PERSPECTIVES

International Postgraduate Journal of Philosophy

Vol. 6, Issue 1 **September 2016**

Special Issue
Social Cognition

UCD School of Philosophy

PERSPECTIVES

International Postgraduate Journal of Philosophy

Volume 6
Autumn 2016

SPECIAL ISSUE:

Social Cognition

UCD School of Philosophy

PERSPECTIVES
International Postgraduate Journal of Philosophy

ACKNOWLEDGEMENTS

The Editors gratefully acknowledge the support of
University College Dublin, School of Philosophy.

FUNDING

UCD School of Philosophy
UCD Seed Fund

THANKS TO

Maria Bahgramian
Board of Reviewers
UCD School of Philosophy
UCD College of Human Sciences

Contents

SPECIAL ISSUE

Social Cognition

EDITORIAL

INTERVIEW

ARTICLES

BOOK REVIEW

ABOUT THE CONTRIBUTORS:

Genevieve Hayman (University College Dublin, Ireland)

Genevieve Hayman received her B.A. with Proudian Honours from the Johnston Centre for Integrative Studies at the University of Redlands. Her thesis was titled, "The Tangible and Intangible Aspects of Personhood." She completed her Master's degree in Consciousness and Embodiment at the School of Philosophy at University College Dublin in 2015. Her master's dissertation was titled, "Time for Phenomenology and Science: On Francisco Varela's Naturalization of Husserlian Time-Consciousness."

Ane Engelstad (University College London, Northern Ireland & UK)

Ane Engelstad is a graduate student at University College London. Her research is concerned with questions about social ontology, especially the assumed conception of personhood and social identity in egalitarian political theory. She is supervised by Professor Jonathan Wolff.

Max Skipper Griffiths (University of Bristol, Northern Ireland & UK)

Max Skipper Griffiths is taking his MA in Biological and Cognitive Sciences at University of Bristol. Within philosophy, he is interested in questions concerning language acquisition, neural plasticity in relation to learning, and the role of implicit bias.

Tony Cheng (University College London, Northern Ireland & UK)

Tony Cheng is doing a PhD at University College London. In the past few years he has been focusing on the relationship between visual attention and awareness. In his PhD project, he explores various aspects of sense perception and space, e.g. objectivity, the nature of sensory fields in different modalities, Molyneux's Question, the Kantian Spatiality Thesis, perceptual demonstratives, animal cognition. He is also interested in various areas in psychology, such as attention, cross-modal interaction/multi-sensory integration, and cognitive development.

David Markwell (University College Dublin, Ireland)

David Markwell received his Bachelor of Arts from Hamline University in philosophy and religion. He completed a Master's degree at University College Dublin before starting a PhD at the same institution. His PhD research attempts to construct an explicit ethical philosophy in the phenomenology of embodiment put forth by Maurice Merleau-Ponty.

Editorial **Open Access**

Austin Dwyer, Meline Papazian, David Markwell

Editorial

DOI 10.1515/pipjp-2016-0001

It is our pleasure to welcome you to the sixth volume of *Perspectives: International Postgraduate Journal of Philosophy. Perspectives* is a peer-reviewed annual publication, featuring articles, book reviews and interviews covering a broad range of issues and topics in philosophy and related disciplines. *Perspectives* aims to reflect the diverse interests of the graduate philosophy community at University College Dublin, publishing works from various philosophical traditions including analytic and continental. The journal also welcomes submissions in social and political philosophy, history of philosophy, philosophy of cognitive science and ancient philosophy. Our goal is to offer a platform for those in the early stages of research in academic philosophy, from Masters and PhD students to recent graduates, to gain experience of the process of academic publishing at the highest level.

The editors of the first volume, Anna Nicholson, Luna Dolezal and Sheena Hyland co-founded Perspectives with support from Prof. Maria Baghramian at UCD School of Philosophy. The inaugural issue of Perspectives was a special issue on the theme of the body and embodiment published in Autumn 2008 reflecting the papers presented at two annual postgraduate conferences at University College Dublin. The founding editorial board also published the second volume in the Autumn of 2009, a special edition on continental philosophy demonstrating the broad spectrum of the research interests of the editorial board and various contributors. Volume 3, was published in Autumn of 2010 with a new editorial board. Since then each volume is published by a different editorial board providing an excellent opportunity for postgraduate students in UCD School of Philosophy to learn about all the various stages involved in publishing an academic journal in philosophy. Volume 4, published in 2012, was a special issue on the theme of philosophy and nature and was followed by volume 5, a special issue on personhood and identity in Autumn of 2014.

