter, bite their lips, groan, and dig their fingernails into their flesh. These were characteristic rather than exceptional responses to the experiment. … Fourteen of the 40 subjects showed definite signs of nervous laughter and smiling. … Full-blown, uncontrollable seizures were observed for 3 subjects. (1963, 375)

8. In Experiment 1, approximately two-thirds of the subjects complied with the instructions of the experimental confederate, and continued to deliver shocks all the way (i.e., pulling all 30 levers, including the one delivering the highest degree of shock). The subjects continued to increase the voltage despite the fact that after the 20th question, the learner apparently receiving the shocks would bang on the wall and then stop providing answers.

After the experiment, when the maximum shocks had been delivered:

> many obedient subjects heaved sighs of relief, mopped their brows, rubbed their fingers over their eyes, or nervously fumbled cigarettes. Some shook their heads, apparently in regret. (1963, 376)

Such levels of distress are indicative of the fact that the subjects were conflicted about their actions and about continuing with the experiment, and that they were appropriately consciously aware of their reasons to stop delivering the apparently lethal shocks.

Not every case of being influenced by situational factors is thus of a kind where people lack the relevant conscious awareness, yet those kinds of situations seem to be abundant nonetheless. In the following section, I explore the implications of lacking such conscious awareness on considerations about agents' control.

3. Lack of Conscious Awareness and Measures of Control

In this section, then, I shall examine four different measures of control and how an agent's lacking conscious awareness, resulting from the influence of situational factors, can affect these measures of control. Recall that by "measure of control" I mean a feature such that the greater degree to which an agent has this feature, the greater degree of control the agent exercises over her behavior (all other things being equal). The features I examine below include the ability to act on one's sufficient reasons, the ability to directly combat pernicious influences on one's behavior, reasons-responsiveness, and the effectiveness with which one translates one's long-term values into action. Each of these is a measure of control—having these features (or having them to greater degrees) enhances one's control, while lacking them decreases one's control. I shall argue that a lack of conscious awareness (either of the influence of situational factors on one's behavior or of one's reasons) adversely affects *each* of these measures of control.

3.1 Ability to act on (sufficient) reasons

The first measure of control we shall consider is the ability to act on sufficient reasons. In my terminology, having a *sufficient* normative reason to perform an action entails having an *obligation* to perform it. In many of the experiments I discuss in section 2, agents have sufficient reasons—for example, the seminarians *ought* to help the person at the side of the road; the subjects in the smoke bystander experiment *ought* to alert someone of the potentially dangerous situation. In this subsection, I shall set out why

agents' lacking conscious awareness of their sufficient reasons rids them of their ability to act on such reasons.

Before this, however, it is worth mentioning why such an ability is a measure of control in the first place. One simple reason is that *abilities in general* are measures of control. Broadly-speaking, an agent has more control the more she is able to do. Another reason is that the ability to act on sufficient reasons is a particularly *significant* ability—it is the ability to be guided by reason—by what one ought to do. If someone *lacks* this ability—be it a psychopath, someone who is severely schizophrenic, etc.—we judge that she is also *less in control* of her actions—rational considerations simply cannot move her.

Why does an agent's lacking conscious awareness prevent her from being able to act on her sufficient reasons? Roughly put, in order to act on one's sufficient reasons, one must know about these reasons (one must know, at the very least, that they are facts). If one does *not* know about one's sufficient reasons, then, one cannot act on them. The blind person who walks obliviously past a person in medical need cannot help this person because she has no idea at all that there is anyone near her who needs help. Of course, should the blind person become aware of the person in need (perhaps because the person manages to shout for help), *then* she is able to act on her reasons to help. But, up until this time, she is not able to help.

Similarly, then, an agent who, due to the influence of situational factors, is not consciously aware of her sufficient reasons to act, *cannot* act on these reasons. She lacks the ability to act on her sufficient reasons because she is not conscious of these reasons. The seminarian who, due to being in a hurry, fails to (consciously) notice that the person at the side of the road (apparently) needs help, cannot act on the basis that the person needs help. The subject who is not consciously aware that there is a potentially dangerous situation cannot act on this fact.

Perhaps, one might argue, an agent need not be *consciously* aware of her sufficient reasons to be able to act on them, but rather simply *aware* of them—consciously *or* unconsciously. It is, however, deeply unlikely in the cases that we are considering that being merely unconsciously aware of sufficient reasons would enable the agent to act on these reasons. When a person unconsciously acts on a reason, she cannot say why she is doing what she is doing (indeed, she may not even be conscious of *what* she is doing). Situationist experiments such as the bystander studies and the Good Samaritan experiment concern actions that one can *only* perform for a reason *if* one is consciously aware of this reason. Those subjects who *do* help someone in need, or alert people of a potentially dangerous situation, can *of course* say why they are doing so. Consider how strange it would be if someone were unable to tell you that they were helping a person

because this person needed help, or how bizarre it would be if someone could alert the authorities after seeing smoke, but simply could not report that the smoke (and potential fire) were *why* she alerted them. In the kinds of cases relevant to our discussion, then, agents are unable to act on reasons unless these agents are consciously aware of these reasons. Because situational factors can block such conscious awareness (as section 2 spells out), situationist agents are often not able to act on their sufficient reasons.

3.2 Ability to directly combat pernicious influences

Another measure of control affected by the undue influence of situational cues is one's ability to *directly* combat or counter pernicious influences on one's behavior. This may affect those situationist agents who lack conscious awareness of being influenced by situational factors (rather than of their reasons for action). This is because directly (and effectively) combating negative influences on one's behavior requires that one is consciously aware of such influences—otherwise one does not (consciously) know *that* there is anything to combat or counter in the first place. Consider, for example, combating the bystander effect. In order to be able to directly attempt to eliminate this effect on an agent's behavior, the agent must be consciously aware that she is (or can be) so influenced. This enables her to undertake direct measures to counter this effect. For example, she may purposefully direct her attention away from other people, exert more effort in overcoming any social pressure she might feel, or remind herself that the presence of other people ought not to make a difference to what she should do/what the right thing to do is, etc. Without conscious awareness of the effect bystanders may have on one's behavior, one cannot directly employ any such strategies which eradicate (or at least lessen) this effect on one's behavior.

Even if we assume that being *unconsciously* aware of the potential negative influence of situational factors might *too*, indirectly, allow the agent to employ some strategies against this influence, such strategies will be certainly less effective. Being consciously aware of the pernicious influences of situational factors gives an agent more, and more effective, ways in which she can combat this influence. Given that some situationist agents do indeed lack conscious awareness of this influence (as set out in Section 2), we can conclude that their ability to directly combat pernicious influences is eliminated. Such an agent thus only retains an indirect ability of this kind (which is arguably a lot less effective).

3.3 Reasons-Responsiveness

Ascertaining how responsive an agent is to her reasons is another way by which we might measure the control an agent has over her behavior. Roughly-speaking, the more responsive an agent is to her reasons, the more she is in control of her actions. This is because to be so in control is, in part, to be *guided* by one's reasons.[9] Exactly how the idea of reasons-responsiveness should be spelled out is a difficult and interesting question.[10] For our purposes, however, we do not need to rely on a particular theory of reasons-responsiveness. It will suffice to say that an agent is more reasons-responsive in a particular situation the greater her capacity to recognize, understand, deliberate about, reflect on, and act on the basis of her reasons.[11] Thus, for example, a psychopath who is simply unable to grasp moral reasons for action is (far) less reasons-responsive than the average person—she does not recognize the moral reasons she has, she does not understand the idea that they *are* reasons, she does not act on their basis, etc.

Reasons-responsiveness obviously comes in degrees (one can recognize more or fewer reasons, one can have greater or lesser understanding of them, etc.). In this subsection, I shall present two arguments that situational factors, and the lack of conscious awareness they bring about, *decreases* agents' reasons-responsiveness (I do not claim, however, that agents' reasons-responsiveness is eliminated entirely).

As stated above, situational factors can cause agents to lack conscious awareness of at least two things—first, agents might be rendered unaware of their normative reasons for action (such as when the bystander effect leads agents to interpret smoke as harmless, and thus causes them to be unaware of their reasons to alert someone), and second, agents might be made unaware of the very fact that these situational factors are influencing them (agents subject to the bystander effect are often not conscious of the fact that their actions are highly influenced by their being accompanied). *Both* of

9. Fischer and Ravizza 1998 spell out their notion of *guidance control* as an agent's being reasons-responsive, while Wolf 1990 conceives of the type of control required for freedom as being tightly connected to an agent's ability to be guided by her reasons.

10. The most influential such account is that of Fischer and Ravizza 1998. See Herdova and Kearns (MS) for a close study of how the influence of situational factors affects agents' reasons-responsiveness as conceived of by Fischer and Ravizza.

11. Reasons-responsiveness does not simply amount to the *ability* to act on one's sufficient reasons. An agent may have the above-mentioned capacities without being able to act on her sufficient reasons because, for example, external obstacles prevent her from exercising these capacities. In such a case, the agent may count as reasons-responsive without being able to act on her sufficient reasons. The measure of control considered in this subsection is thus different from the measure of control considered in 3.1.

these ways in which agents can lack conscious awareness can decrease agents' reasons-responsiveness. Let us consider them in turn.

Why does the fact that an agent lacks conscious awareness of her reasons make her less reasons-responsive? Simply put, an agent's lacking conscious awareness of her reasons is at least partly *constitutive* of her having a lower degree of reasons-responsiveness than someone who *has* such conscious awareness. In subsection 1.1, I highlighted various features of conscious awareness. These included the fact that when an agent is consciously aware of X, X can readily serve as the basis for her overt behavior, reasoning, inferring, etc.—X can be incorporated into the agent's plans with ease and flexibility. They also included the fact that the agent can reflect on X, and the fact that she can report on X. If an agent is *not* consciously aware of X, she does not have all of these capacities. But it is exactly these capacities, amongst others, that make up an agent's reasons-responsiveness. The more easily an agent can base her behavior on her reasons, can reflect on them, deliberate about them, report them, etc., the more reasons-responsive she is. Thus having conscious awareness of reasons increases the degree to which one is reasons-responsive.

A *lack* of conscious awareness of one's reasons, then, results in a lower degree of reasons-responsiveness. And because, as I have argued in Section 2, certain situational factors often *cause* such a lack of awareness, these factors thereby reduce agents' reasons-responsiveness. In so doing, these situational factors reduce the control agents have over their behavior.

Why might the fact that an agent lacks *conscious* awareness of the influence of situational factors make her less reasons-responsive? The idea is simple enough. If we are not consciously aware of the influence of situational factors on us, then, partially *because* of this lack of awareness, many such factors can (and do) make us worse at forming beliefs about reasons on the basis of evidence. Being worse at this is *itself* one way of being less reasons-responsive. Thus when we are not consciously aware of the influence of situational factors on us, we are less reasons-responsive than we otherwise would be.

I take it that the second premise of the above argument (that being worse at forming evidence-based beliefs about reasons translates to being less reasons-responsive) is relatively obvious—part of what contributes to one's degree of reasons-responsiveness is how well one forms beliefs about reasons on the basis of one's evidence. What of the first premise—that it is precisely our lack of *conscious* awareness of the influence of situational factors which allows these situational factors to adversely affect how we form beliefs about reasons? It is clear that situational factors *do* adversely affect the manner in which we form beliefs about reasons. Those subjects in the bystander experiments who

are accompanied have just as much evidence that someone is in medical need, or that there is a potentially dangerous situation, as those who are unaccompanied. Despite this, many such subjects fail to realize these facts. Situational factors often do, then, make us worse at forming evidence-based beliefs about our reasons.

Part of why this is so is that we cannot directly combat the influence of situational factors, and part of why we cannot directly combat this influence is that we are not *consciously aware* of it (see 3.2 for a more in depth defense of these claims). In essence, because we are not consciously aware of the ways in which situational factors influence us, we cannot effectively counter the negative ways in which these situational factors affect how we form beliefs about reasons. Thus by lacking conscious awareness of the influence of situational factors, these factors can render us less reasons-responsive than we otherwise would be.

I conclude, then, that reasons-responsiveness is often diminished due to an agent's lacking conscious awareness of either her reasons or the influence of situational factors on her. Given that reasons-responsiveness is a measure of control (because the more reasons-responsive one is, the more control one enjoys), we may further conclude that an agent's control can be diminished when she is not consciously aware of her reasons or how situational factors affect her.

3.4 Translating long-term goals and values into action

Being affected by situational factors and lacking the relevant kinds of conscious awareness also makes us less effective in translating our long-term goals and values into action. These goals and values may include helping others, acting compassionately or with kindness, having certain religious values, helping oneself or self-preservation, etc.

In the first instance, translating long-term goals or values into action can be negatively affected by one's lack of conscious awareness of *reasons*. This is because the implementation of such goals and values requires that the agent *perceives* the relevant situation as an *occasion* for their execution. For example, implementing one's goal of assisting others in need requires that one is aware that one is presented with an opportunity to assist someone. If an agent lacks awareness of normative reasons, she is rather unlikely to perceive her situation as an occasion to translate the corresponding long-term goals and values into action. This is because recognizing that one has a normative reason to A just *amounts to* recognizing that A-ing is needed or justified in the given situation. For example, recognizing that one has a (normative) reason to help amounts to recognizing that one is in a situation where help is needed.

Now, if an agent is *unconsciously* aware of the relevant normative reasons, and thus unconsciously recognizes that she is facing a situation where her goals or values can be implemented, this gives her *some* opportunity to translate these into action (in comparison with a case when she lacks awareness altogether). However, being *consciously* aware of one's reasons, and, correspondingly, consciously recognizing that one has an opportunity to translate one's goals into action, significantly enhances one's effectiveness or chance of doing do so (due to increased flexibility, etc.).

Now, as I have shown above, at least in some experiments, agents do lack conscious awareness of their reasons for action due to being influenced by certain situational factors. It is thus, minimally, more difficult for these agents (and other agents in relevantly similar situations) to translate their values and goals into action, in comparison with those agents who *are* consciously aware of their reasons for action. (One might even suggest that some of the former agents are *unable* to translate their goals and values into action altogether if they lack conscious awareness of the relevant reasons).

Translating goals and values into action may be negatively affected not only by one's lack of conscious awareness of reasons, but also by one's lacking conscious awareness of being influenced by situational factors. Suppose that an agent values not being influenced by some normatively irrelevant factor. For instance, she might strongly disvalue that her decisions about whether to help out in an emergency situation should be based solely (or at all) on things such as the clothes the person in need of assistance is wearing, or, relevant to the discussion above, whether there are other bystanders around. Now, if this agent is exposed to such situational factors and ends up being influenced by them (due to the fact that she lacks conscious awareness of their influence and thus fails to combat it), she will then fail to act in accordance with her values. An agent's implementation of long-term goals and values into action may be, then, negatively affected by her failing to become consciously aware of the influence of situational factors as well.

4. Moral Responsibility

I have shown above that the lack of conscious awareness which may occur when agents are influenced by certain situational factors diminishes various measures of control. All four measures of control that I discuss in the previous section are indeed negatively affected by such situational influence. Recall that both the ability to act on sufficient reasons as well as the ability to directly combat pernicious influences on one's behavior are arguably completely eradicated if one lacks the relevant conscious awareness (possibly leaving the agent only with an indirect—and a lot less effective—ability of the

latter kind). Further, the effectiveness with which one translates long-term goals and values into action, while maybe not completely eliminated, is significantly decreased. With regards to reasons-responsiveness, this measure of control is also diminished given the fact that part of what makes a person reasons-responsive is *precisely* that one is consciously aware of the relevant reasons.

What does this mean for the *overall* amount of control of those agents whose behavior is influenced by situational factors in this way? The most straightforward conclusion is that the overall level of control that such agents have is *diminished*. After all, the different measures of control are what *constitutes* an agent's having control, and so diminishing one or more measures of control available to the agent will also diminish her overall amount of control.

Why not then say, in the light of the above observations about different measures of control, that a situationist agent, whose measures of control are affected by an undue situational influence, lacks control of her behavior altogether? Simply because the situationist data do not warrant this kind of strong conclusion—while situational cues may diminish the control an agent has over her behavior, they do not make her completely powerless. First, there are some measures of control which, even if somewhat negatively affected, are not completely eradicated (such as reasons-responsiveness). Second, there are arguably some measures of control which are not affected at all by the situational influence and the attendant lack of conscious awareness.[12] The situationist agents thus retain some control. However, in comparison with those who are not so influenced, agents who do lack conscious awareness of the situational influences on them, or of their reasons for action, will have, keeping everything else equal, *less* control over their actions.