This year has been a time of important changes for *Perspectives*. We are delighted to announce that starting with Volume 6, *Perspectives* will be published as an open access journal with De Gruyter Open, a leading open access publisher of 435 journals in various areas and disciplines. The opportunity to increase the international reach and reputation of *Perspectives* by moving the journal to an online open access platform is exciting and welcome. As always *Perspectives* will continue to publish high quality philosophical research from graduate, post-graduate, and early career researchers but now it has the opportunity to bring this research to scholars worldwide. Moving the journal to open access with De Gruyter makes *Perspectives* visible on over 300 online databases, including ProQuest and Google Scholar, and makes it a fully searchable and citable publication. The chance to publish *Perspectives* with De Gruyter Open was made possible by the generous contributions of University College Dublin Seed Funding and the UCD School of Philosophy. Through international calls for papers, the journal will continue to elicit high quality submissions that will then be subject to rigorous blind refereeing. The contributors will not be liable for the costs of publishing with De Gruyter Open and all such charges will be borne by the journal.

The journal will continue with its tradition of publishing thematic volumes drawn from a wide range of philosophical topics and approaches. The present volume focuses on the Philosophy of Cognitive Science in general and the theme of 'Social Cognition' in particular. Emerging trends in Cognitive Science in recent years have given rise to stimulating philosophical debates regarding processes of mind and cognition both in

Austin Dwyer, Meline Papazian, David Markwell, University College Dublin, Ireland

individuals and between them, as well as between embodied persons in dynamic relations with the environment. Furthermore, issues related to social cognition, understanding other minds and intersubjectivity are currently under vigorous investigation. In keeping with the journal's pluralist approach, we have welcomed submissions from various philosophical traditions and from a wide range of topics, including philosophy of psychology; philosophy of mind and cognitive science; classical as well as enactive, distributed, extended, situated, and embodied approaches to cognitive processes; animal cognition; social cognition, intersubjectivity and group behavior, social perception, empathy and other minds.

The volume opens with interviews with two original and authoritative voices in contemporary discussions of enactivism, Prof. Tony Chemero (University of Cincinnati) and Prof. Fred Cummins (University College Dublin). Two sets of questions were posed by the Chief Editors of the volume, Austin Dwyer and Meline Papazian. One set explored subjects relevant to the ongoing research of Chemero and Cummins and the other consisted of common questions aimed at comparing their views on topics such as, the role of philosophy and its different traditions in Cognitive science. Prof. Chemero starts by outlining what he thinks the role of philosophy should not be. He intends to incorporate phenomenology with radical embodied cognitive science. For Prof. Cummins cognitive science is closer to natural philosophy than science or philosophy and cannot be done from a "vantage point of pure disinterest". As for the topic of representation, for Chemero, human thoughts and experience work in more than one way. Cummins proposes notions of 'presence' and 'co-presence'. Regardless of their own theoretical preferences both suggest more inclusiveness in theoretical frameworks of Cognitive Science and both are for pluralism. Regarding his own research Chemero goes on to discuss synergies in complex systems where he also uses mathematical and experimental techniques to explore phenomenological concepts such as 'being-with'. Cummins explains his study of participatory joint speech, using dynamical systems in joint action, challenging our current notions of subject, agency and intentionality.

In addition to the interviews, the current volume publishes four articles on varied topics relevant to the theme of the journal. In her article 'Mirror Neurons, Husserl, and Enactivism: An Analysis of Phenomenological Compatibility', Genevieve Hayman critically examines the impact of mirror neuron research on aspects of social cognition. She uses Dan Zahavi's work to highlight incompatibilities between mirror neuron research and a phenomenological understanding of intersubjectivity. Then by examining Fuchs and De Jaegher's take on mirror neuron research, the author examines the possibility of phenomenologically sound mirror neuron research by these enactivists. She concludes that mirror neuron research should be contextualized in a broader phenomenologicaly compatible framework, using an Husserlian description of empathy and intersubjectivity, for which Fuchs and De Jaegher's account is a good candidate.

In 'False Consciousness and the Socially Extended Mind', Ane Engelstad presents a problem for the Marxist idea of false consciousness, as being vulnerable to accusations of dogmatism. She suggests that the concept needs to be further developed by incorporating Shaun Gallagher's theory of the socially extended mind. The article concludes with the proposal that, by using the notion of extended mind expanding towards external objects and systems, false consciousness as a reliable tool for the analysis of cognitive dynamics within power structures can be reinstated.

In 'Review of Carruthers' Massive Modularity Thesis' Max Skipper Griffiths discusses Peter Carruthers' Massive Modularity (MM) thesis according to which the central systems of the mind are widely encapsulated and operate via heuristics and approximation techniques. The article suggests that Carruthers faces an explanatory burden and must explain how the operations of those central systems are encapsulated in this 'wide-scope' sense. The author argues that by using a weakened definition of encapsulation, Carruthers thesis faces a dilemma. It is either a misnomer and unrecognizable as a version of MM or it is not and has to put forward a convincing claim of MM. The author suggests that Carruthers' commitment to his claim about central systems can meet this challenge by adopting an understanding of central systems whose information-frugal and processing-frugal operations allow for feasible tractability. He concludes that the Central Wide Scope thesis could provide a plausible account of Massive Modularity Thesis.

Tony Cheng, in 'The Situational Structure of Primate Beliefs' develops the *situational* model of primate beliefs using the Prior-Lurz line of reasoning. The author criticizes the skepticism of analytic

philosophy concerning primate beliefs where it is argued that non-human animals do not have beliefs because beliefs are propositional (e.g., Davidson 1975, 1982). The author offers two arguments against a propositional model, an *a priori* argument from referential opacity and an empirical argument from varieties of working memory. Secondly, the Prior-Lurz situational model is introduced with modifications. The author suggests that with this model, some progress is possible in understanding certain mindreading capacities of other primates.

This issue closes with a book review of Anthony Steinbock's *Moral Emotions: Reclaiming the Evidence of the Heart* by David Markwell. In his review Markwell praises Steinbock's project of carving out a space for emotions in moral discourse while remaining critical of Steinbock's chosen 'moral emotions' and their monotheistic biases.

Perspectives is a collaborative process between postgraduate editors, authors and referees, a voluntary labor of love. Many thanks to our referees who graciously offered their time to anonymously review the articles. We also thank Professor Chemero and Professor Cummins for kindly agreeing to respond to our questions, and our contributors for their efforts and cooperation. Special thanks to Professor Maria Baghramian for her continuous support, encouragement and invaluable advice.

Editors

Austin Dwyer, Meline Papazian, Georgios Petropoulos, Giulio Di Basilio, Maeve Kelly, Fionn McGrath

Open Access

Interviews

DOI 10.1515/pipjp-2016-0002

Professor Anthony Chemero
Anthony Chemero is Professor of Philosophy and Psychology and a core member of the Center for Cognition, Action, and Perception at the University of Cincinnati. His work is both philosophical and empirical and includes resources drawn from ecological psychology, phenomenology, artificial life and complex systems. He is author of more than 70 articles and the book *Radical Embodied Cognitive Science* (2009, MIT Press), which was a finalist for the Lakatos Award. His second book, *Phenomenology: An Introduction* (2015, Polity Press), co-authored with Stephan Kaufer, assesses the relevance of phenomenology to contemporary psychology, philosophy of mind and cognitive science. He is currently editing the second edition of the *MIT Encyclopedia of the Cognitive Sciences*.

Professor Fred Cummins
Fred Cummins is a cognitive scientist and Lecturer at the School of Computer Science, University College Dublin (UCD). His work focuses on joint or collective speech and vocal communication, covering a wide range of collective speech phenomena from prayer to protest. He has particular interests in coordination within and among people, in the links between rhythm, speech, and language, and in philosophical foundations such as enactive theory. He works from an enactive perspective, incorporating the role of the body and the environment in accounting for the phenomenon of collective speech. Professor Cummins has published numerous papers on collective speech and strives to develop a cognitive science that goes beyond the limitations of a Cartesian view of mind and matter and views his work as a contribution to the exploration of the thesis that mind and life are inseparable.

Questions for both Anthony Chemero and Fred Cummins:

Austin Dwyer and Meline Papazian (AD/MP): *What role do you see for philosophy in Cognitive Science? More specifically, what is your view on the contributions of different trends in philosophy, such as analytic philosophy of mind or phenomenology, to scientific studies of cognitive processes?*

Anthony Chemero (AC): I want to begin by saying what I think the role of philosophy should *not* be. In my 2009 book, I rant about what I call 'Hegelian arguments'. These are conceptual arguments against the very possibility of empirical research programs. Philosophers have traditionally loved arguments like these, and seem to love them still. Consider a few "classic" examples: Dreyfus and Searle argued against the possibility of AI; Fodor and Pylyshyn argued against the possibility of Gibsonian ecological psychology and against the possibility that connectionist networks could model cognition. I could go on with examples about original intentionality, qualia, incompleteness, unbridgeable intermediaries, etc. These arguments are the antithesis of 'contributions' and practicing scientists quite rightfully ignore them. If philosophers want to have an actual role in the cognitive sciences, they have to contribute in a positive way, and there are plenty of ways to do that. But to do so, they need to really dig in and understand the science. Only then is it possible to make the sort of contributions we should expect from good philosophy, such as conceptual clarifications, historical perspective, and novel hypotheses. This is true for both analytic and continental philosophers. Analytic philosophers of mind have a long history of contributing to the cognitive sciences, and were involved from the very beginning. Nowadays, it seems that the analytic philosophers who contribute most to the cognitive sciences are primarily philosophers of science.