What of implications for moral responsibility? Assuming volitionism, the view on which moral responsibility *requires* control, decreased behavior control correlates with decreased responsibility. In other words, the more control one has, the more responsible one is. Conversely, the less control one has, the less responsible one is. So, those agents who are influenced by situational factors in the way outlined above will be *less* responsible for what they do, in comparison with someone not so influenced who acts in a similar

12. These may include, for instance, *self-control* and *the ability to do otherwise*. Of course, some may try to claim that even these measures of control might be significantly affected when one is influenced by situational factors in the way set out above. It is, unfortunately, beyond the scope of my paper to engage with this point here. It should be noted though that defending myself against this worry is not essential to my argument—as long as the situationist agent retains some relevant proportion of at least one of the measures of control considered in Section 3, then one cannot claim that this agent lacks control altogether.

way (again, keeping everything else equal). Take someone who fails to help due to the bystander effect. According to my line of reasoning, this person will be less responsible than someone who fails to help but who is not subject to the influence of the relevant situational factors. This is because the first agent will have less control over what she does than the latter agent. However, it is important to bear in mind that situationist agents ought not be excused altogether for what they do—given that their control is not completely diminished, neither is their moral responsibility.

Some may think this last claim is too quick. Perhaps, such people may venture, some of the measures of control that are eliminated by a lack of conscious awareness caused by situational factors are *required* for having *any* amount responsibility whatsoever. Of the four measures of control discussed above, only two are plausibly eradicated completely (these are the abilities to act on sufficient reasons and to directly combat pernicious influences).[13] In my opinion, of these two measures of control, only the first is a plausible candidate for being required for moral responsibility.[14] Indeed, Susan Wolf claims that "an agent is responsible if and only if the agent can do the right thing for the right reasons." (Wolf 1990, 68). In essence, Wolf claims that an agent is responsible for an action only if she is able to act on her sufficient reasons. (One argument for this position runs as follows: free will is required for responsibility [an agent is responsible for an action only if she performs it freely], and free will is best understood as the ability to do the right thing for the right reasons/sufficient reasons; thus such an ability is a necessary condition of responsibility.)

Any view, however, which requires of moral responsibility that an agent possesses certain *abilities* should be treated with considerable caution. This is because, since Harry Frankfurt (1969), various cases have been concocted that (at least seem to) show that responsible agents need not possess some specific abilities. Consider the following case, based on Frankfurt's, that specifically targets the claim that an agent is responsible only if she is able to act on her sufficient reasons:

> Ethan has sufficient reason to help someone nearby to him—Warren—who is in medical need. Unbeknownst to him, if he even shows signs of

13. Agents adversely affected by their lack of conscious awareness may be *less* reasons-responsive and *less* effective in translating their values into action, but they are not *totally* unresponsive to reasons, nor *completely* ineffective at translating their values into action.

14. The ability to directly combat pernicious influences on one's behavior is not *required* for responsibility in part because one may lack this ability and yet still succeed in *indirectly* combating such influences, in which case one would be praiseworthy (and hence responsible) for one's actions.

> choosing to help Warren, a spell cast by a powerful witch, Willow, will make him instead choose to walk past Warren, without doing anything to help. As it happens, however, Ethan decides entirely on his own to walk past Warren without doing anything to help—the spell did not need to kick in at all.

In such a case, Ethan is *responsible* (indeed, blameworthy) for not helping Warren (and thus responsible for failing to act on his sufficient reasons). After all, Ethan's choice not to help Warren is made perfectly under his own steam—the spell has nothing to do with him making the choice. Indeed, had Willow not cast the spell, Ethan would have made the same decision, for the same reasons, and in the same way. Ethan is nevertheless *unable* to help (or even choose to help) Warren, and thus unable to act on his sufficient reasons. This is because, should Ethan show any sign at all of choosing to help, the spell would kick in and prevent him from doing so. Therefore, such an ability is not required for moral responsibility.

I conclude, then, that a lack of conscious awareness of the influence of situational factors, or of one's reasons for action, brought about by the situations one faces, can diminish the degree of control one exercises over one's behavior. In turn (assuming volitionism), this decrease in control mitigates one's moral responsibility—one is less responsible that one would otherwise have been. One does *not*, however, bear *no responsibility at all* for one's behavior. This is because one still exercises some degree of control over one's actions. Though lacking conscious awareness of certain things excuses us to some extent, we are still accountable for what we do.

References

Baron, R.A. 1997. "The Sweet Smell of … Helping: Effects of Pleasant Ambient Fragrance on Prosocial Behavior in Shopping Malls." *Personality & Social Psychology Bulletin* 23: 498–503.

Bateson, M., D. Nettle, and G. Roberts. 2006. "Cues of Being Watched Enhance Cooperation in a Real-World Setting." *Biology Letters* 2 (3): 412–414.

Darley, J.M., and B. Latané. 1968. "Bystander Intervention in Emergencies: Diffusion of Responsibility." *Journal of Personality and Social Psychology* 8: 377–383.

Darley, J.M., and C.D. Batson. 1973. "From Jerusalem to Jericho: A Study of Situational and Dispositional Variables In Helping Behavior." *Journal of Personality and Social Psychology* 27: 100–108.

Doris, J.M. 2002. *Lack of Character: Personality and Moral Behavior*. New York: Cambridge University Press.

Emery, N.J. 2000. "The Eyes Have it: the Neuroethology, Function and Evolution of Social Gaze." *Neuroscience and Biobehavioral Reviews* 24: 581–604.

Fischer, J.M., and M. Ravizza. 1998. *Responsibility and Control: A Theory of Moral Responsibility.* New York: Cambridge University Press.

Frankfurt, H.G. 1969. "Alternate Possibilities and Moral Responsibility." *The Journal of Philosophy* 66 (23): 829–839.

Haxby, J.V., E.A. Hoffman, and M.I. Gobbini. 2000. "The Distributed Human Neural System for Face Perception." *Trends in Cognitive Sciences* 4: 223–233.

Herdova, M., and S. Kearns. 2015. "Get Lucky: Situationism and Circumstantial Moral Luck." *Philosophical Explorations* 18 (3): 362–377.

Herdova, M., and S. Kearns. MS. "This is a Tricky Situation: Situationism and Reasons-Responsiveness."

Isen, A.M., and P.F. Levin. 1972. "Effect of Feeling Good on Helping: Cookies and Kindness." *Journal of Personality and Social Psychology* 21 (3): 384–388.

Latané, B., and J.M. Darley. 1968. "Group Inhibition of Bystander Intervention in Emergencies." *Journal of Personality and Social Psychology* 10 (3): 215–221.

Lieberman, M. 2005. "Principles, Processes, and Puzzles of Social Cognition." *NeuroImage* 28: 746–756.

Mele, A.R., and J. Shepherd. 2013. "Situationism and Agency." *Journal of Practical Ethics* 1 (1): 62–83.

Milgram, S. 1963. "Behavioral Study of Obedience." *The Journal of Abnormal and Social Psychology* 67 (4): 371–378.

Milgram, S. 1974. *Obedience to Authority: An Experimental View.* New York: Harper & Row.

Miller, C.B. 2013. *Moral Character: An Empirical Theory*. New York: Oxford University Press.

Nahmias, E. 2007. "Autonomous Agency and Social Psychology." In *Cartographies of the Mind: Philosophy and Psychology in Intersection*, eds. M. Marraffa, M. De Caro, F. Ferretti, 169–185. Berlin: Springer.

Nelkin, D.K. 2005. "Freedom, Responsibility and the Challenge of Situationism." *Midwest Studies in Philosophy* 29 (1): 181–206.

Rigdon, M., K. Ishii, M. Watabe, and S. Kitayama. 2009. "Minimal Social Cues in the Dictator Game." *Journal of Economic Psychology* 30 (3): 358–367.

Smith, A. 2008. "Control, Responsibility, and Moral Assessment." *Philosophical Studies* 138 (3): 367–392.

Vargas, M. 2013. "Situationism and Moral Responsibilty: Free Will in Fragments." In *Decomposing the Will*, eds. J. Kiverstein, A. Clark, T. Vierkant, 325–349. New York: Oxford University Press.

Wolf, S. 1990. *Freedom within Reason.* New York: Oxford University Press.

Journal of Cognition and Neuroethics

Physicalism and the Privacy of Conscious Experience

Miklós Márton
Eötvös Loránd University Budapest
Center for Theory of Law and Society

János Tőzsér
Research Centre for the Humanities
Hungarian Academy of Sciences

Biographies

Miklós Márton (Phd) is currently assistant professor at Eötvös Loránd University Budapest, Center for Theory of Law and Society. His research interest includes the philosophy of mind and philosophy of language.

János Tőzsér (PhD) is currently senior researcher at the Research Centre for the Humanities, Hungarian Academy of Sciences. His research interest includes the philosophy of mind and metaphilosophy.

Acknowledgements

We are grateful to Gergely Ambrus, Katalin Farkas, Gábor Forrai, Zoltán Jakab, Dávid Márk Kovács for their helpful comments on an earlier version of this paper, and to the participiants of the "Consciousness Conference" in Flint October 2015 for their inspirative comments. The research leading to this paper was supported by OTKA (Hungarian Scientific Research Fund), grant no. K109638, and by Bolyai Research Scholarship, grant no. BO/00028/13/2.

Publication Details

Journal of Cognition and Neuroethics (ISSN: 2166-5087). April, 2016. Volume 4, Issue 1.

Citation

Márton, Miklós, and János Tőzsér. 2016. "Physicalism and the Privacy of Conscious Experience." *Journal of Cognition and Neuroethics* 4 (1): 73–88.

Physicalism and the Privacy of Conscious Experience

Miklós Márton and János Tőzsér

Abstract
The aim of the paper is to show that the privacy of conscious experience is inconsistent with any kind of physicalism. That is, if you are a physicalist, then you have to deny that more than one subject cannot undergo the very same conscious experience. In the first part of the paper we define the concepts of privacy and physicalism. In the second part we delineate two thought experiments in which two subjects undergo the same kind of conscious experience in such a way that all the physical processes responsible for their experiences are numerically the same. Based on the thought experiments and their interpretations we present our argument for the inconsistency of the privacy of experience with physicalism in the third part of the paper. In the final part we defend our argumentation against some objections.

Keywords
Privacy of Conscious Experiences, Subjectivity, Physicalism, Property Dualism, Modularity of the Brain

Introduction

In this paper we would like to show that the privacy of conscious experience is inconsistent with any kind of physicalism. However, we do not conclude from this that physicalism is mistaken, we merely generate a dilemma. On the one hand, if one is a physicalist, then one has to deny our common sense conviction that only one subject can have a specific conscious mental state, that is, more than one subject cannot undergo the very same conscious experience. On the other hand, if one does hold this common sense conviction, one has to accept substance dualism and claim that conscious experiences are modifications of the immaterial mind.

Our paper divides into four parts. In the first part we define the concepts of privacy and physicalism and we formulate our thesis. In the second part we delineate two thought experiments and interpret these. In the third part we present our argument for the inconsistency of the privacy of experience with physicalism. In the final part we defend our argumentation against some objections.

1. The thesis

In order to formulate the thesis, we need to clarify the two concepts in question, so we have to define privacy and physicalism shortly.

There are two different senses of the concept of privacy. In one sense of the term, conscious experiences always belong to a subject. As Gottlob Frege puts it in his 'Thought':

> It seems absurd to us that a pain, a mood, a wish should rove about the world without a bearer, independently. An experience is impossible without an experient. The inner world presupposes the person whose inner world it is. […] [I]deas need a bearer. (1918/1956: 299.)

In other words: conscious experiences always need an owner for their very existence. They cannot exist in their own right, that is, without the subject. To wit: for every experience *e* there is at least one subject *S* who has *e* and *e* cannot exist without *S* having it. This kind of necessary ownership constitutes the first kind of privacy or subjectivity of conscious experiences.

However, according to some philosophers (for example Michael Tye or Ronald de Sousa) there is nothing extraordinary about the privacy of conscious experience in this sense, since it is not just her pains, fears and anxieties which belong necessary to a subject *S*, but her laughter, walk and state of health, as well. What is more, the falling of a stone belongs necessary to the stone in question. If this is true, then it will show that this conception of privacy has nothing to do with the 'mental' *per se*. Conscious experiences or occurrent states are events that happen, just as the above examples. So, according to these philosophers, they are private or subjective entities just because they are occurrent states of the owner, not because they constitute some special kind of entities (see e. g.: Tye 1995, 84-92; de Sousa 2002).

Nevertheless, we rather focus on the other sense of the term, since we think this second sense of the concept of privacy plays a more essential role in the common sense conception of conscious experience. In this sense of the concept, every conscious experience can belong only to one subject. For example: Mary's pain can be felt only by Mary and Juliette cannot feel it, or vica versa, Juliette's pain can be experienced only by Juliette and Mary cannot experience it. As Frege wrote it a few paragraphs later:

> It is so much of the essence of each of my ideas to be the content of my consciousness, that every idea of another person is, just as such, distinct from mine. […] No other person has my idea but many people

> can see the same thing. No other person has my pain. Someone can have sympathy for me but still my pain always belongs to me and his sympathy to him. He does not have my pain and I do not have his sympathy. […] [E]very idea has only one bearer, no two men have the same idea. (1918/1956: 300.)

To wit: for every experience *e* there is at most one subject *S* who has *e*. What does constitute this essential character of conscious experiences? The well-known answer is that only *S* can *directly experience* her conscious mental states; only *S* can undergo her particular conscious experience, and so, only *S* can access her conscious mental states in a way that nobody else can. In other words, whereas it can be true that anyone can have access to *S*'s pain in some way, only she can *feel* it. That is, *S* has a private path to it. This kind of private access constitutes the second kind of privacy or subjectivity of conscious experiences.

One clarification: The sentence "*S* has private access to her conscious mental states" does not say that there are two entities, namely *S* and her conscious mental state, whereas both of them exist in their own right and *S* has private access to the latter one. That is, private access is not a relation between two independent entities, since the entity to which the subject has this special kind of access, cannot exist without the subject having the access. In a certain sense, in the case of conscious experiences the very act and the result of it are just two aspects of the same thing. When we speak about private access we mean the first one, and when we speak about the entity to which the subject has private access we mean the second one.

Let us compare the two conceptions of privacy. While the first one states that the occurrence of every conscious mental state presupposes a subject as an owner of it; according to the second conception, only one subject can experience a conscious mental state directly. So, while the former one does not exclude the possibility that more than one subject could experience the very same conscious mental state, the latter one does exactly that.[1]

1. There are several other formulations of the common sense thesis of privacy. For example, one can speak about the necessary subjective quality or inalienability of conscious experiences, or the *esse est percipi* character of them. In our opinion, these phrasings are either more opaque or ambigious than the two above, or can be subsumed under them. For example, it seems obvious that to claim that conscious experiences have an esse est percipi character is nothing more than to claim that they cannot exist without a subject who experience them, which is exaclty the content of the 'necessary ownership' sense of the thesis.

Let us turn to the concept of physicalism. As it is well-known, the precise content of the physicalist thesis is difficult to explicate, but everyone agrees about the following. Physicalism is the metaphysical thesis that every phenomenon in our world is physical. Naturally the above-mentioned difficulty arises from the fact that we have no consensual answer to the question of what physical properties are.

Consequently, we do not wish to take a stand on the debate concerning the details of physicalism, so we will work with the following modest conception. There are fundamental physical phenomena (for example bozons, fermions and spin or charge), and *all* other phenomena *depend* on them for their existence.

This dependence could be ontological. This means that the fundamental physical entities or some configuration of them bring about all the others with metaphysical necessity. In the usual phrasing: there is no possible world in which all facts about these fundamental physical entities hold, but some facts of the actual world do not.

According to the orthodoxy, this ontological dependence of mental phenomena on the physical can be understood in three ways. (1) Mental properties are identical to physical (neurophysical) properties. As the good old (and empirically false) example says: pain = C-fiber firing. (2) Although mental properties are not identical to physical ones, there obtains a necessary supervenience or constitution relationship between them. In this conception, mental properties depend on physical ones in the sense that metaphysically there can be no difference in the former ones without a difference in the latter ones. In other words: if one determines a neurophysical entity together with its properties then one *eo ipso* determines the mental entity supervene on it together with its properties. (3) Mental properties necesseraly supervene on or are constituted by not merely some neurophysical properties, but also by further relevant facts of the external world.

In a strict sense only these three conceptions should be called physicalism. However, there is another form of dependence that can be found in the theory of property dualism. Here is a standard formulation of the main thesis of this view:

> [C]onscious experience involves properties of an individual that are not entailed by the physical properties of that individual […]. Consciousness is a *feature* of the world over and above the physical features of the world. This is not to say it is a separate „substance" […] All we know is that there are properties of individuals in this world — the phenomenal properties — that are ontologically independent of physical properties. (Chalmers 1996: 125, italics in the original)

As we can see from the quotation, property dualism is also a substance-monist theory which differs from strict sense physicalism in the sense that this theory does not commit itself to the ontological dependence of mental properties on the physical ones. Some kind of dependence nevertheless obtains between the two kinds of properties in this conception as well, namely there must be contingent psychophysical laws which connect them. As Dean Zimmerman says: „[t]here would have to be laws governing causal relations between microphysical events and emergent mental properties, laws that are sensitive to differences in microphysical duplicates […]" (2003: 506.). In a word, the property-dualist denies the ontological dependence, but states nomic dependence instead (see also: Chalmers 1996: 240.).

In spite of the main difference between any kind of strict physicalism and property dualism, they agree on a crucial point. Namely both claim that physical phenomena determine the mental ones, that is, mental properties depend on the physical ones. Therefore, they cannot allow the following: (1) there are mental states that do not connect to any physical entities at all (such as Descartes' clear acts of thinking); (2) there are mental states which connect to physical entities merely randomly. From this point of view, the difference of the two theories consists merely in the fact that in property dualism the dependence of mental properties from physical ones is assured by some nomological relation rather than a metaphysical one.

We are finally in a position to state our thesis. The common sense conviction that a subject has private access to her conscious mental states, that is, for every experience *e* there is at most one subject *S* who can undergo *e*, is inconsistent with any theory according to which mental phenomena depend on physical ones in one of the above senses. Consequently, the privacy of conscious experiences is inconsistent with all three versions of strict physicalism and property dualism as well.

2. Two thought experiments and their interpretation

Let us imagine that the parts of Mary's brain which are responsible for the pain in her lower back are damaged, so Mary cannot feel this kind of pain when she has lumbago. And let us also imagine that Mary's nerves are wired across to a healthy person's brain. Let us call this person Juliette.

The wiring works in the following way: when Mary's nerves have been pinched in her lower back the neural information arrives from her waist to the parts of Juliette's brain which are responsible for lower back pain. Then, the information flows further to those parts of both Juliette's and Mary's brains which are, in the case of healthy people,

directly connected to the parts responsible for the pain in question. In this situation Mary feels her lower back pain *through* Juliette's brain. (Of course, poor Juliette would also feel lower back pain, since her appropriate brain-parts would be active.)

Or imagine that in the future, technology of neurosurgery can produce a device (an implant) which is suitable to satisfy the function of the damaged brain parts. So, imagine that there will be such an implant which is wired into Mary's brain so that this device will be responsible for Mary's ability to feel lower back pain when her nerves would have been pinched in her waist. However, Mary and Juliette, who suffer from the same condition, will get a common implant which is wired into the both of their brains. From here, the story is similar to the one told above: when Mary's nerves have been pinched in her lower back, the neural information first travels to the common implant, and then flows further to those parts of both Juliette's and Mary's brains which are, in the case of healthy people, directly connected to the parts responsible for the pain in question. In this situation both Mary and Juliette feel lower back pain through the implant.

We have created these thought experiments in such a way that in both of them each physical entity responsible for Mary's and Juliette's pain is the same. One and the same physical entities (the brain parts, the implant, the wiring, etc.) are responsible directly for their conscious experiences.

The empirical plausibility of this claim hangs on the modular make-up of the human brain. According to this (oversimplified) modularity thesis, when a conscious experience occurs, only a certain part or parts of the brain and their connections are responsible for it. Or, in a reverse formulation: there are parts of the brain the activities and interconnections of which are not necessary conditions of the occurrence of a certain conscious experience (though they can be necessary for the occurrences of other kinds of experiences). In our case this means that we have to suppose merely that the other, uncommon parts of the two subjects' brains do not play any role in the occurrence of the conscious experience in question.[2]

However, one can say that the conception of the modular brain in itself does not support our interpretation of the thought experiments, namely that in the case of Mary and Juliette every relevant physical factor is common. While it can be true that the local

2. The modularity thesis can be supported by considerations concerning the possibility of evolutionary psychological explanations of mental functions. As Leda Cosmides és John Tooby write: "[…] natural selection will ensure that the brain is composed of many different programs, many (or all) of which will be specialized for solving their own corresponding adaptive problems. That is, the evolutionary process will not produce a predominantly general-purpose, equipotential, domain-general architecture". (Tooby – Cosmides 2005: 17)

neural or "implantic" bases of their conscious experiences are the same, these common bases have different connections. In the situations described above, the same physical entity is connected to two different brains, therefore their appropriate relationships are quite different. The modularity thesis actually claims that it is not only particular parts of the brain, but also their appropriate relationships to other brain-parts which are necessary for the occurrence of some conscious experience. Consequently, since the latter differs in Mary' and Juliette's cases, it is not true that all the relevant physical factors are common.

We think this possible objection misses its target. Remember for example the famous case of Phineas Gage (see e. g.: Damasio 1994: ch. 1). As it is well-known his brain's left frontal lobe was seriously damaged, and that injury had strange effects on his personality and behavior for the rest of his life. What is the moral of Gage's case? On the one hand, even such complex properties of a person as his personality or behavioral patterns can be associated to particular parts of the brain and their interconnections. It is plausible to suppose that this is also true in the case of a much more simple particular experience. On the other hand, and this is more important, in Gage's case the most remarkable fact is that his basic and several complex mental abilities remained intact after the injury. For example, according to the testimonies, his basic cognitive, linguistic and practical abilities survived the brain-damage to a great extent. One have to infer from this that the seriously damaged parts of his brain and its connections were *not necessary* for these abilities to work. In other words, Cage's case shows that a mental ability can survive the loss of some parts of the neural network, therefore the latter and its connections to other parts were not responsible for this ability. It is not the whole brain with its extremely complex neural interconnections which serves as the physical basis of a mental state. Consequently, it must be possible that two subjects share the relevant physical bases (brain parts plus interconnections) and individually possess only those brain-parts and connections (wirings) which are not necessary for the mental state in question. The above thought experiments show exactly this arrangement, so our interpretation seems to be empirically plausible in light of modularity.[3]

3. There may be a further worry about the correctness of our interpretation if there is some part or parts of the human brain which are necessary for all kinds of conscious experiences, even for all kinds of mental functions. If such a central processing unit really exists (and it is questionable) then it will be more difficult to conceive that each parts of the brain that are responsible for the pain in question are common, since in this case even this central universal parts must be shared by the two subjects, so we have to conceive them as totally incapable of any conscious experiences, or even any mental functions before the operation. However, we think that this possibility has no serious theoretical impact for our argumentation, though we acknowledge that it makes our interpretation empirically less plausible.

In sum, we interpret the thought experiments as follows. It is true in both fictive cases that (1) the *two* subjects equally *feel* pain and feel this pain as coming from her lower back; and (2) *all* the physical processes responsible for the experiences are the *same*. (1) seems phenomenologically evident in light of the situations described in the thought experiments, and (2) seems obvious, if we commit ourselves to the modularity of the brain. Of course, although our example of conscious experiences in these thought experiments was a certain kind of pain, you can substitute it for any other kind of conscious mental states (e. g.: perceptual experiences, moods, other bodily sensations, etc.).[4]

3. Arguments for inconsistency

As we defined in the first section, the privacy of conscious experiences in the second, more interesting sense is the thesis that every experience can be directly experienced or felt only by one subject, and not more. It follows from this thesis that if two subjects have some conscious experiences, as in the above cases of Mary and Juliette, these experiences are numerically different. There are two conscious pain-experiences, rather than Mary and Juliette feeling literally the same pain.

Let us see, why this common sense conviction is incompatible with any kind of physicalism, including property dualism. Maybe, the result will seem strange to some theorists since it is a widely held view in the literature that the question of privacy and that of physicalism are conceptually independent ones (see e. g.: Farkas 2008b: 15). Our argument against this view is, in a certain sense, quite simple: since every mentioned theory is committed to some kind of dependence of mental properties on some physical phenomena, and because all the relevant physical factors are the same for the two subjects' experiences, they depend on the same physical basis and therefore cannot differ from each other. Let us see the details of the argumentation.

(i) As for the case of identity theory, it seems obvious that this kind of physicalism is incompatible with the privacy of conscious experiences. The situation is this: if two properties, a mental and a physical one, are identical, then any instances of them are also

4. We think that all this is not just empty philosophical phantasy. Consider the case of craniopagus conjoined twins. They are joined at their head and so have some smaller or larger brain parts in common. One of the most famous cases is that of Krista and Tatiana Hogan. According to the medical reports, the tickling of one of them triggers laughter in the other, or one of them stops crying when somebody puts a teat into the other's mouth (see: Dominus 2011). In our opinion this situation is similar to the fictive cases delineated in our thought experiments: two subjects have supposedly similar conscious experiences and one and the same physical (neurophysical) processes are responsible for them.

identical. For example, any instances of pain are identical to an instance of C-fiber firing. Now, Mary's and Juliette's experiences, as instances of a certain type of pain-experience, are equally identical to the common physical basis responsible for them. It follows from this that the two subjects' experiences will be also identical to each other, regarding the transitive nature of the identity-relation. So, they feel numerically the same pain. In other words: since there are no physical differences between the facts relevant for their experiences, one can say that they undergo the same experience. Consequently, the supposition of the obtaining of the identity-relation is conceptually inconsistent with the privacy thesis.

(ii) As for the case of supervenience or constitution theory, it is important to see that these views are also committed to the identity of instances of mental and physical properties. A particular conscious experience as an instance of a certain type of mental property is identical to a particular neurophysical state or event in the subject's brain. Therefore, in the cases of particular mental events such as a conscious experience, it does not matter, whether one is a supporter of reductive or non-reductive physicalism, because both theories accept the thesis concerning the identity of property instances. The only difference between them lies in the fact that non-reductive theories allow the possibility that different instances of the same mental property can be identical to (realized by, supervene on, etc.) different instances of different physical properties. Now, since the cases in our thought experiments are about particular experiences, such a supervenience physicalist have to think that an identity relation obtains between Mary's and Juliette's pain-experiences on the one hand and the appropriate common physical basis on the other hand. (This latter could be some neurophysiological process of the relevant brain parts of one of the subjects, or some state or process of the common implant.) Consequently, the situation in the case of these kinds of physicalism is the same as it was in the identity theory. Given the transitivity of the identity relation, the two subjects undergo numerically the same experience, therefore non-reductive physicalist theories are also inconsistent with the privacy thesis.

(iii) There are other physicalists, who think that although particular conscious mental property instances depend on some physical entities, these entities are constituted not merely by inner neurophysical states or processes but by some further physical factors of the environment, too. Such an externalist physicalist can therefore claim that Mary's and Juliette's pain experiences will differ due to these factors.

In order to examine this possibility, we have to distinguish between two kinds of externalist approach. The first one is usually called 'phenomenal externalism', and the essence of it can be summerized in the claim that the phenomenal quality which is,

at least partly, constitutes the conscious experience in question, can be found in the experienced object itself, not in the subject's state of experience (see e. g.: Dretske 1996, Fish 2009). The second one is usually called 'content externalism', and the main thesis of it consist in the famous claim that the content of a mental state, which is also a constitutive part of it, is not in the subject's head.

As for the former one, we think it is easy to see that this theory does not threaten our thesis. Such phenomenal externalist physicalism should include the objects of experiences and their qualitative properties in the physical factors on which conscious mental property instances are supposed to depend. And it is plausible to suppose that the object of the experience is nothing more than the causal starting point of the sensational process. However, in the case of Mary and Juliette, these factors are all common. There is only one pinching of a nerve, which serves as the common causal starting point of both subjects' pain-experience. Or, if we would take an example of sensory experiences rather than pain in the thought experiments above, then there will be only one object that serves as the causal starting point of their sense impressions. Consequently, the situation is the same as in the case of the internalist physicalist: all the physical entities to which the subjects' conscious experiences are identical are common; therefore, they are identical to each other, too.

The situation is a little bit more complicated in the case of content externalism. There surely can be some facts in the context of the occurrences of conscious experiences in question which differ from each other. To mention just the most obvious one: the content of Mary's and Juliette's pain-experience is the same in a sense, that is, both Mary and Juliette experience that their lower back hurts. However, these contents are, at least for the content externalist, different in another sense: the content of Mary's experience is that Mary's lower back hurts and the content of Juliette' experience is that Juliette's lower back hurts. So the two contents differ, which makes their experiences also different.

We think this consideration has nothing to do with the concept of privacy we are interested in. As it was stated in the first section, this concept is connected to the notion of access, and makes conscious experiences private insofar as the subject's access to them is private. If we keep this in mind, it will become obvious that the difference between the two subjects' experiental content is a difference to which the subjects have absolutely no access. The concept of the content to which we have access is the concept of narrow content, and narrow content is, in turn, the content on which the external factors of the context have no impact. For example, Mary and Juliette have the same narrow experiental content in the above situation, namely: "my lower back hurts". As it is well-known, all external considerations are concerned with the broad content, but this very

broad content (if there is any, see: Farkas 2008a, Pitt 2013) does not make the conscious mental states private.

(iv) If we are right, strict physicalism is incompatible with the private nature of conscious experiences due to the simple reason that all forms of strict physicalism are committed to the identity of conscious mental property instances and the appropriate physical entities. However, we think that property dualism is also incompatible with the privacy thesis.

What should a property dualist say about the case of Mary and Juliette? Perhaps the following: There are two numerically different conscious experiences present in these situations, that of Mary's and that of Juliette's, because there are two numerically different phenomenal properties which have emerged from the common physical causal basis. Nevertheless, these conscious experiences as phenomenal properties of some physical phenomena are connected by some contingent psychophysical laws to it, in our case to some physical properties of the common brainparts or the implant.

Given the above, we think the property dualist has to commit herself to the claim that Mary's and Juliette's experiences are qualitatively alike. They feel their lower back hurting in equally the same way. Although psychological laws are contingent ones according to the property dualist, there could not be two different laws in the same world, i. e. two psychophysical laws by which two different kinds of phenomenal properties emerge from the same physical ones. In this respect there is no difference between nomic and ontological dependence. In sum, since Mary and Juliette live in the same world and all the relevant physical bases of their experiences are the same, their experiences are of exactly the same kind.

So, the property dualist opponent has only one thing to say, namely that although Mary and Juliette have qualitatively identical conscious experiences, these experiences are numerically different. The two subjects' conscious mental states differ from each other merely in a numerical sense. It seems to us that this idea is a rather implausible one. If the defender of property dualism claims that the conscious experiences of Mary and Juliette are merely numerically different while qualitatively identical, then she will have to allow that the same could be true in the case of one single subject, for example you. She has to allow that when the nerve in question pinches in your waist, two qualitatively identical but numerically different phenomenal properties will emerge; or, in simpler words; you will have two qualitatively identical but numerically different conscious experiences, so you will feel two qualitatively exactly the same but numerically different pains. There is no theoretical difference between your case and that of Mary's and Juliette's, so if

someone allows the possibility in question in the latter case, then she will also have to allow it in the former one.

However, it is not the end of the story yet. If one allows that you have two exactly alike pain-experiences, then one will have to allow that you have more, say 342 or 234 or 454627 merely numerically different, but in all other respects exactly alike conscious experienes. In a situation like this you could not tell the pain experiences apart, because they are qualitatively identical and so there is nothing by which you could distinguish one from the other.

We do not assert that the obtaining of such a situation is logically or metaphysically impossible. It is without any doubt metaphysically possible for a subject to have several indistinguishable and merely numerically different pain-experiences at the same time.[5] It is rather implausible for the reasons of phenomenology and theoretical parsimony. Why should we include several qualitatively identical but numerically different conscious experiences or phenomenal properties in our ontology if this move is not supported either by any consideration regarding the phenomenology of our mental life or by any theoretical benefit? Consequently, if we do not want to commit ourselves to such implausible claims then we have to acknowledge that even from a property dualist point of view, Mary and Juliette undergo numerically the same conscious experiences. In a word: it is not just strict physicalism, but also the other substance materialist theory, that is property dualism, which is inconsistent with the privacy of conscious experiences.