Although they are often criticized by their analytic colleagues as being too vague or ignorant of science, continental-style philosophers can also contribute to the cognitive sciences. First, it is simply false that there is anything vague about the work of phenomenologists such as Husserl, early Heidegger, and Merleau-Ponty. These thinkers made very concrete and specific claims about the nature of experiences, many of which are testable in the lab. Second, many of the continental philosophers who work in the cognitive sciences are much more well-informed about science than their analytic colleagues. A good example of this is Evan Thompson: his philosophical positions are inspired by phenomenology and Indian philosophy, but he defends these positions using his detailed and up-to-date knowledge of neuroscience and the physics of self-organizing systems.

Fred Cummins (FC): Before there was a view that philosophy and science are different things, there was simply Natural Philosophy, which employed both abstract argument and empirical evidence in the framing and answering of substantial questions about what is, and what our place among that is. Cognitive science, to me, is better viewed as natural philosophy, than either as science or philosophy. It simply can't be done by separating the two enterprises. As well as asking "What do we see?", which is the sort of question that scientists are happy with, we must necessarily ask "What are we that we should see such?", which is much trickier, as we become embroiled in our own questioning.

Here, the nonce division between science and philosophy becomes more difficult, as political, theological, and legal questions impinge on our debates. Cognitive science cannot be done from an aseptic vantage point of pure disinterest, because we are negotiating our own being. This will, of course, have consequences for science, for philosophy, and for many many other spheres of human activity and self-description.

As for the territorial disputes within academic philosophy, I don't really find them interesting. There is analytic philosophy within cognitive science, and continental philosophy aplenty. We could probably do with a bit more critical theory, and the like, as we recognize the necessary links between the public view of "knowledge" and the structures of authority within education, medicine, and the state generally. And we could certainly do with a richer input from Daoist, Vedantic and Buddhist traditions, to counteract what I see as a very strong Abrahamic, and even Christian, core to contemporary science, and the psychological sciences in particular.

Phenomenology plays a very significant role in contemporary embodied and enactive approaches, but this is something that must be treated very sensitively. Early Husserl, in particular, seems to me to be thoroughly beholden to a set of solipsistic Christian assumptions about being and autonomy that I find very problematic. The Husserlian act of reduction has something in common with Descartes' flight from his senses, both of which bring into being a spectator subject who believes him or herself to be distinct from the world. Merleau-Ponty seems to be useful for embodied cognitive science, but that does bring the danger of mistaking bodies for people, which to my mind is not much better than mistaking brains for people.

AD/MP: *Some cognitive scientists reject the classical explanation of the interface between the mind and the world in terms of a relation between representations and things they represent. Could you tell us about your views on, and possible solutions to, what has come to be known as the problem of representation? What do you think of the suggestion that in some of our activities we use representations and/or computations and in some others we do not? Is it acceptable to say that we don't engage in representations in our daily interactions with the environment, whereas scientific work may require modelling, using representations, in order to explain these daily behaviours?*

AC: I can't see any reason to think that human thought and experience work just one way. Gilbert Ryle was right in *The Concept of Mind* when he said that lots of the words we use to describe the mind are success verbs. Ryle's point was that seeing is not a special, sparkly mental state; it is just successful looking. My point in bringing this up is less the 'success' than the 'verb'. Seeing, thinking, experiencing, remembering and the like are things that we do, and not things that happen to us (or, worse, to our brains). Just as jumping and reading are very different from one another, so too are seeing and remembering. In the case

of things like seeing, perceiving, choosing, which involve interacting with the world surrounding you, you don't need to build and maintain a representation of the things you see, perceive, and choose. They are there with you. In contrast, when you remember the number of windows on the front of the house you grew up in, you seem to be doing so in terms of a representation of the house, maybe a mental image. You can make the same point in Heidegger's terms: when we are smoothly coping with the ready-to-hand, we don't need representations; when things go wrong, and we are dealing with the unready-to-hand or present-at-hand, we might need representations.

FC: Nearly every big dispute that lands within our remit seems to revolve around differing assumptions about what the referent of the personal pronouns "I", and especially "we" are. These are the obligatory themes that will be negotiated, and I believe, must be continuously negotiated.

So for some value of "I" that has arisen within a tradition strongly oriented to the views of Descartes, Hume and Kant, there is a need to postulate "representations". One can view the person through this lens, and it is of some use. There are ways to frame our observations of humans that seem to require the notion of representation. But this particular orthodox tradition does not exhaust the person, and contemporary embodied and enactive accounts are often trying very hard to overcome the baggage of this specific tradition, which sees persons as atoms, each possessed of a single body and a single mind, where a transcendental ego has no direct knowledge of a transcendental world, where bodies are machines in a mechanical universe, and where minds are nowhere to be observed.

From the tone here, you can see that I'm not finding representation to be of much use myself in my own work, because I self-consciously address contexts, phenomena, and activities that speak of different ways of understanding subjects and their worlds. I'm happy for representational work and non-representational work to both go forward, funded, and interacting, but I find it lamentable when one side or another demands that they have the one true way to lay out the issues. Inevitably, much of the reality of the person and of lived lives gets lost in such debates.