If our argument succeeds in supporting the incompatibility thesis, then it has two important consequences. Firstly, it shows that from a physicalist point of view, the private nature of conscious experiences is not a conceptual truth, but only a contingent feature of our physical make-up. We, normal human beings are built in such a way that the physical bases of our conscious mental states do not extend over our bodies. The neural networks which are responsible for conscious experiences are usually not connected to any other fellow's brain, and this is the reason why at most one subject can undergo a particular experience.

Secondly, from an anti-physicalist point of view, the incompatibility of privacy and physicalism can serve as a possible starting point of a new kind of argument against

5. As far as we can judge it, one needs to argue for the metaphysical impossibility of this situation in the following way: we have such special access to our conscious experiences that excludes any error in the individuation of them. That is, if it appears to us that a conscious experience *A* is identical to another one, namely *B*, then they will be necessary identical. The appearance of identity implies the identity of appearances, so to say. However, this kind of argumentation does not work. To see this, one only has to think about the examples of the phenomenal sorites problem.

the latter. The hitherto presented and much debated arguments usually allude to the supposed subjective nature of conscious experiences. They regularly emphasize that the very existence of conscious experiences supposes a subject with a special viewpoint who undergoes such experiences, and this essential subjectivity cannot be explained by any physicalist theories. In contrast to this, our argument alludes to the other sense of privacy, namely that more than one subject cannot undergo the same experience, and claims its incompatibility not just with physicalism in the strict sense, but with property dualism as well.

4. Some downright objections and replies

As far as we can see it, our argument can provoke some *prima facie* plausible or downright objections. We consider the following three of these.

(i) One can say that in order to individuate mental states, we have to allude to the whole stream of consciousness of the subjects. If this is true, our interpretation of the thought experiments will be false, because Mary's and Juliette's pains are parts of different streams of consciousness. So they cannot be the same.

Reply: nothing exludes the possibility that a part of a stream of consciousness be also a part of another one. The fact that a conscious experience essentially belongs to some subject's stream of consciousness does not imply that it is conceptually inconsistent for it to belong to more than one stream. In other words, streams of consciousness can be shared with each other, and we think that in the cases of Mary and Juliette the situation is exactly that.

(ii) Probably, Mary's earlier mental life was quite different from Juliette's. Perhaps Mary is an elite soldier who was trained to tolerate heavy pains, so her pain threshold was raised to a very high level. Therefore, even if her appropriate brain part happened to be damaged, and her brain has been wired to Juliette's, who is an ordinary person, in the way described in the first thought experiment, her pain experience will be much less intensive than Juliette's. Naturally, since the intensity of a pain belongs to the phenomenal features of this experience, Mary undergoes a phenomenologically and so numerically different conscious experience.

Reply: we suppose that a physicalist cannot accept the assumption that the level of Mary's pain threshold remains the same after the operation. If this was true, then this mental ability would not depend on any particular neurophysiological state or process. Therefore, if the objector adheres to the story above, she will have to deny physicalism.

Consequently, this objection is not directed against our inconsistency thesis but rather acknowledges it.

(iii) Mary and Juliette cannot undergo the same pain-experience as Mary feels her pain in her$_1$ lower back, Juliette feels pain in her$_2$ lower back, that is, Mary feels her pain as a state of her own body and the same is true for Juliette. This difference can be explained in intentional terms. Mary's and Juliette's pains have different intentional structures, namely their intentional objects are different. Mary's experience directs to Mary's waist; Juliette's experience directs to Juliette's waist.

Reply: The consideration behind this objection is very much like the one we handled in the argumentation for the inconsistency of privacy with the externalist version of physicalism. As we explained there, the concrete particular factors of the context of a conscious experience belong to the broad content of the mental state in question. However, broad content is actually the content to which the subjects do not have any direct access, so one cannot allude to it in order to support the private nature of conscious experiences. All that can contribute to this private nature rather belongs to the narrow content.

We illustrated this by the example of the particular experiencing subjects. The concrete identity of the experiencing subject does not belong to the narrow content; therefore, if two subjects feel the same kind of pain, then the narrow content of their experience will be the same regarding the subject of the pain. Both feel the pain as *their own*. The same is true for the objects of conscious experiences: the identity of the particular object belongs to the broad content and the narrow content will be the same: both subjects feel their lower back pain as belonging to *their own* lower backs. So the difference of external intentional objects does not make any difference in the conscious experiences themselves as the subjects undergo them.

References

Chalmers, David. 1996. *The Conscious Mind*. Oxford: Oxford University Press.

Damasio, Antonio R. 1994. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Putnam Publishing.

de Sousa, Ronald B. 2002. "Twelve Varieties of Subjectivity." In *Language, Knowledge and Representation*, edited by Jesus L. Larrazabal and Luis A Pérez Miranda, 147–164. Dordrecht: Kluwer.

Dominus, Susan. 2011. "Could Conjoined Twins Share a Mind?" *New York Times*, May 25. http://www.nytimes.com/2011/05/29/magazine/could-conjoined-twins-share-a-mind.html.

Dretske, Fred. 1996. "Phenomenal Externalism." In *Philosophical Issues, 7: Perception*, edited by E. Villanueva. Atascadero, CA: Ridgeview Publishing.

Farkas, Katalin. 2008a. "Phenomenal intentionality without compromise." *The Monist* 91 (2): 273–93.

Farkas, Katalin. 2008b. *The Subject's Point of View*. Oxford: Oxford University Press.

Fish, William. 2009. *Perception, Hallucination, and Illusion*. Oxford: Oxford University Press.

Frege, Gottlob. (1918) 2000. "The Thought: A Logical Inquiry." *Mind* Vol. LXV (284): 289–311.

Pitt, David. 2013. "Indexical Thought." In *Phenomenal Intentionality*, edited by Uriah Kriegel, 49–70. Oxford: Oxford University Press.

Tooby, John, and Leda Cosmides. 2005. "Conceptual Foundation of Evolutionary Psychology." In *The Handbook of Evolutionary Psychology*, edited by David M. Buss, 5–67. Hoboken, NJ: Wiley.

Zimmerman, Dean W. 2003. "Material People." In *The Oxford Handbook of Metaphysics*, edited by Michael J. Loux and Dean W. Zimmerman, 491–526. Oxford: Oxford University Press.

Journal of Cognition and Neuroethics

Pre-Conscious Noise

Bradley Seebach and Eric Kraemer
University of Wisconsin-La Crosse

Biographies

Brad Seebach is an Associate Professor of Biology at the University of Wisconsin-La Crosse. Seebach received his A.B. in English Literature from Cornell College and his Ph.D. in Neural Science from Brown University. He is a self-described developmental neuroscientist. Significant influences on his thinking come from years spent in the study of chemical engineering, linguistics, and computational neuroscience, in addition to many years of research and teaching in neurophysiology. His scholarship focuses on the development of neural systems. For many years, he has more specifically focused on the development of central pattern generator circuitry related to mammalian locomotion—one place where development may require unsupervised learning within a network of neurons, and yet be relatively accessible to analysis using the tools of an electrophysiologist.

Eric Kraemer is Professor of Philosophy at the University of Wisconsin-La Crosse. Kraemer received his A.B. in Philosophy from Yale University and his Ph.D. in Philosophy from Brown University. The overall goal of his current research program is to discover and defend a comprehensive naturalistic perspective by showing how the naturalist can rely not only on support from the advancing scientific program but can also successfully appeal to strategies developed and utilized by thinkers from other philosophical viewpoints, including the super-naturalist perspective. A convinced mind-body physicalist, Kraemer now views the newly emerging Hidden Properties Physicalism approach as providing the most helpful perspective for solving the mind-body problem, by explicitly supporting making connections between developments in the philosophy of mind and possibilities that current neurophysiology is making known.

Publication Details

Journal of Cognition and Neuroethics (ISSN: 2166-5087). April, 2016. Volume 4, Issue 1.

Citation

Seebach, Bradley, and Eric Kraemer. 2016. "Pre-Conscious Noise." *Journal of Cognition and Neuroethics* 4 (1): 89–112.

Pre-Conscious Noise

Bradley Seebach and Eric Kraemer

Abstract

Philosophers and neuroscientists take such different approaches to questions surrounding the existence or definition of conscious states that they often dismiss each others' viewpoints as irrelevant or meaningless. Yet there is a historically rich, informative interaction between philosophers and neuroscientists that we believe should be kept within the purview of each field. The traditions of philosophers who have considered the problem of consciousness, and the traditions of neuroscientists who have considered the problem of consciousness are examined in this discussion, in order to see how fundamental differences in the approach to the problem may produce a lack of obvious, common ground for discussion.

Reductive materialism and the dual-properties theory are presented as representative approaches employed by contemporary philosophers of mind, as they consider the problem of consciousness. On the side of the neuroscientists, we examine approaches to the same problem from perspectives used in neurophysiology and in computational neuroscience.

A key understanding emerges in the discussion, which is that introspective analysis of the mind cannot account for the emergence of apparently novel thought, and that this corresponds to the neurophysiologists' bane, which is the presence of apparently spontaneous neuronal activity, uncorrelated with identifiable patterns of behavior within a neuronal circuit, system, or whole animal. This "noisy", apparently spontaneous activity becomes of greater interest when viewed as a common weakness for both the philosophers of mind, for whom it represents the nothingness from which conscious states emerge, and for the neurophysiologists, for whom no sufficient tools exist to find correlated neuronal activity in a great many neurons or circuits in the majority of studies. The goal of discussion becomes to identify whether there is an unidentified precursor of a conscious state buried in the noise ("pre-conscious noise"), and then to identify a plausible explanation for what such a precursor might (roughly) look like.

In order to understand whether there are meaningful patterns within the noise, it is necessary to go beyond tools that are currently available to physiologists and consider what other demands are placed on the brain during behavior. Few modern neuroscientists would claim that a conscious thought that is in progress drives all activity in the brain—there are likely many, many patterns of activity in progress in the brain, some of which may be vying for attention. What rises to consciousness in the next moment of time may lie submerged in a partial pattern of activity that is not yet fully established, but that may become more fully established as the result of a blend of anatomical structure resulting from genetic and experience-driven development and ongoing learning, immediate sensory input, and existing or recent activity patterns in the circuitry of the brain. These are described as proximate and ultimate causes for behavior.

The oscillatory activity of the brain that appears as a fundamental drive for activity patterns and which is visible throughout life in electroencephalogram (EEG) measurements, varying in frequency range in a daily pattern, is considered as a driving force for the rise of partial patterns of neuronal activity to strong, completed patterns that could be correlated with behavior at either a conscious or non-conscious level.

The last piece of the discussion involves the necessity of providing an example of a pattern of activity that might have the qualities of a partial pattern that becomes a strong, completed pattern. For this, the distributed, content-addressable memory of a Hopfield network model is used. A neuronal network that uses

a Hebbian-learning algorithm in an unsupervised manner is briefly examined, as it represents early learning for linguistically-relevant characteristics in a small network of mathematically simple, artificial neurons.

The resulting explanation, arising from principles of neurophysiology and computational neuroscience, is linked back into the philosophy of mind scholarship. "Hidden Natures Physicalism" offers a very interesting framework of consideration, within which our explanation of "conscious noise" or (perhaps more accurately) "pre-conscious noise" is found to have merit for explaining the essential characteristics of the mind-body problem that is at the heart of consciousness studies.

Keywords
Hidden Nature Physicalism, Computational Neuroscience, Content Addressable Memory, Resonant Patterns

Introduction

What is the problem of consciousness, and why is it so hard, both to get researchers outside of philosophy of mind to appreciate that there is a serious problem of consciousness, and, once clarified, to propose a program for how a solution might emerge? The solution proposed in this discussion is one that relates the disciplines of the authors of this paper, neurophysiology and philosophy. Philosophers, when they are thinking about the mind and mental phenomena, typically engage in first-person reflection, concentrate on the robust and varied qualities of the experiences they are having and then try to explain their natures and remarkable sensory and intentional features. Pains, itches and tingles, sensations of red, hallucinations and after-images, beliefs about non-existent beings, and desires for non-obtainable states-of-affairs all constitute typical subject matter for philosophers of mind contemplating consciousness. Neurophysiologists (and psychologists), on the other hand, engage in third-person observation, and, once they are confident with having established specific relations between types of mental states and events in the brain (or in the social context), then turn to postulating detailed physical (or social) mechanisms to explain how these states arise. For most of the past two decades, philosophers of mind have referred to the biologist's project as the "Easy Problem", while insisting that their own project is the "Hard Problem" (Chalmers 1995). How can the mechanisms of the neurophysiologist, wonderfully detailed though they are, account for the qualitative and intentional features of conscious experience? It is this lack of how a connection might be made between the two concerns that is often referred to as *the* "Explanatory Gap": exactly how is the itchiness of one's currently itching mosquito bite to be explained by the current structure and functioning of one's nervous system (Levine 1983)? But there is more to the story.

Reduction or Dualism?

Reductive materialism and the dual-properties theory are the two basic alternative approaches employed by contemporary philosophers of mind to account for consciousness. According to reductive materialists, all of reality is, at base, physical in nature, including conscious mental states, and all mental features, including consciousness, must somehow be reduced to, that is, completely explain in terms of, more basic physical features (Smart 1959, Armstrong 1968, Lewis 1966, Churchland, 1998, Hill 2009). This approach faces the problem just mentioned of elucidating how the basic entities and forces of physics and chemistry could give rise to the phenomenal or "felt" aspect of consciousness. This problem, therefore, seems better referred to as the Problem of the *Phenomenal* Explanatory Gap. This problem is neatly avoided by the dual properties theorist who postulates a special aspect of reality expressly to account for this felt gap between consciousness and the physical world.

Defenders of the Dual Properties Theory (DPT) come in three different varieties, the two most common versions of which are substance dualism and property dualism. The most important defender of the original version of substance dualism in the modern philosophical period was, of course, René Descartes (1648). Descartes claimed that the proper account of consciousness required postulating a second non-physical substance, a soul, in addition to the physical body. This version of the DPT has largely fallen out of fashion, being perceived as metaphysically non-parsimonious, and most current defenders of the DPT accept the simpler view that the features of consciousness, while non-physical, are nonetheless, features of a physical body. So, instead of there being two metaphysically different kinds of substances there is only one kind of substance with two metaphysically different kinds of properties. (See Nagel 1974, Campbell 1984, Jackson 1986, Levine, 1983, Chalmers 1995, Kim 2007.)

For the dual properties theorist, in order to account for the remarkable features of consciousness what is required is that we postulate, in addition to the basic materials of the physical sciences, a second, different aspect of reality to account for consciousness. The problem, of course, for defenders of the dual properties theory (DPT), is that they are equally challenged: they cannot explain [1] how this second aspect of reality arose, [2] how it appears to be caused by and causally influences physical reality, and [3] how this special aspect of reality actually creates consciousness. So, to be fair, defenders of the DPT need to admit that on their own view there remains an equally huge, if not larger "explanatory gap" between the two kinds of reality. Let us refer to this as the Problem of the *Metaphysical* Explanatory Gap: how can the physical world contain two different kinds of features, physical and non-physical?

There is also a third variant on the dual properties view, namely that of the non-reductive materialist. Proponents of this approach claim that while all of reality is ultimately physical, nevertheless there are some features of living beings, such as consciousness, which cannot be reduced to physicochemical properties (Cornman 1983, Searle 2004). Defenders of this approach face their own serious worries, namely [1] explaining how normal development from purely physical systems can give rise to two very different kinds of properties, (2) explaining how these two different kinds of properties are causally related to each other, and [3] explaining how physical features at one level, say the biological level, can themselves give rise to conscious experience at the psychological level. So again, there is an explanatory gap to be filled. Let us call this the *Level* Explanatory Gap: how can different, irreducible levels of physical reality arise? While it was once expected that functionalism combined with supervenience would provide the means to fill this gap, the current philosophical consensus seems is arguably that this approach has not delivered on its initial promise. (See Putnam 1999, Kim 2007; but compare Lewis 2004.)

From the perspective of scientific researchers who think that there may well be something of empirical value to be discovered from the scientific investigation of consciousness, however, all versions of the DPT are an unmitigated disaster. This is because defenders of all forms of the DPT place an impregnable metaphysical barrier between the philosophical and scientific explorations and explanations of the mind. Scientific researchers, on the other hand, continue to suggest that there must be further approaches that might be tried. And, some reductivist philosophers of mind would agree. In response we would like to suggest that the solutions of these two problems, the easy problem and the hard problem, might well go together. Why so? How is neuroscience research relevant to questions about the nature of consciousness?