At least, on a good day, I feel ecumenical along those lines. However when I look at the structure of contemporary science, and the sciences of the living and the person, in particular, I am often less generous and sanguine. The association of human experience with the activity of the brain alone is so pervasive that I feel that a more combative stance is necessary. Because the way we view the person is not merely a scientific question, but also a political one, an ethical one, a medical one, and a legal one, it is not easy to maintain a pleasant "can't we all just get along" demeanor. So on a bad day, I come out with both guns blazing. I admit that I like bad days too.

But rather than adding to the immense literature on the merits and demerits of "representation" in theories of perception, thought, cognition, etc, I would prefer to introduce the theme of "presence" instead. Presence is a term that speaks to the immediacy of the here and now, and that is strongest when we turn to co-presence, as in chanting or even collective silence. Here we need some new vocabulary. I think even the notion of "perception" needs to be abandoned as we turn to presence, for talk of "perception" invites a spectatorial view of being that separates a singular subject from an objectified world, while "presence" invites consideration of how subject(s) and world(s) co-arise. (The notion of "direct perception" beloved of ecological approaches attempts this, but by remaining tied to the term perception, it remains beholden to a subject that is then fissioned off from its world.)

AD/MP: *Where do you stand on debates about theoretical frameworks in Cognitive Science? Is a single framework possible and/or sufficient to explain both natural and artificial minds and cognitive processes? Alternatively, might theories of cognition be better explained through a pluralistic approach to the interaction between mind, body and the world? Do you have any preference for certain framework(s) over others?*

AC: As you might expect, given my answer to question 2, I am never a fan of –isms. Computationalism, for example, is the view that the brain is a computer and mental activity is computation. Full stop. Similarly for other views in cognitive science, like mechanism, connectionism, enactivism, and so on. These views all suggest that we engage with the world in just one way. But that is silly. See, reach for, grasp, and drink

from the coffee cup on your desk. Now, think about meeting friends for drinks this evening. I think it is obvious that you did two very different things, both probably successful, both richly experiential, but nonetheless dramatically different from one another. So: yes, pluralism.

That said, I do have theoretical preferences. I have always worked within a broadly ecological and dynamical research program. I prefer it for a lot of reasons, among which are moral and political reasons. But I don't pretend that I am doing the One True Psychology. My version of psychology is much better at explaining your interaction with the (ready-to-hand) coffee cup than it is at explaining your making plans for the future.

FC: Pluralism is absolutely essential if we are not to do violence to the subject matter, which is ourselves as subjects. One might argue that a single naturalistic framework might serve in physics, where measurement contexts are rigorously controlled and agreed by physicists before they debate data and models. I have my doubts, as there is plenty of contemporary physics that is non-reductionistic. So maybe, in physics. But once we enter the realm of the living, and I mean biology as well as any of the psychological and social "sciences", no justice can be done to the object of study without acknowledging that those objects are subjects with values. A single cell cannot be understood without consideration of the world from the perspective of the cell. How much more, then, as we approach human experience. Imposition of a single theoretical framework on such questions is, to me, no different than religious fundamentalism that insists one set of values must triumph over all others. So again, the question becomes political, demands negotiation, and resists dogmatism.

Questions for Anthony Chemero:

AD/MP: *You have suggested that the same explanatory strategy holds for a range of phenomena, from brain areas coordinating with each other all the way to humans interacting with one another and that this strategy is not computational or representational, but is based on a dynamical model. Also, after the coordination process, all elements involved go back to the way they acted before. How would you apply this account in social cognition? Also, how would you explain this model scientifically?*

AC: The answers to these questions are pretty complicated, and require a brief digression into complex systems theory. Sorry... There is a lot of work in complex systems that indicates that collections of entities that are far from an equilibrium state often form what are called "synergies". A synergy is a collection of entities that is temporarily constrained to act as a single unit. The notion comes from mid-20th century physics, and was initially used to explain how spontaneous order appears in systems like lasers. You don't need a laser to see a synergy, though. Go look at your toilet. The water molecules in the bowl are at energetic equilibrium with their surrounds, and each water molecule moves around more or less randomly. When you flush the toilet, though, you release the potential energy of the water stored in the tank, and create an energy gradient, which is to say that the water in the bowl is no longer at equilibrium with its surroundings. The whirlpool that forms in the bowl is a result of this energy gradient. It appears not because all the water molecules decided to move in helical patterns rather than randomly, but because their activity was constrained by being part of a synergy. The synergy exists for only as long as the energy gradient exists—when the water has run out of the tank, the molecules will go back to their more or less random motions. It's important to realize that synergies are mundane, and appear all over nature. In fact, every cell in your body is a synergy. Cells are collections of parts that are constrained to work together temporarily. Cells are not at equilibrium with their surrounds—the semi-permeable cell membrane guarantees that, and keeps the cell far from equilibrium for extended periods of time. If a cell wall fails, the chemicals that made up the cell will disperse toward equilibrium with their surrounds. Organs are synergies, where the components temporarily constrained to form a unit are cells; organisms are synergies, where the components temporarily constrained to from a unit are organs. When organisms are at chemical and thermal equilibrium with their surrounds, they are no longer living.

Complex systems scientists have developed mathematical and experimental techniques to study synergies in the physical world, in biological systems, and in physiology. For the last several decades, cognitive scientists have been applying these techniques to the brain and behavior, including social behavior. Fred Cummins has done important work here, as have my colleagues Mike Riley and Mike Richardson at the University of Cincinnati. I have only recently gotten involved in experiments with Mike and Mike on interpersonal synergies, especially when people play video games or improvise music together. I hope to join with Fred for some experiments on joint speech in the near future. The basic idea in this work is a straightforward application of synergies. Just as the water molecules are temporarily constrained to form a whirlpool by an energy gradient, groups of people can be temporarily constrained to form synergies by task demands. The exciting part of this is that the same mathematical tools that apply to whirlpools and lasers also apply to groups of humans.

AD/MP: *You have shown a willingness to engage with phenomenological writers in your work. You have mentioned Merleau-Ponty as a role model in appreciating interdisciplinary work. You have also specifically applied the Heideggerian structures of ready-to-hand and unready-to-hand in your work. With respect to the scientific study of experience and social cognition in particular, how would you incorporate phenomenology into your version of radical embodied cognitive science?*

AC: This has actually been an important focus of my research recently. I have been very interested in what Heidegger called 'being-with', and whether we can study it scientifically in terms of interpersonal synergies. Heidegger said that part of the essence of being a subject of experience is being-with others. This being-with others goes well beyond the fact that we are typically near other people and able to think about them; instead, the very possibility of experience of and engagement with the world requires implicit awareness that others can experience and engage with it in similar ways.

We can think of the importance of being-with in terms of the traditional philosophical problem of other minds: how do I know what anyone else is thinking, or even that they are thinking, when all I can see is their behavior? It seems that I can't. However this problem can't arise if the very possibility of my experience depends upon the possibility of experiences for others. The very fact I can have experiences, including the experience of wondering whether other people have experiences at all, requires that others could wonder the same thing. If Heidegger is right that being-with in this sense is required for experience of the kind that we have, we should wonder how we are-with others and what it requires. Importantly, we are not with others in virtue of consciously thinking about their experiences and what they are like. Being-with requires that we are able to implicitly and unreflectively be aware of how others experience.

The ease with which humans participate in interpersonal synergies, and that they do not explicitly intend to coordinate with others, suggests that we are implicitly and unreflectively aware of how others experience. Remember the last time you danced or jogged with someone: you and your partner formed a temporary unit, in which each of you allowed your behavior to be constrained by the interpersonal synergy you were part of. This required each of you to adjust your behavior so that you could work together. Recent experimental work has shown similar interpersonal synergies arising when people sit next to one another on rocking chairs, when people converse, when people collaborate in novel video games, and when professional jazz musicians improvise together. As I mentioned above, these last two cases are from my own recent empirical research, in collaboration with students Ashley Walton, Patrick Nalepka, Chris Riehm, and my colleagues Mike Richardson and Peter Langland-Hassan.

This is how I have been incorporating phenomenology into radical embodied cognitive science: I have been using the mathematical and experimental techniques from complex systems science to try to explore concepts from phenomenology, such as being-with. In earlier work, with Dobri Dotov and Lin Nie, we have studied the transition from readiness-to-hand to unreadiness-to-hand.

AD/MP: *Your view of affordances is that they do not belong to the environment or to the agent, 'they are relations between abilities and some features of a situation', that are 'perfectly real and perceivable'. What is the relation between perceiving affordance and experience? Also, is our understanding of others at least partly due to the fact that they provide affordance for this understanding?*

AC: I would say that perceiving affordances is having experiences, but that not all experiences involve perceiving affordances. Affordances are tricky, and it is easy to expand the notion of 'affordance' so that art works afford something like visual-enjoyment-while-pondering-the-history-of-depictions-of-the-virgin-Mary. This would make all experiences perceiving of affordances, but would stretch the concept well beyond situations in which it really is useful. Pluralism in the cognitive sciences is, among other things, a refusal to stretch notions beyond their usefulness in order to defend an –ism.