The Relevance of Neuroscience

If we assume that humans are wholly biological beings, then at this point in the development of the history of science it seems that that neurophysiology, particularly when augmented with tools of computational neuroscience, may be well positioned to answer questions about how a human body can give rise to consciousness. Consciousness is a exceedingly difficult concept because it cannot be examined at a level that available scientific tools within any one discipline can manage. A synthesis of neurophysiology and computational neuroscience can, we believe, significantly advance our understanding. We will be investigating multiple causes for neuronal activity that appears to be correlated

with behavior. We will also examine neuronal activity that appears to be uncorrelated with behavior and has therefore often been referred to as *noise*, or as being of random nature. As several classes of causal agents may exist for any neuronal activity, we will make use of a conceptual framework for the study of causation within the biological sciences established by Ernst Mayr (1961) and developed over the last half century by many others.

Neurophysiology provides tools that can examine nervous system activity that exists in the immediate context of a behavioral decision. These tools range from the extracellular observations of electrical activity in nerves that became common in the 19th century, to 20th century innovations such as the electroencephalogram (often credited to Hans Berger circa 1924) intracellular recording of action potentials (Hodgkin and Huxley 1939), and activity in ensembles of neurons making use of improvements in intracellular calcium-concentration imaging documented by Tsien (1980) in combination with advances in optics and in the handling of live nervous tissue specimens. Each of these techniques captures activity at a particular anatomical scale and time base of analysis, and is capable of finding correlated regularity in the midst of tremendous "noise" at the associated scale. Yet the analysis at the smaller anatomical scales is likely blind to emergent patterns of activity that appear in the analysis of larger circuits, and the larger-scale analysis is blind to regional sub-patterns that are likely to be of critical importance to attention selection and for (either) conscious perception or triggering reflexive responses. Each has its own effective time scale, also. Neuroscientists are, generally, quite aware of these limitations in scope.

Computational neuroscience can augment neurophysiology due to its ability to incorporate aspects of so-called "ultimate" causes derived evolutionarily. These include anatomical structure and mathematical rules that govern the interactions between neurons within a circuit. Such interactions will incorporate an individual's historical experience due to patterned sensory and contextual input through processes that can be broadly referred to as "learning", and will build upon the genetically-determined capabilities of the nervous system. In computational neuroscience, the experimental design of the sensory inputs and context can substitute for real-world experience that drives developmental change. These genetic and historical causes interact with immediate context to drive immediate decision-making processes in neuronal circuits that are responsible for recall and behavior. Consciousness sometimes has been attributed to the presence of synchronous, iterative activity in neuronal circuits and systems that produces strong learning in development phases and may trigger awareness thereafter (Llinàs et

al. 1998). We will focus on the transition between (so-called) noise and stable patterns of activity that can be correlated easily with behavior.

The Conceivability Objection Considered

At this point some philosophers will object that all of these approaches to understanding consciousness are fundamentally flawed and for the same basic reason, viz. whatever discovery we claim to make regarding any such procedure that is then used to make the claim that the nature of consciousness is ultimately to be explained using that discovery we can then easily conceive of a creature who is just like the one upon which we are experimenting with the following difference: instead of being in a particular conscious state C^1 when the particular neurophysiological state or condition is present the creature either lacks consciousness altogether or is in a different conscious state, say C^2. From this thought experiment these theorists conclude that no neurophysiological account or explanation can fully explain consciousness (Nagel 1974, Campbell 1984, Jackson 1986, Chalmers 1995). But, we think that this argument, though often cited, is ill-considered. That there is yet no simple generally accepted neuroscience account yet of consciousness fails to prove that consciousness must be something other than neurophysiological activity. Consider this analogy. While physiologists before Harvey disagreed about how the blood circulated and could claim to refute their opponents' views with similar conceptual arguments, once Harvey established the correct pattern amidst the considerable physiological "noise" that persisted, all such contrary arguments fell by the wayside. We think that something similar is likely to happen with respect to consciousness, at some point.

A second useful analogy for the worry about consciousness is to be found in the debates between mechanists and vitalists shortly after the turn of the 20th century. The mechanists can be viewed as playing the role of the reductive materialists and the vitalists as playing the role of the dual properties theorists. The vitalists' (and later the organismalists') sole argument consisted in pointing to remarkable features in organic development which, it was claimed, just could not be explained by any purely mechanical process (Driesch 1908, Ritter 1919). Clearly, turn-of-the-20th-century mechanists had to be reductive mechanists, while their vitalist contemporaries were "dual property" vitalists. While the analogy may be comforting, in the meantime, how should one proceed?

What about "Noise"?

One promising answer counsels that we turn directly to "noise" itself. What is noise? We conceive of noise as a rich state of interaction in which activity patterns have not settled into resonance in the monitored anatomical region. Yet these suboptimal patterns of activity are continually influencing other, connected regions, offering momentary biases into patterned activity in those other regions. These momentary biases may combine in a manner that produces positive feedback, creating a resonant stable pattern, or they may fail to do so. So then, what does the noise represent? It represents multiple, potentially resonant patterns that are competing for attention or to trigger reflexive action. Also, for the neuroscientist, the noise is often what one tries desperately to filter out of one's results through either a hardware or software frequency filter or through post hoc statistical analysis, in order to reach a core pattern of some significance. One cannot easily publish uncorrelated data that cannot produce answers to useful questions!

Yet when we examine the nature of consciousness, the supposed unity of conscious experience, and the flow of attention from one focus to another, the noise may be precisely where we should be looking. In the period of time in which no recognizable, resonant pattern has yet been established, there is interaction occurring that involves evolutionary, developmental, and immediate causes for a behavioral decision. This interaction may produce strong influences that arise from each of these sources in a manner that masks a deterministic process. We believe it is therefore useful to consider causation more fully. And we need to use computational neuroscience in order to examine what might create "noise" while *en route* to stable, resonant patterns of activity.

Biological Causation, Including Evolution and Development

Mayr (1961) expands the Aristotelian view of causation into an explicitly "biological" framework, in which he describes proximate cause as things that a functional biologist might study. A neurophysiologist is our primary example of a functional biologist. Ultimate cause includes a broad variety of historical causes for a biological event, and includes the things that an evolutionary biologist might study. Mayr used an example of the proximate versus ultimate causes of a biological event such as the migration of a bird to illustrate his point. One of Mayr's great contributions was to provide an explicit distinction between these types of causal agents, and this has been an important contribution to the advancement of science (Beatty, 1994).

The apparent, attractive dichotomy between proximate and ultimate causes in biological science has been extended (perhaps appropriately) in a manner that ignores

cautions that Mayr expressed (Laland et al. 2011), particularly with regard to the "muddle" of development (Amundson 2005). This "overextension" makes it more difficult to place developmental considerations into the proper framework with proximate and ultimate causes. In the words of Laland et al. (2011), "Mayr's proximate/ultimate distinction has proven problematic because it builds on an incorrect view of development that fails to address the origin of characters and ignores the fact that proximate mechanisms contribute to the dynamics of selection." Causation includes feedback loops that cannot be adequately described by a strict separation between proximate and ultimate causes.

We find it necessary, in the consideration of the neuroscience of consciousness, to carefully consider proximate causes for a behavioral choice, ultimate causes for that behavioral choice, and how developmental constraints and influences shape each of these. This is, essentially, a fundamental viewpoint in evolutionary-developmental approaches to biological science (Amundson, 2005). Because development incorporates the lifetime of experience of an individual into the background context of immediate decisions, we choose to include development as an ultimate cause for behavior, recognizing that distinctive time boundaries between ultimate and proximate causes are, in some cases, lost due to learning processes being capable of producing rapid change in neuronal circuitry under some circumstances. We find this to be an acceptable confusion, as it would only be applicable in earlier stages of development for many regions of the brain, as primary sensory regions and many of the non-cortical regions of the brain lose their plasticity after a critical period of development passes.

Neurophysiology tools can describe several layers of proximate cause (Calvin 1998, 2004), depending on the level of analysis and the tool being used. Computational neuroscience tools make explicit the constraints of memory storage and retrieval, and are therefore a required part of the consideration for how development (and incorporation of memory as a shaping influence on behavior) shapes both proximal and ultimate cause.

Conscious decision-making in a traditional sense may be thought of as a cause of behavior that would override ultimate causes, allowing a human to choose to ignore those ultimate causes that might arise through evolutionary shaping of the human body and nervous system, and to also ignore, at will, the developmental shaping of the body and its nervous system that is accomplished in preceding years through integral processes of learning. A strong bias towards this view likely develops from the natural border of levels of analysis – consciousness is limited by attentional focus and cannot be divided easily into multiple streams that can keep track of the minute contributions made by individual cells or even systems of cells (an anatomical and functional size barrier), nor

can it be diverted from the behavioral time scale of seconds or tenths of seconds to analyze events occurring in only a few milliseconds or extending to minutes, hours, and days—these abilities are in the province of recorded language, which can bring context back to us in compressed or expanded time scales through the utilization of memory recall.

Consciousness and Oversimplification

Another strong bias towards the explanation that consciousness directs our activities comes from the evident creativity of conscious thought. However, consciously-directed creativity may be an incorrect explanation if there is lack of ability to understand underlying processes. Nate Silver (2012) describes a pattern for why humans consistently fail in necessary understanding for the accurate analysis of scientific, experimental results: a failure to eliminate personal viewpoint or bias from experimental design or analysis. We are, of course, blind to those things that we cannot see. Silver states that "we forget—or we willfully ignore—that our models are simplifications of the world." Biology is full of hidden patterns that have predictive power. Witness the rise of genetic explanations for human behavioral patterns, individual differences, and illness (Bargmann and Gilliam, 2013) that followed the discovery and developmental understanding of DNA. To which causal agents are we blind?

Calvin (2004) explicitly uses Mayr's proximate versus ultimate cause descriptions to describe layers of neurophysiological analysis from chemical-molecular events that determine a neuron's electrical state, to intercellular activity determined synaptically between two neurons, to a network of cooperating neurons, to the system of neurons and glial cells that govern a behavior. He compares this layering to an examination of the fibers that are stitched together in a pattern to make cloth, which then may be trimmed and organized to create an item of clothing. An example of a neuronal 'system' with such layering might be the visual system centers of the eye and brain, in which many neurons, made from similar components but exhibiting different morphologies and a variety of arrangements into networks, may together provide a sensory stream that can be integrated with a historical record of some kind maintained in memory, in order to produce an awareness of the visible world.

The Time Basis for Neural Circuits and Behavior

Less well-explored, perhaps, are the time bases for conscious decision-making, and the equally complex structuring of time. This is of great interest to us, because it offers

an opportunity to explore what happens between the molecular and neuronal time scale of activity (nanoseconds to milliseconds), the behavioral time scale of hundreds of milliseconds to tens of seconds, and how this relates to observable chemical and electrical patterns of activity within nervous circuits. This is the realm of behavioral "noise" that has been evident to several generations of neurophysiologists, and which is generally ignored by others who make use of what the field of neurophysiology has been able to show at the more understandable, behavioral time scale. On a time scale, the proximate cause of a behavior may be an established, correlative pattern of activity among a large ensemble of neurons that represents a predictive state, as established by neurophysiologists and neuroscientists using EEG, fMRI, or other tools and generally detectable with a minimum duration of several hundred milliseconds or more. Yet that pattern of activity arose through many cycles of activity on a smaller time scale.

Those very brief cycles of activity have been viewed as 'noise' by many, but are they noise? Or do they represent a rich state of interaction that would be predictive of the final behavioral outcome, if we were not simplifying our model of analysis on both the time scale and the anatomical scale? If they have no purpose, then why do they exist as a nearly universal subscale activity across all orders of complex organisms and in simpler pattern generating neuronal circuits, such as those that drive locomotor activity?

The general problem may be thought of in this way. A complex pattern of activity in a highly-interconnected network of neurons exists in the nervous system at all times during a human life. The failure of such patterns is a common definition of death. The basic patterns reflect an anatomical substrate that is strongly influenced by evolution of the human body, but is also strongly influenced by interaction between evolutionary mechanisms and a lifetime of contextually-dependent change in the interconnections, and indeed, in the number and type of cells within many anatomically-identifiable brain regions. These are identifiably some of the ultimate causes for a behavior. Immediate context (proximate cause) produces a barrage of fluctuating inputs that drive change in the complex pattern of activity in this interconnected network, though the pre-existing state of activity and the developmental and evolutionary shaping of the neuronal and glial cell networks make the influence of immediate context dissimilar from one person to the next (Finn et al. 2015). In fact, because of the ability of the nervous system to reshape its connections through learning, the immediate context, if it reoccurs, may produce a dissimilar effect on the same person the next time.

Content-Addressable Memory

Computational neuroscience has described how a system of common elements in an interconnected network can store and recall memory patterns in a way that overlaps, so that no single memory trace is held discretely, yet each can be recalled as a form of "content-addressable memory". We will use the primary example of a Hopfield network (Hopfield 1982).

To illustrate the idea of content-addressable memory and its possible use in a neuronal circuitry, we will follow two paths of explanation. The first is to explain a general method of approximation that uses iterative processes of calculation, common to some types of mathematics and engineering methods, and which we believe to be a good analogy for a function of the recurrent, looping neuronal circuits that produce iterative activity in the brain. Such circuits retransmit information between regions of the cortex and the thalamus (and also involving various basal nuclei). The second path of explanation is to provide an example of useful information that could be distributed across a network of interconnected neurons to create overlapping memory in synaptic connections. The specific example we will use is to represent syllabic speech. Such sensory information may be involved in both memory formation and recall processes through a loop between cortical and subcortical auditory centers of the brain that can sustain iterative, auditory and cognitive processes and produce stable activity patterns.

Iterative Methods and Recurrent Neuronal Circuits

Iterative methods of approximation have a long history in mathematics. Characterizations of different, iterative methods are credited variously to Newton, Gauss, and many other mathematicians, and scientists. They were extended into common use in nonlinear systems in the computational sciences through the work of David Young (e.g., Young 1950; Kincaid et al. 2010) and others. The general idea is to use an equation that describes known parameters (representing boundary values) of a problem, then insert into the equation a hypothesized, possible value x. Solving the equation using the hypothesized value returns a closer approximation to the real value of x. This new value can be substituted back into the same equation, and reiteration of the calculation will produce a new, even more accurate approximation to the real value of x. This gradual process of approximation is often referred to as a "relaxation" process when it is used to identify one or more stable states for a complex system, in which values may represented in two- or three-dimensional arrays, and for which the iterative mathematical techniques are often drawn from linear algebra.

We will use the term "relaxation" to describe the iterative approximation techniques that are used in computational models and that we believe are employed in the nervous system for purposes of memory recall and selection of behavioral responses. The term "relaxation" is derived from original conception of a reduction in chaos of a highly-variable system (in terms of energy) to a more coherent, stable energy state. In a Hopfield network, if part of a pattern that has been stored in content-addressable memory is presented as an input to the network (of artificial neurons), those neurons are excited in a manner that will return a closer approximation of the stored pattern. This closer approximation can be re-presented as an input to the network, and the following output should produce an even better representation of the stored pattern. Hence, when used in an iterative fashion, the recall process gradually improves the result as it "relaxes" to a coherent, pattern stored in memory, and the input and output of the network come to match one another more and more closely – the minimization of differences is thought of as representing a minimal, more "relaxed" energy condition as compared to the initial response state.

A system of 100 neurons with simple interconnections in a Hopfield network is estimated to be capable of storing without error up to 15 memory traces (Hopfield 1982) in which the state of each neuron is important to the correct learning and recall of each of the 15 memory traces. Storing additional memory traces produces "errors" in recall due to pattern interference that allows confusion of patterns. This type of error is actually of great interest to our discussion as well – for the confusion of one pattern with another may be deemed an error or it may be the basis of fruitful, creative processes. It is worth noting that the Hopfield network can produce an error in recall if the initial input either does not match any of the stored "memories" or if it is indeterminate between two or more stored memories.

Many connectionist models have been built upon the ideas represented in the Hopfield network combined with variations on the concept of Hebbian learning (Hebb 1949), and we briefly examine one such use.