Even without this stretching, the term 'affordance' is used in lots of different ways. For example, most classical experimental studies of affordances involve yes-or-no decision tasks. Participants are asked "Can you climb that step?" or "Can you fit through that doorway?" for a series of differently sized steps or doorways. Their yeses and nos define a body-scaled boundary at which they perceive the presence of climbability or passability. I would argue that most affordances that we perceive are not like this, and even making up new ugly English words does not suffice to describe them. Consider: how do you adjust your gait when you walk out of the carpeted entryway and onto the icy sidewalk? When we smoothly and unreflectively change the way we walk on a surface, we are perceiving and acting on affordances. Another example: Right now, stop reading this, go to your kitchen, and open a jar. (Seriously: go ahead. I'll wait.) Before this time of opening the jar, did you realize that you began with a full hand, power grip, and then switched to a fingertips-only precision grip? In order to do that, you perceived and acted on affordances for those different styles of action. This is having an experience. I would argue that climbability and grip-switching are cases in which the idea of affordances has real explanatory value, but discussing the affordances of paintings in museums is much less useful. One way to say this, but one that would take a long time to defend adequately, is to say that perceiving affordances is experiencing the ready-to-hand.

The second part of this question is about social cognition and social affordances. For the last few questions, I discussed interpersonal synergies and being-with. In order to participate in an interpersonal synergy, people perceive affordances for subtly adjusting their behavior so that it fits with what another person is doing; they also continuously perceive and act upon new affordances made available by the simultaneous adjustments that their partner is making to fit with them. Again, consider having a conversation with someone you are jogging with. Both joggers subtly adjust and re-adjust their speed and step length to maintain a fluid and variable coordination that makes a conversation possible. We also perceive affordances for interacting with other people, based upon uncountably many things. We perceive others as affording a passing nod, a vocal greeting, a request for directions, and the like. These are more like the yes-or-no classic affordances that have ugly English names. (Brian Eiler, a PhD student at University of Cincinnati is doing research on the factors that determine the affordance intractability.) As with the case of the painting in the museum, I think that we stretch the notion of affordance beyond its usefulness if we think about someone as affording a-gift-from-me-of-tickets-to-see-a-black-metal-concert or a-proposal-of-marriage-after-a-few-more-months-of-dating.

Questions for Fred Cummins:

AD/MP: *Will the problem of agency become more and more significant in an embodied-enactive cognitive science? How does agency feature in instances of closely coupled joint action?*

FC: That's two questions. To the first, yes, clearly, agency is precisely the nub of the issue when we begin to negotiate subjectivities. If subjects were simple, discrete, and entirely autonomous, there would be nothing to discuss, but they are not simple, discrete, or entirely autonomous. I am increasingly frustrated at scientific approaches to behaviour that act as if subjects were simply given, and worse, as if subjects

were automata of various kinds. For embodied cognitive science, there is the danger of elevating the body to the sole role of subject, and hence locus of agency. That is a mistake I fear we are currently encountering frequently. More intellectualist, rationalist accounts tend to place agency in the brain/mind, whatever that means. Both run the danger of essentialism with respect to agency, and that seems just wrong to me. I am very concerned that we don't reduce the subject to a single naturalised entity, and the same concern applies to agency, for agency and subjectivity are bound at the hip.

Note this is not some kind of poorly disciplined relativism. It is rather a recognition that in identifying agents, and subjectivities, there is an ineluctable ethical charge to our framing observations that we must acknowledge.

In closely coupled joint action we see very clearly that participants are not distinguishable from each other, that no individual entity (mind, brain, person) controls the unfolding of movement, and that manifest action is a joint achievement. We don't need fancy scanners or squid arrays to see this, we can see it in a handshake, a reciprocal smile, or a conversation. Here, dynamical systems provides a technical vocabulary that does not condemn us to the attribution of agency to a single domain, but provides us with a means for acknowledging emergent phenomena that must be taken seriously in their own right. That is not a solution though; it is only a starting point. The work starts here.

AD/MP: *You have conducted significant research in joint action particularly joint speech, what do you consider the most important empirical as well as philosophical insights you have been able to draw from this research?*

FC: At the moment I feel like a kid in an Aladdin's Cave, having stumbled onto a trove of phenomena that demand attention, that can radically inform us about our own nature, and that are simply fascinating. And they are all in full public view. The absence of scientific research in the area of joint speech is, to me, bizarre. I estimate there is about 10 times more scholarly work done on the marginal, rare, and highly context bound phenomenon of speaking in tongues than has been done on unison speech, yet the latter is *billions* of times more common (7 billion people, everybody takes part in joint speech regularly), and is potentially informative about how we come to use our voices to establish a common ground from which a common order emerges.

As a coordinative act, speaking in synchrony challenges our notion of coordinated action and demands that we acknowledge a far greater nomological relation between speakers than is available with present models. Gibsonians will be comfortable with the manner in which speech of two or more people becomes entangled, but cognitivists will be in trouble. This was the aspect of speaking in unison that first got my attention: a highly plastic channel: the voice, that we adapt flexibly to a wide variety of contexts, interlocutors, and emotional charge, can be disciplined to fall in precise lockstep with the voice of another. There is no beat, as in music and dance. There is no physical scaffolding, such as that which permits tight synchronisation of rowers, divers or trampoliners. And this happens without practice, and with ease.