Recurrent Neuronal Circuits and Oscillatory Activity

Iterative, synchronous activity in the mammalian brain has been identified in thalamic neurons that project to many regions of the cortex (Hunnicutt et al. 2014). Synchronous, oscillatory neural activity in the gamma-wave frequency range has been postulated to be critical to binding together neural centers that work together for behavioral purposes in development, consciousness and attentional modulation (Llinas et al. 1998, Miltner et

al. 1999) using mechanisms of Hebbian learning (Hebb 1949, Caporale and Dan 2008) and accretion of neuronal activity that shapes alpha-wave oscillatory activity (Bollimunta et al. 2011). For our purposes, it is useful to note that gamma-wave frequency range is 40-100 Hz, offering the possibility of many iterations of activity, each producing greater synchrony in neuronal activity to perhaps cross a threshold to produce behavior, within the time frame of several hundred milliseconds that is associated with even rapid, primed voluntary behaviors in response time experiments. Oscillatory brain wave activity in the slower beta-wave frequency range (12-40 Hz) is associated with awake, active mental behavior that corresponds with conscious thought, and may represent a more standard pace of "iterative approximation" that human brains could use for complex processes—in which current sensory conditions and internal processes might be matched with memory in order to produce recall and to choose an appropriate action.

In studies of "readiness potentials" (Libet et al. 1982, 1983) that precede voluntary activity, a gradual increase in synchronicity of neuronal activity is noted and reaching a threshold level of excitatory synchronicity seems to be associated with awareness (Mathewson et al., 2009). A useful, stochastic "accumulator model" that links readiness potentials (represented in EEG traces) with behavioral tasks has been published by Schurger et al. (2012).

The type of information that the nervous system needs to learn to work with, and to match during a process of recall, can be represented with a matrices of numbers that represent excitatory activity in synaptic connections that represent an orderly, topographic map of a type of information. Our example will use auditory information and basic assumptions about peripheral auditory processing used to create a model of prenatal speech perceptual development (Seebach et al.1994). Speech samples that were sufficient for the development of perceptual discrimination of elementary consonant-vowel syllables, differing only in the initial stop consonant's place of articulation, have acoustic energy intensities represented in different frequency bands in the range of human hearing, spread across a brief period of time (Figure 1A). Sounds such as these, presented repeatedly as through the apparatus of the auditory system to an interconnected group of artificial neurons (see Bienenstock et al. 1982) whose connections (representing synapses) change in a Hebbian learning process, will shape responsive patterns of those neurons (Figure 1B) in a manner that produces discriminative ability (Seebach et al.1994). Different syllables can be discriminated by the presence or absence of acoustic energy and energy transitions at specific times and frequency regions. The purpose of the particular study was to show that this type of discrimination could be learned by a neuronal circuit in an "unsupervised" manner—simply as a matter of experience, with no "teacher" or

confirmation of right or wrong responses—in other words, a learning process that could explain developmental changes that can occur very early, prior to the time of full, social interaction.

This is a very limited example of a memory process that could incorporate aspects of ultimate causes such as evolutionary shaping of the initial, anatomical network of neurons and the historical, developmental shaping of regions that become involved in more specific processing tasks. Iterative calculations representative of oscillatory brain wave patterns can be used to aid the development of the memory traces. For example, in the Seebach et al. study mentioned earlier, several thousand presentations of the CV syllable stimuli were needed in order to produce stable processes of learning. At first, this seems to be a rather high number of presentations—but the iterative, oscillatory circuits of the human brain would greatly reduce the number of real-world presentations that might be needed. The brain supplies many, additional iterations of each behavioral presentation. This amplification of learning and memory could be a reason for the natural, oscillatory processes noted by Llinas et al., (1998), Miltner et al., (1999), Bollimunta et al., (2011) and many others. Oscillatory patterns provide robust, intensifying patterns that can underlie Hebbian learning, and also could account for observed intensification of readiness potentials. It is perhaps unsurprising, then, that each human being appears to have a unique "neural fingerprint" (Finn et al., 2015), given that ultimate causes for brain activity and decision-making will be unique for each individual, given unique genetic heritage (excepting identical twins) and a unique life-time of experience for each. Within the 'noise', therefore, are unique sub-patterns of activity that may produce different responses across individuals or within an individual as context changes.

Conscious Noise and Hidden Nature Physicalism

What we find both interesting and promising about this approach to consciousness is that it indicates how one might get past the current impasse in philosophy of mind with regard to how the reductive materialist project should be pursued. It should be noted that, just as there are three versions of the DPT, there are at least two versions of reductive materialism (RM) as well. Some versions of RM make explicit claims regarding which physical features or mechanism with which we are to identify mental states, and some do not. Let us call the former Explicit Reductive Materialism or (ERM) and the latter Hidden Nature Physicalism. Early-twentieth century attempts by behaviorists to account for mental states solely in terms of behaviors and behavioral tendencies would count as one version of ERM. Mid-twentieth century thinkers then proposed to identify

mental states and their features with specific neurophysiological states and their features (a view known as the "Identity Theory"). This is also a version of ERM. (Smart 1959) This view soon had to confront the conceptual objection that one could not expect exact neurophysiological identity between two individuals in the same mental state only similarity, and such similarity needed to be explained as functional identity. (Fodor 1981) There now seems to be empirical confirmation for this possibility. (Finn et al. 2015]

The latter part of the twentieth century was then dominated by attempts by functionalists to account for mental phenomena in terms of what sorts of functional roles particular physical states might instantiate.[1] But, while initially enormously promising, the Functionalist Turn has not proven able to deal successfully with the Hard problem. (Ludlow, et. al., 2004) The inability to produce a generally convincing functional account of consciousness has led some recent philosophers of mind to embrace instead a non-explicit version of reductive materialism which is now often referred to as "Hidden Natures Physicalism". But, what is Hidden Nature Physicalism and why does the program for the study and discovery of the physical nature of consciousness here proposed involving 'Noise" seem to support it?

Hidden Nature Physicalists claim that the nature of mental states is not completely revealed in our experiences of them, which entails that that there is more to consciousness than its experiential nature. For example, Chris Hill says the following:

> "If the Cartesian argument is to succeed…the essential nature of pain must be fully accessible to us when we experience pain. But it is precisely this thesis about the essential nature of pain that is called into question….There is no guarantee that experiential representation of pain will do full justice to its essential properties." (Hill 2009, 118)

Similarly, Patricia Churchland comments:

> What is troublesome is the idea that all the reality there is to a sensation is available through sheerly having it….I suggest, instead, a rather simple alternative: A sensation of pain is real, but not everything about

1. Whether Functionalism is itself a version of Reductive Materialism or a form of Nonreductive Materialism is not a settled matter. Some functionalists consider themselves to reductivists, other regard themselves as non-reductivists. The key to the disagreement is which and how strict an account of reduction one accepts. Those who accept Ernst Nagel's account of reduction will count themselves as non-reductivists; those inclined to the Kemeny-Oppenheim approach will consider themselves to be reductivists. (Nagel 1961, Kemeny and Oppenheim 1956)

> the nature of pain is revealed in introspection—its neural substrate, for example, is not so revealed." (Churchland, 1998, 117).

Similarly, Michael Tye says he rejects the view that "experience itself reveals red or canary yellow as simple, as not having a hidden nature". (Tye, 2009, 142). Thus for Hidden Nature Physicalists there must be something in addition to mere experience to account for consciousness, and this something is its physical nature, about which we are currently ignorant.

Why is HNP difficult for some physicalists to accept? Many physicalists are reluctant to make claims defending their views that are ultimately based upon current ignorance and the issuing of empirical promissory notes. Such a strategy seems not to be much an improvement over the various forms of the DPT discussed earlier. But, even a casual study of current brain research makes one painfully aware of the woeful state of our ignorance regarding how the brain might work to produce consciousness. If such ignorance really is our current epistemic condition, then, we should act accordingly. If we take the objections to all forms of DPT to be conclusive, and, if we think that objections to all other forms of RM are also convincing, and if we think that some version of physicalism is the only palatable alternative, then, as distasteful as some may find it, the most reasonable alternative seems now to be to embrace HNP.

While variants on HNP have been proposed in the past, these variants typically involved claiming either that there was a conclusive epistemological limitation that prevented humans from ever figuring out what physical properties consciousness might be,(McGinn, 1989) or that consciousness was to be conceived as analogous to anti-matter and could not be studied scientifically (Chalmers, 1995). Both of these alternatives, unlike the sort of approach embrace by the conscious noise proposal defended here, are seriously unsatisfying from the perspective of the scientific researcher.[2]

Let us take HNP, then, simply to be the general thesis that [1] although humans are wholly physical beings with no special levels, messily emergent features or spooky stuff, simple mind-brain identity theories are inadequate to account informatively for all mental phenomena; [2] functional accounts, while somewhat helpful in accounting for similarities across individuals and species, are nonetheless unable to account in an enlightening way for phenomena such as consciousness, and [3] there is some, as yet, undiscovered purely physical aspect of human existence that does explain such

2. McGinn's Mysterianism seems under-motivated and sells comparative scientific procedures short, (Kraemer 2007); and the Chalmers' anti-matter approach deliberately flirts with pan-psychism, a view which seems evolutionarily under-motivated as well as unsupported by our best current evidence.

phenomena. It should be clear from what has been said that there are at least two version of HNP, those variants which claim that, for one reason or another, mental states are physical but their physical natures are not discoverable by human beings, and those which urge that for all we know we may very well be able to discover and adequately explain the physical nature of consciousness. Let us call the first variant, Forever Hidden Nature Physicalism, (FHNP), and the second variant Currently Hidden Nature Physicalism, (CHNP). From the perspective of scientific researchers working on consciousness the latter variant is the more interesting, and the one which relates directly to the noise account proposed above.

What the pre-conscious noise account actually makes plausible is why the specific nature of consciousness should have remained hidden, and why philosophers of mind and other interested parties should be happy to support neuroscientists in their efforts to try to make clear just what sorts of developed patterns different forms of consciousness might turn out to be. And this is precisely what the pre-conscious noise proposal does: [1] it suggests a source for items out of which consciousness might be constituted; [2] it suggests a mechanism or physical procedure whereby consciousness might be seen to develop over time; and, [3] it also makes it very clear why the kind of enlightening account of phenomenal features that we think is currently missing from the physicalist camp has not been readily apparent to all; which [4] in turn explains explicitly the current and contingent hiddenness required by the CHNP. These are, we think, very serious advantages that the CHNP offers with respect to solving the problem of consciousness.

Conclusion

The conscious noise approach to consciousness outlined here promises a satisfying reply to a requirement set down some years ago by Donald Davidson with respect to the mind-body problem. (Davidson 1980) Davidson claimed that any adequate solution to the mind-body problem must be able to provide not only a convincing account of mental phenomena, but must also be able to provide a convincing explanation as to why it took so long for human beings to figure out how to answer the mind-body problem. And, the pre-conscious noise proposal is certainly able to provide a most satisfactory explanation on that score.

So, one might then ask, if the project of determining the mechanics of consciousness are to be turned over to the neurophysiologists and their scientific allies, what role should philosophers of mind who support the CHNP continue to play? Another way to pose the same question is to ask: what intellectual burdens must CHNNP defenders assume?

There seem to be at least three. First, defenders of CHNP need to provide convincing arguments for to articulate the numerous advantages of the CHNP approach. Second, there are serious objections that have been raised specifically against this view (Fumerton 2013, Robinson 2014), and defenders of the CHNP approach need to provide convincing philosophical rebuttals to them, as well as trying to anticipate and respond to other philosophical worries for CHNP that are likely to be raised. And, third, supporters of CHNP further need to be ever on the look-out for a plausible empirical strategy that supports the view by demonstrating not only how consciousness might arise but also by indicated how the currently hidden nature of consciousness might come to be made public. This discussion has attempted to provide some initial support for all three of these projects.

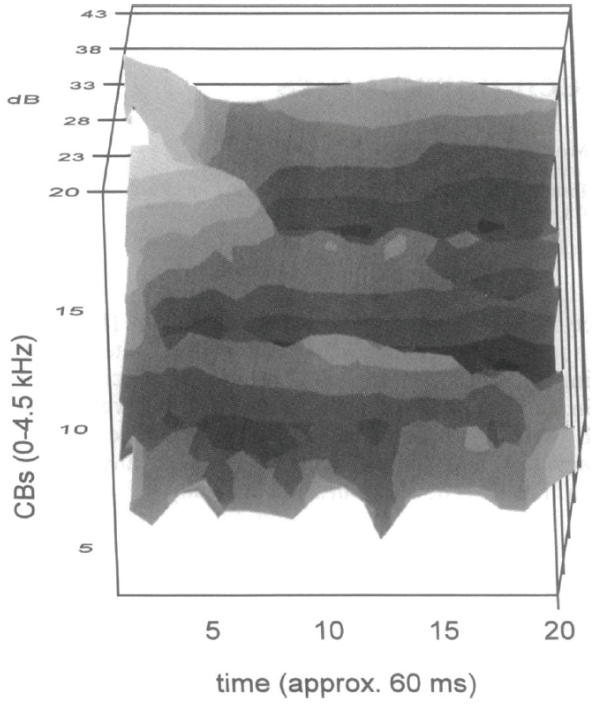


Figure 1. Modified from Seebach et al. 1994. (A) Average energy contour for each of the three training syllable types for speaker BR, shown as gray-scale images. The lighter areas of these images represent the presence of greater acoustic energy, with the ordinates of each image representing increasing frequency on a critical-band scale, and the abscissa of each image representing increasing time, which goes from 0 to 38 ms as marked by the start of each sampling window. (B) Gray-scale images of resulting synaptic weights for the five cells (neurons) of a BCM artificial neuronal network following developmental training using the inputs shown in (A). Collective responses provide a clear ability to discriminate among the different training syllables, and among the same syllables as spoken by other people.

Philosophy References

Armstrong, David. 1968. *A Materialist Theory of the Mind*. London: Routledge.

Campbell, Keith. 1984. *Body and Mind*, 2nd ed. Notre Dame, IN: Notre Dame University Press.

Chalmers, David. 1995. "Facing Up to the Problem of Consciousness." *Journal of Consciousness Studies* 2: 200–219.

Cornman, James. 1981. "A Non-Reductive Identity Thesis about Mind and Body." In *Reason and Responsibility*, edited by Joel Feinberg, 285–296. Belmont, CA: Wadsworth Publishing Co.

Churchland, P.S. 1998. "Brainshy: Nonneural Theories of Conscious Experience." In *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates*, edited by Hameroff et al., 109–126. Cambridge, MA: MIT Press/Bradford.

Davidson, Donald. 1980. "The Material Mind." In *Essays on Actions and Events*, edited by Donald Davidson, 245–61. Oxford: Clarendon Press.

Descartes, Rene. 1648. *Meditations on First Philosophy*.

Driesch, Hans. 1908. *The Science and Philosophy of the Organism: Gifford Lectures delivered at Aberdeen University, 1907-1908*. London: Adam and Charles Black.

Fodor, Jerry. 1981. "The Mind-Body Problem." *Scientific American* 244: 114–225.

Fumerton, Richard. 2013. *Knowledge, Thought and the Case for Dualism*. New York: Cambridge University Press.

Hill, Christopher. 2014. *Meaning, Mind and Knowledge*. Oxford: Oxford University Press.

Jackson, Frank. 2004. "Post Scripts." In *There's Something about Mary*, edited by Peter Ludlow, Yugin Nagasawa and Daniel Stoljar, 407-444. Cambridge, MA: MIT Press.

Jackson, Frank. 1986. "WhatMary Didn't Know." *Journal of Philosophy* 83: 291–295.

Kemeny, J. and Paul Oppenheim. 1956. "On Reduction." *Philosophical Studies* 7: 6–19.

Kim, Jaegwon. 2007. *Physicalism, or Something Near Enough*. Princeton, NJ: Princeton University Press.

Kraemer, Eric.2006. "Moral Mysterianism." *Southwest Philosophy Review* 22: 69–77.

Levine, Joseph.1983. "Materialism and Qualia: The Explanatory Gap." *Pacific Philosophical Quarterly* 64: 354–361.

Lewis, David. 1966. "An Argument for the Identity Theory." *Journal of Philosophy* 63: 17–25.

Ludlow, Peter, Yugin Nagasawa, and Daniel Stoljar (eds.). 2004. *There's Something about Mary.* Cambridge, MA: MIT Press.

McGinn, Colin. 1989. "Can We Solve the Mind-Body Problem?" *Mind* 98: 349–366.

Nagel, Ernst. 1961. *The Structure of Science*. New York: Harcourt, Brace and World.

Nagel, Thomas. 1974. "What is it like to be a Bat?" *The Philosophical Review* 83: 435–450.

Nagel, Thomas. 2012. *Mind and Cosmos*. Oxford: Oxford University Press.

Putnam, Hillary. 1999. *The Threefold Cord: Mind, Body and World*. New York: Columbia University Press.

Robinson, William. 2015. "Hidden Nature Physicalism." *Review of Philosophy and Psychology* 7 (1): 1–19.

Ritter, William. 1919. *The Unity of the Organism, or the Organismal Conception of Life*, two volumes. Boston: Richard G. Badger.

Searle, John. 2004. *Mind*. Oxford: Oxford University Press

Smart, J.J.C. 1959. "Sensations and Brain Processes." *The Philosophical Review* 68: 141–156.

Tye, Michael. 2009. *Consciousness Revisited*. Cambridge, MA: MIT Press.

Biology / Neuroscience References

Amundson, R. 2005. *The Changing Role of the Embryo in Evolutionary Thought: Roots of Evo-Devo* (Cambridge Studies in Philosophy and Biology). Cambridge: Cambridge University Press.

Bargmann, C.I., and T.C. Gilliam. 2013. "Genes and Behavior." In *Principles of Neural Science*, 5th Edition, edited by Kandel et al., 39–65. New York: McGraw Hill.

Beatty, J. 1994. "Ernst Mayr and the Proximate/Ultimate Distinction." *Biology and Philosophy* 9: 333–356.

Bienenstock, E.L., L.N Cooper, and P.W. Munro. 1982. "Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex." *Journal of Neuroscience* 2: 32–48.

Bollimunta, Anil, Jue Mo, Charles E. Schroeder, and Mingzhou Ding. 2011. "Neuronal mechanisms and attentional modulation of corticothalamic alpha oscillations." *Journal of Neuroscience* 31 (13): 4935–4943.

Calvin, W.H. 1998. "Competing for Consciousness: A Darwinian Mechanism at an Appropriate Level of Explanation." Journal of Consciousness Studies 5: 389–404.

Calvin, W.H. 2004. *A Brief History of the Mind: From Apes to Intellect and Beyond*. Oxford: Oxford University Press.

Caporale, Natalia, and Yang Dan. 2008. "Spike Timing–Dependent Plasticity: A Hebbian Learning Rule." *Annual Review of Neuroscience* 31: 25–46.

Finn, E. S., X. Shen, D. Scheinost, M. D. Rosenberg, J. Huang, M. M. Chun, and R.T. Constable. 2015. "Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity." *Nature Neuroscie*nce 18 (11): 1664–1671.

Hebb Donald. 1949. *The organization of behavior*. New York: Wiley & Sons.

Hodgkin, A.L., and A.F. Huxley. 1939. "Action potentials recorded from inside a nerve fiber." *Nature* 144: 710-711.

Hopfield, J.J. 1982. "Neural networks and physical systems with emergent collective computational abilities." *Proc. Natl. Acad. Sci. USA* 79: 2554–2558.

Hunnicutt, B.J, Long, B.R, Kusefoglu, D., Gertz, K.J., Zhong, H., and T. Mao. 2014. "A comprehensive thalamocortical projection map at the mesoscopic level." *Nature Neuroscience* 17: 1276–1288.

Kincaid, D.R., R.S. Varga, and C.H. Warlick. 2010. "The life and times of Dr David M. Young, Jr." *Numer. Linear Algebra Appl.* 17: 743–757.

Laland, K.N., K. Sterelny, J. Odling-Smee, W. Hoppitt, and T. Uller. 2011. "Cause and Effect in Biology Revisited: Is Mayr's Proximate-Ultimate Dichotomy Still Useful?" *Science* 334: 1512–1516.

Libet, B., C.A. Gleason, E.W. Wright, and 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* 106: 623–642.

Libet, B., E.W. Wright Jr., and C.A. Gleason. 1982. "Readiness-potentials preceding unrestricted 'spontaneous' vs. pre-planned voluntary acts." *Electroencephalogr Clin Neurophysiol* 54: 322–335.

Llinàs, R., D. Contreras, and C. Pedroarena. 1998. "The neuronal basis for consciousness." *Phil.Trans. R. Soc. Lond. B* 353: 1841–1849.

Mathewson, K.E., G. Gratton, M. Fabiani, D.M. Beck, and T. Ro. 2009. "To See or Not to See: Prestimulus α Phase Predicts Visual Awareness." *Journal of Neuroscience* 29: 2725–2732.

Mayr, Ernst. 1961. "Cause and Effect in Biology." *Science* 134: 1501–1506.

Miltner, W., C. Braun, M. Arnold, H. Witte, and E. Taub. 1999. "Coherence of gamma-band EEG activity as a basis for associative learning." *Nature* 397: 434–436.

Schurger, A., J. Sitt, and S. Dehaene. 2012. "An accumulator model for spontaneous neural activity prior to self-initiated movement." *Proc. Nat. Acad. Sci. USA* 109: E2904–E2913.

Seebach, B., N. Intrator, P. Lieberman, and L.N Cooper. 1994. "A model of prenatal acquisition of speech parameters." *Proc. Natl. Acad. Sci. USA* 91: 7473–7476.

Silver, Nate. 2012. *The signal and the noise: why so many predictions fail—but some don't*. New York: Penguin Books.

Tsien, R.Y. 1980. "New calcium indicators and buffers with high selectivity against magnesium and protons: design, synthesis, and properties of prototype structures." *Biochemistry* 19: 2396–2404.

Young, D.M., Jr. 1950. *Iterative Methods for Solving Partial Difference Equations of Elliptic Type*. Ph.D. thesis. Cambridge, MA: Harvard University, Mathematics Department.

Journal of Cognition and Neuroethics

The Insignificance of Empty Higher-order Thoughts

Daniel Shargel
Lawrence Technological University

Biography

Daniel Shargel is an Assistant Professor of Philosophy at Lawrence Technological University. He works on the philosophy of mind and cognitive science, and defends an embodied theory of emotion. He is exploring implications of this view for a wide range of emotional phenomena, such as intentionality, normativity, and the relationship between emotions and desires. His other philosophical interests include moral psychology, consciousness and perception.

Publication Details

Journal of Cognition and Neuroethics (ISSN: 2166-5087). April, 2016. Volume 4, Issue 1.

Citation

Shargel, Daniel. 2016. "The Insignificance of Empty Higher-order Thoughts." *Journal of Cognition and Neuroethics* 4 (1): 113–127.

The Insignificance of Empty Higher-order Thoughts

Daniel Shargel

Abstract
A crucial move in Kripke's modal argument is his assertion that awareness of pain is essential to pain. Lycan has argued that Kripke's assertion is not consistent with higher-order theories of consciousness. Ironically, Lycan's defense of higher-order theories against Kripke's argument is predicated on the fact that they allow for empty higher-order states: states of higher-order awareness that represent the presence of non-existent lower-order states. This very feature has been the focus of recent critics of higher-order theories, including Ned Block, who argue that it leads to absurdities. So the possibility of empty higher-order states is taken by different sides to be both the salvation and the destruction of higher-order theories. I will argue that both sides are mistaken. First, empty higher-order states only seem problematic when higher-order theories are misconstrued. Second, I will argue that Lycan's appeal to empty higher-order states is not ultimately effective. His critique is successful against Kripke's argument as he presented it, since Kripke does not address the case of empty higher-order states. However, it is possible to adjust Kripke's argument so that it is compatible with that possibility.

Keywords
Consciousness, Higher-order Theories, Modal Argument

1. Introduction

If you follow recent discussion of higher-order thought theories of consciousness, it seems clear that empty higher-order thoughts are their biggest threat. According the simplest gloss, David Rosenthal (2005, 2011) says that mental states are conscious when they are the object of a higher-order thought. A prominent criticism of higher order theories is based on the apparent possibility that higher order states can either misrepresent first order states, or occur without the first order states that they represent.[1] In the later case the higher order states are typically called 'empty'. There is an ongoing debate about whether such cases render such theories incoherent or implausible.[2] Reading

1. I am referring to higher-order theories, such as Rosenthal's (2005) and Lycan's (1987), in which the higher order content belongs to a distinct state. When I refer to higher-order theories I mean this kind unless I specify otherwise.

2. See Neander (1998), Block (2011) and Wilberg (2010) for versions of this critique, and Rosenthal (2011) and Berger (2014) for a defense, though the literature is extensive.

this literature gives the distinct impression that empty higher-order thoughts ('empty HOTs') are the biggest obstacle to the acceptance, or at least acceptability, of higher-order theories.

Participant in this debate over empty HOTs typically overlook the other, very different role that empty higher-order states play in a response to Saul Kripke's (1971) modal argument. Kripke claims that our knowledge of external objects is mediated, but we have immediate knowledge of conscious mental states like pains. Being in pain is itself sufficient for being aware of your pain. This is a crucial step in his argument that pains cannot be identified with any type of physical state. Lycan (1974, 1987) pointed out that according to higher-order theories such as his own, our knowledge of our own mental states is in fact mediated by distinct higher-order states. Since it is possible to have a higher-order state without the first-order state it represents, you can seem to have pain without really being in pain. Lycan is describing nothing other than empty higher-order states, and arguing that higher-order theorists have a uniquely effective response to the modal argument because their theories allow for these empty states.

Are empty higher-order states a vulnerability for higher-order theories, or are they a strength? I will argue that the answer is 'no'. First, defenders of higher-order theory are right to dismiss the empty higher-order state issue as a pseudo-problem. I will argue that once you properly understand the reasoning behind higher order thought theory, you will see why there is nothing strange about empty HOTs for a higher-order thought theorist. Second, Lycan's move certainly does defeat the specific argument Kripke presents in *Naming and Necessity*. I will show, however, that Kripke's argument can be adjusted to counter it. A materialist cannot refute Kripke's modal argument without addressing his deeper claims. The debate over whether pains are epistemically mediated is also a red herring. In order to properly evaluate higher order theories we need to set aside debates over empty HOTs.

2 Empty HOTs as a Weakness

2.1 Motivating Higher-order Theories

One way to describe the contemporary status of theories of consciousness is to describe the theories themselves. First, of all, there are first order theories and higher-order theories. Focusing only on the higher-order theories, there are occurrent and dispositional theories, theories of higher order thought and higher order perception, and theories where the higher order state is intrinsic and distinct from the first order

state. However, this sort of taxonomy obscures what is really at stake, and why anyone endorses any of these theories.

It is more enlightening to begin with debate over the explanandum rather than the explanans. On one end of the spectrum is Galen Strawson (2006), who thinks consciousness is essentially phenomenological, so that any attempt to reduce consciousness to non-conscious phenomena reveals a failure to grapple with consciousness itself. On the other end are explicitly eliminativist theories, such as those defended by Paul Churchland (1981) and Patricia Churchland (1986) which reject the notion that our folk notions correspond to any mental reality.

Higher-order theories are based on a conception of consciousness that falls somewhere in between. David Rosenthal (2011) begins his reply to Ned Block (more on Block later) by saying that, "A state's being conscious is a matter of mental appearance – of how one's mental life appears to one. If somebody is in a mental state but doesn't seem subjectively to be in that state, the state is not conscious" (431). As I look out my window I see leafless tree branches against a grey sky. The fact that I see those tree branches is a mental phenomenon that requires explanation. However, there is a second phenomenon that also requires explanation: that *it seems to me that I am seeing* those tree branches. It would be possible for me to see, and yet, for it not seem to me that I see anything. We can find a clear example that involves audition. As I type right now I can hear a dishwasher running. I'm sure that I have been hearing it, without interruption, for the last several minutes, but I only just now realized that I am hearing it. That is paradigmatic example of the transition from nonconscious to conscious perception, since I just gained a new mental appearance of perceiving.

One could attempt to reinterpret my example of non-conscious perception in two different ways. First, one could claim that I didn't truly hear it before, since hearing must be a conscious state. My non-conscious sound-detection falls short of hearing in some key respect. Second, one could claim that I did hear it before, but since all hearing is conscious, I actually heard the washing machine consciously. Both moves are motivated the assumption that all mental states are conscious, and this assumption leads them astray. The first strategy runs aground on the fundamental similarity between the conscious and non-conscious cases of sound detection. Both types of perception provide the hearer with the same types of information, although in nonconscious cases the signal is often weaker (Lau 2008). The second reply faces the objection that it didn't seem to me beforehand that I was hearing any dishwasher, so it must not have been conscious. One could insist that it did, in fact, seem to me that I heard it at the time, despite my denial, but without corroborating evidence that move is just not compelling.

The critical starting point for Rosenthal (2011) is that I seem to have (we can set aside for now the question of whether I really do) unmediated access to my current state of mind (432). By seeing the tree I gain access to facts about the tree. That is visual perception. In addition, I also seem to have unmediated access to my state of mind. That is consciousness.[3] The job of a theory of consciousness is to explain why it seems to me that I have unmediated access to my state of mind. If that is the job, then it seems very tractable. In general appearances/seemings and reality can diverge, and consciousness is just a special class of seeming – it is the way my mind seems to me. We should expect that the states responsible for my mental reality are distinct from the states responsible for mental seemings.[4] The tree is distinct from the perception of the tree, and similarly, perceiving is distinct from seeming to perceive.

Higher-order awareness theories have a very simple explanation for consciousness. We can call my perception of the tree a first order state. This state makes me aware of the tree, but does not, by itself, make it seem to be that I have that awareness. This is called a first-order state. In order for it to seem to me that I see the tree – for it to seem to me that I am in that first-order state, I need to have a higher-order state. We call this a higher-order state because it represents the occurrence of another state. In this case, the higher-order state represents the occurrence of my perception of the tree. If I represent myself, via this higher-order state, as currently perceiving the tree, then it will seem to me *that I perceive the tree*, just as my perception of the tree itself makes it seem to me *that there is a tree*.

There is extensive debate about the nature of the higher-order state. Lycan (1987) argues that it is a perception of the first order state, while Rosenthal (2005) argues that it is a thought about that state. I will not discuss that disagreement further, since their theories are, for the sake of this paper, similar enough. They both deny that mental states are intrinsically conscious, and they both argue that mental states become conscious in virtue of a distinct mental state with assertoric mental attitude. I will often focus on higher-order thought theory, but the moves that I lay out on either side would be relevant for either.

---

3. To be more precise, this is what Rosenthal (2005) calls 'state consciousness', the phenomenon of having conscious mental states. State consciousness is sometimes confused with creature consciousness (being awake) and transitive consciousness (awareness of an intentional object), theories of consciousness are generally theories of state consciousness.

4. It is important to keep in mind, however, that mental seemings are themselves part of mental reality. This point will be very important as the discuss develops.

2.2 Full and Empty HOTs

I consciously hear the dishwasher running, but according to higher order theories my auditory perception itself is not responsible for my conscious experience. Instead, that experience is entirely determined by a distinct higher-order state. This may seem to be a fundamental mistake. How could my perception be metaphysically divorced from my perceptual experience?

The case of empty HOTs is designed to make this problem more vivid. If the higher-order state is entirely responsible for conscious experience, then it should be possible to have the higher-order state without the first-order state that it represents – an empty HOT. The higher-order state has the content, perhaps, that "I am in pain," despite the fact that I am not. What does Rosenthal say in this situation? As long as my higher order thought does not seem to have arisen via observation or inference (Rosenthal 2011, 423), I will have a conscious pain. So, in that situation, I am not in pain, but I have a conscious pain. When you put it like that it is hard to dispute Ned Block's (2011) claim that the view is unworkable.

However, higher-order theorists can simply respond: "Don't put it that way!" That way of framing the empty HOT case is not quite inaccurate, but it is highly misleading. Jacob Berger (2014) pointed out very clearly what critics of Rosenthal typically misunderstand. State consciousness, despite the misleading term, is not a property of states. When my auditory perception of the dishwasher becomes conscious, the higher-order state does not have any effect on the first-order state. Instead, the higher-order state has an effect on me: it makes me aware that I have an auditory perception. This follows directly from Rosenthal's conception of consciousness as the phenomenon of mental appearances. When a mental state becomes conscious, your mind now appears to be in that state, when before it did not appear to be.

What about empty HOTs, those conscious states that paradoxically do not exist? When you frame them in terms of mental appearances the paradox disappears. If you have a HOT with the content, "I am in pain," then it will seem to you as though you are in pain. If, at the same time, you lack the first order state, then you are not really in pain. There is no need to say that there is a non-existent state that is nonetheless conscious. Instead, just say that your mind appears to be different from the way it really is.

Once we avoid misleading characterizations, it becomes clear who should and who shouldn't accept higher-order theories. First of all, higher-order theories reject the Cartesian view that the mind is necessarily the way that it appears to be. If you accept the Cartesian view, then that is already sufficient reason to get off the boat. Second, higher-order theories take the phenomenon of consciousness to be nothing other than

the phenomenon of mental appearances. Once mental appearances are explained, there is nothing more for a theory of consciousness to do. If you reject this conception, then you should not be a higher-order theorist. Third, if you think that having an intentional state with an assertoric mental attitude, and content about one's own mind, is sufficient for having a mental appearance (for making your mind seem to be a certain way), then you should be a higher-order theorist. If not, you probably need some other sort of theory.

If you want to argue against higher-order theories, you would do well to argue that the mind is identical to the way that it appears to be. Or, argue that consciousness is something other than mental appearances. Or, argue that higher-order states are not sufficient to create mental appearances. Any of those could lead to a productive discussion.

3 Empty HOTs as a Strength

3.1 The Modal Argument

After concluding that empty HOTs do not pose any sort of threat to higher-order theories, we will now consider whether they might instead provide salvation. Specifically, does the fact that higher order theories allow for empty higher-order states give them a unique and effective response to anti-materialist arguments? That is exactly what Lycan (1974, 1987) proposed.

Kripke argued in *Naming and Necessity* (1972) that proper names and natural-kind terms are rigid designators, and therefore all identity statements that use two of these terms are necessarily true if true at all. Furthermore, Kripke takes conceivability to imply possibility. If someone can conceive of A's existing without B, then it is possible for A to exist without B.[5] Taken together, these claims appear to undermine claims of *a posteriori* identity. 'Heat' and 'molecular motion' are presumably natural-kind terms, so if 'heat = molecular motion' is true at all, it is true in all possible worlds in which heat occurs. But it may seem conceivable that heat could exist without molecular motion. Given Kripke's assumptions, this would falsify the identity.

5. This interpretation of Kripke is wide-spread, though still controversial. For alternatives see Byrne 2007 and Papineau 2007. Both deny that Kripke is committed to conceivability's implying possibility, though they disagree about the actual nature of his argument. I take the usual interpretation to be accurate, but I will not defend it. For present purposes it is sufficient that I capture the argument as Lycan, Rosenthal and many others have seen it.

Kripke has a stock response for dealing with such cases. Any individual who claims to imagine the occurrence of heat without molecular motion is confused. The sensation of heat mediates our knowledge of heat itself. What the challenger actually imagines is the sensation of heat, which is an epistemic mediator of heat, without any molecular motion. In general, when someone claims to imagine the occurrence of A without B, he or she might really be imagining the epistemic mediator of A occurring without B.

Identity theorists identify pain with C-fiber firing (or some other type of neural state). Kripke claims that both 'pain' and 'C-fiber firing' are rigid designators; so if 'pain = C-fiber firing' is true, then it is necessarily true. It seems that we can imagine a pain that is not a C-fiber firing. Given Kripke's assumptions, this is a prima facie reason to doubt that pain is really C-fiber firing. Can this problem be resolved in the same way as with heat and molecular motion?

It could if we were not actually imagining a pain that is not a C-fiber firing, but an epistemic mediator of pain occurring without any C-fiber firing. However, Kripke claims that there is no distinct epistemic mediator for pain. To be in pain is to be aware of having a pain, and vice versa. If so, the strategy that works for other cases of necessary identities known a posteriori fails for pains, and some other mental states as well. Kripke concludes that these mental states are not identical with any physical states.

3.2 Lycan's Response

If Lycan's higher-order view is correct, then he can defend the theory that pain is C-fiber firing in the same way that Kripke defends the theory that heat is molecular motion. Kripke denied that anyone could imagine the occurrence of heat without molecular motion. Instead, what the challenger really imagines is the occurrence of heat without the sensation of heat. Analogously, Lycan asserts that anyone who claims to imagine having a pain without any C-fiber firing is confused. The challenger is really imagining the awareness of pain, which on Lycan's hypothesis is a suitable higher-order representation, and can occur without any C-fiber firing. The higher-order representation could occur in the absence of any actual pain, which would be the case (by hypothesis) if there were no C-fiber firings. So the challenger is actually imagining a state of affairs perfectly compatible with the identification of pains with C-fiber firings.

This is not the only critique of the modal argument that Lycan makes. He also contests the view that 'pain' is rigid (1987: 14). This is a very different kind of objection. When Lycan asserts that pains are epistemically mediated he makes a delicate surgical defense of materialism - denying one feature of Kripke's argument while leaving the rest

of the apparatus intact. Denying that psychological terms are rigid, by contrast, is more like amputating a limb.

It is often more appealing to make a minimally invasive critique, so it would be preferable for higher order theorists if the former move were sufficient by itself. They are already committed to denying pains are intrinsically conscious, so this defense against the modal argument seems to come for free. The remainder of the paper explores whether the higher order move really is sufficient, only considering more aggressive strategies at the end.

3.3 Retreat to Higher Ground

Kripke never presents a response to this move, perhaps because he finds the identification of pain with awareness of pain so obvious. But there are effective moves that Kripke could make which follow naturally from Lycan's application of the apparatus in *Naming and Necessity*.

Follow Lycan in taking awareness of pain to be distinct from pain. This gets around the problem for pains, since by hypothesis they do have distinct epistemic mediators. But at the same time, it suggests a new problem. What about the awareness of pain? According to materialists it too is identical with some type of physical state or other. Let's call those physical states D-fiber firings, for lack of a better term. The identification of awareness of pain with D-fiber firing raises problems parallel to those we had with pain and C-fiber firing. 'Awareness of pain' and 'D-fiber firing' are presumably natural-kind terms, so 'awareness of pain = D-fiber firing' is necessary if true at all. And It seems as though we can imagine having an awareness of pain without any D-fiber firings. This gives us a *prima facie* reason to deny that awareness of pain really is D-fiber firings.

So the question arises, is awareness of pain itself epistemically mediated? Materialists face a dilemma. If they hold that awareness of pain is not epistemically mediated, and follow Lycan's application of the *Naming and Necessity* apparatus, then Kripke immediately wins. If you seem to imagine having an awareness of pain without having D-fiber firing, then that really is what you imagine, and it really is a possibility. The awareness of pain cannot be D-fiber firing after all. And since 'D-fiber firing' is just a stand-in for whatever neuroscientists will eventually tell us is the neural correlate for awareness of pain, awareness of pain cannot be identical to any type of physical state.

The other option is to claim that our awareness of pain is also epistemically mediated. Perhaps it is mediated by a third-order representation, resulting in a kind of

introspective awareness.[6] Bracketing any dispute over whether such states exist, this proposal only postpones defeat. The same move that Kripke makes concerning pain, and could make concerning awareness of pain, he could make yet again for third-order awareness of pain. According to materialists, states of third-order awareness must again be identical with brain states. However, we can imagine that they occur without any proposed neural correlates. The materialist is back in the same place again, no better off than before. Either third-order awareness has no epistemic mediator, or it does have one. In one direction lies immediate defeat, and in the other a vicious regress.

Lycan (1987: 13) does anticipate that his initial move might lead to a regress. In response, he appeals to Armstrong's (1981) view that each level of higher order awareness requires a distinct physical mechanism to implement it, and any individual will have a finite number of such mechanisms. This argument shows that no one has infinite levels of higher order awareness, which seems to be the regress that he meant to address. However, this does address the dilemma presented above. Lycan claims that some level of awareness is as high as we go, but it is still necessary to explain how we are aware of those highest-order states. A challenger might claim to imagine being in such a state without the proposed neural correlate. Lycan cannot reply in the standard higher order manner, that epistemic access is mediated by higher order states, since in this case there are no higher order states. He also cannot say that these states lack a distinct epistemic mediator, since given the established rules that would amount to conceding defeat.

3.4 Another Round

There is one more move that Lycan or a like-minded theorist could make without contesting substantive features of Kripke's apparatus. Perhaps, following Lycan and Armstrong, there is some level of higher-order awareness that is as high as we can go, given the limits of our psychology. Let it be the third-order awareness mentioned above, but which level it might be makes no difference here. Lycan can avoid the first horn of the dilemma, immediate defeat, by denying that we are directly aware of our third-order states. At the same time he can avoid the second horn of the dilemma, the regress, by denying that we are ever aware of third-order states via a fourth-order state, and indeed that we ever could be.

Instead, we become aware of the existence of third-order states in a third-person manner, by inferring their occurrence from our own behavior. This could work in different

6. David Rosenthal (2005, 28-29) explains introspective awareness by appealing to this sort of third-order state.

ways, but perhaps the easiest would be by listening to our own speech. We might hear ourselves say, 'I am aware of my pain.' The awareness of pain is itself, by hypothesis, a second-order state, since it is the epistemic mediator of pain. If we are aware of that second-order state then we must have yet another state, a third-order state. The fact that we verbally reported a second-order state implies that we are aware of it, so we can infer on the basis of that speech act that we are in fact in a third-order state.[7] This may seem a rather arcane inference to make, but then again, we are rarely aware of our third-order states. Perhaps this explains why.

This story suggests a way to avoid both horns of the modified modal argument. When a dualist asks whether third-order awareness of pain is identical with some type of brain state, a materialist can say yes. The dualist then says that we can imagine having such third-order awareness without its neural correlate, giving us prima facie reason to doubt that identity. A materialist, however, could deny that we really imagine having third-order awareness without the relevant brain state. Instead, lacking first-person access to our third-order awareness, we imagine inferring from one of our own speech acts that we are in such a state.

However, this might be a false inference. It is possible to say, 'I am aware of my pain' without actually being in a state of third-order awareness. Normally we will make that type of utterance when we really are aware of being aware of our own pain - in other words, when we are in a state of third-order awareness. But in some cases we may speak insincerely, or our speech may result from self-deception. Any inferences based on those sorts of speech acts will be mistaken. So if a challenger claims to imagine having third-order awareness of pain without the relevant neural correlates, Lycan could say, 'You have no direct epistemic access to third order states. You must be imagining inferring the existence of a third-order state from a speech act, and that speech act might be insincere or self-deceptive. Therefore, you might not be imagining being in such a state after all.`

This move would successfully avoid the regress, but a challenger is not likely to be satisfied. When Kripke says in *Naming and Necessity* that we are really imagining one thing rather than another, he is careful to propose an alternative that sounds plausible. It is not completely implausible that there are cases where we apparently imagine heat but really imagine the sensation of heat. But it is quite another thing to be informed that

7. This is reminiscent of Dretske's (1994) displaced-perception theory of introspection. Dretske, however, takes introspection to be a special case of displaced perception, in which subjects have privileged access to their mental states. It is crucial to this account that subjects access their mental states in a third-person, fallible manner.

we did not imagine having third-order awareness, but instead imagined inferring the existence of such a state on the basis of an insincere or self-deceptive speech act. Would we not be aware that we were imagining hearing a speech act? Why was it insincere or self-deceptive? Perhaps answers could be provided, but the whole line of reasoning seems dubious. Mind-body materialism deserves a stronger defense.

3.5 A Brutal Finish

As we have seen, the regress can be prevented, though at the cost of testing our credulity. However, even this move is vulnerable to another, more ruthless dualist attack, put forward by Kripke himself. Forget C-fibers and D-fibers. Kripke (1971: 161) says that we can apparently imagine pain without any neurons whatsoever. A materialist might reply that Kripke only seems to imagine having a pain without neurons. Instead, he imagines having the awareness of pain without having any neurons.

This response worked before, when the question was whether we can imagine pains without C-fibers. Lycan suggested that we only seem to imagine pain without C-fiber firings, while really imagining the awareness of pain without C-fiber firing. The latter is perfectly compatible with the necessary identity of pain with C-fiber firing. However, the awareness of pain, according to the materialist, is identical with some type of neural state.[8] If the Kripke really imagines the awareness of pain without any neurons whatsoever, then that would, on Kripke's apparatus, falsify any such identity, and with it identity theory in general.

Recall the assumptions that Lycan accepted from Kripke. Whatever we can imagine is possible. If it seems that we can imagine something, we can be mistaken only if we confuse the presence of something with the presence of its epistemic mediator. Kripke says we can imagine a being that has pain without a human brain, perhaps without any body at all. Lycan, according to the rules he accepted, can only deny this by claiming that he is imagining a being that is aware of pain without having a brain. But for a materialist this is no improvement. It does not matter whether pain is nonphysical, or awareness of pain is nonphysical. Neither conclusion is acceptable to a materialist.

8. Strictly speaking, materialist versions of functionalism do not require that mental states be identical to neural structures. However, if it is possible to imagine mental states without any neurons, it is presumably also possible to imagine them without any physical structures that have a suitable functional organization, so the argument should be equally applicable to functionalist theories.

4 Conclusion

It is natural that Lycan, in his attempt to defeat the modal argument, began by taking on-board the model of conceivability and possibility that Kripke developed in *Naming and Necessity*. In the years since its publication this model has become something of an industry standard, and it is generally preferable when making an argument to avoid unpopular commitments.

Lycan's response to Kripke proves no more effective than using empty HOTs to attack higher-order theories. Adopting a higher order theory of consciousness is not sufficient for defending materialism against Kripke's argument. Kripke framed his modal argument in a manner that begs the question against higher order views, but it can be reframed to address this weakness. Defenders of materialism need to dig deeper, and contest some of Kripke's more popular views. Just as critics of higher-order theory ought to redirect their attacks, higher-order critics of Kripke's argument need to do the same. Does conceivability imply possibility? Are 'pain', and similar psychological kinsd terms, rigid designators? Lycan himself asks these sorts of questions, though he does so after making a more broadly palatable critique based on his higher order theory. If they desire to defeat the modal argument, even higher order theorists need to lead with these less palatable critiques.

References

Armstrong, David. 1981. *The Nature of Mind and Other Essays*. Ithaca: Cornell University Press.

Berger, Jacob. 2014. "Consciousness is not a property of states: A reply to Wilberg." *Philosophical Psychology* 27 (6): 829–842.

Block, Ned. 2011. "The higher order approach to consciousness is defunct." *Analysis* 71 (3): 419–431.

Byrne, Alex. 2007. "Possibility and imagination." *Philosophical Perspectives* 21 (1): 125–44.

Chalmers, David J. 1996. *The conscious mind: In search of a fundamental theory*. New York: Oxford University Press.

Churchland, Paul M. 1981. "Eliminative materialism and the propositional attitudes." *Journal of Philosophy* 78: 67–90.

Churchland, Patricia S. 1986. *Neurophilosophy*. Cambridge, MA: MIT Press.

Dretske, Fred. 1994. "Introspection." *Proceedings of the Aristotelian Society* 94: 263–78.

Jackson, Frank. 1982. "Epiphenomenal qualia." *The Philosophical Quarterly* 32 (127): 127–36.

Kripke, Saul. 1971. "Identity and necessity." In *Identity and individuation*, edited by Milton K. Munitz, 135–64. New York: New York University Press.

Lau, Hakwan. 2008. "Are we studying consciousness yet?" In *Frontiers of Consciousness: Chichele Lectures*, edited by Lawrence Weiskratz and Martin Davies, 245–258. Oxford University Press.

Levine, Joseph. 1983. "Materialism and qualia: The explanatory gap." *Pacific Philosophical Quarterly* 64 (4): 354–61.

Lycan, William G. 1974. "Kripke and the materialists." *The Journal of Philosophy* 71 (18): 677–89.

Lycan, William. G. 1987. *Consciousness*. Cambridge, Mass: MIT Press.

Nagel, Thomas. 1974. "What is it like to be a bat?" *The Philosophical Review* 83 (4): 435–50.

Neander, Karen. 1998. "The division of phenomenal labor: A problem for representational theories of consciousness." *Nous 32*: 411–434.

Papineau, David. 2007. "Kripke's proof is ad hominem not two-dimensional." *Philosophical Perspectives* 21 (1): 475–94.

Rosenthal, David M. 1986. "Two concepts of consciousness." *Philosophical Studies* 49 (3): 329–59.

Rosenthal, David M. 2005. *Consciousness and mind.* New York: Oxford University Press.

Rosenthal, David M. 2011. "Exaggerated reports: reply to Block." *Analysis* 71 (3): 431–437.

Strawson, Galen. 2006. "Realistic monism: Why physicalism entails panpsychism" *Journal of Consciousness Studies* 13: 3–31.

Wilberg, Jonah. 2010. "Consciousness and false HOTs." *Philosophical Psychology* 23 (5): 617–638.

cognethic.org