But if the "physical" act of synchronisation is rich, how much richer is the embedding of joint speech practices in our lives! Prayer, protest, and the enactment of the collective identity of football supporters make a merry bunch, and they positively shout to us of the importance of collective intentions, collective sentiments, collective subjects. Each domain has its own characteristics. Prayer leads to the study of rituals and their role in establishing a stable common order. Protest has a naked urgency to it. Sports chanting has its own distinctive features in which we see a blending of tribal and more ephemeral characteristics. But there are many facets to joint speech that transcend these domains. Shall I begin a list: the obliteration of speaker/listener distinctions, the obliteration of a speech/music distinction, the role of musically tinged prosody in the stereotypical intonation patterns and the rhythmically enhanced beats, the redundancy and repetition, over and over, the stability of idiosyncratic forms that recur across diverse situations, and even across centuries, the use of common structures such as various kinds of call and response, or the generalised assent of "amen" or "hell, yeah!", and above all the participatory nature of joint speech, in which speakers acquire commitments and obligations through the act of uttering. I could go on, but there is enough here, I hope, to suggest that this is a rich enough playground to keep me and many others happy for a long time to come.

But best of all, this empirical cornucopia challenges our notion of the subject, of intentionality, of agency, and of who we think we are. It is ideal for cognitive science as a locus at which we must ask what we can aspire to say about ourselves.

AD/MP: *You state that social interaction is 'the regulated coupling between two autonomous agents' and, by applying dynamical coupling concepts, you extend the notion of 'participatory sense making' suggested by De Jaegher & Di Paolo and state that two synchronising agents create a higher level domain of coordination mutually. What does your work and your view add to the existing notion?*

FC: The experimental object of two speakers in synchrony illustrates the manner in which the activity of two individuals becomes inexorably fused, leading to the emergence of a higher level domain, the dyad, that has its own conditions of persistence. This is clearly revealed in the regular occurrence of a specific kind of speech error, in which both speakers abruptly stop speaking at the same moment, usually in response to a prior uncertainty or minor speech error. The abrupt cessation is not found in conversational or monological speech, but it occurs simultaneously when two speakers are tightly synchronised. I can find no language to begin to express this phenomenon if we couch our description of action in the highly individualistic vocabulary of cognitivism, but it appears as a wonderfully illustrative example of the emergent view of joint action.

That particular observation is made at the level of speech production. But more interesting questions arise about how we tackle meaning, when the subject doing the uttering must be understood as collective. Most formal theories of meaning will not be able to accommodate any such notion, as they have a priori commitments to individual minds, and individual systems within which meaning is expressed. Here I too am in need of novel concepts, new ways of talking about the collective creation of meaning, and I see the formal device of participatory sense making as a useful first step. There is a long way to go though.

Concluding Common Question:

AD/MP: *In conclusion, what do you think are the central issues in cognitive science today and how does your research relate to these issues?*

AC: I think the most important issue in cognitive science right now is whether it will continue to exist as a vibrant area of research, or whether neuroscience will just take everything over. For a variety of reasons, neuroscience seems much more interesting to many more people right now than psychology does. I have been trying to make people realize that neuroscience without psychology is empty. Of course, your brain is important and of course your brain does different things when you think and do different things. That does not mean, though, that the thinking and doing of these things is really nothing other than a brain event. (To really see this, Fred's recent blog post on the-brain-in-the-ass is worth reading several times.) My whole career has been an argument that we can't replace psychology or cognitive science with neuroscience. Moreover, even if we really just want to understand human brains, we will fail to do so if we forget that human brains always exist in human bodies and bodies always exist in rich and complex behavioral, social, linguistic, and cultural environments.

FC: I have unorthodox views about where cognitive science is going. I would like to see it distance itself from the business of psychology. Scientific theories of psychology have become fixated on the problematic notion of individual cognitive mechanisms. This does violence to the person. It rests on metaphysical assumptions that are specific to Abrahamic theology, which is a problem as long as we don't acknowledge it. Let's leave "Christian Science" to those who call themselves Christian Scientists. Most importantly, it blinds us to our collective nature, and our collective responsibility in bringing a shared world into being. It also creates a kind of impoverished approach to social phenomena which are seen as non-psychological. The social sciences rest on very shaky foundations, but the belief in a positivist scientific or cognitive psychology is

one reason for this. Of course, that makes this discussion inherently political, which it should be as we address our own nature. Given the kinds of tumultuous societal developments and the enormous problems of planetary husbandry we face, we will be forced to face up to very hard questions about our selves that we are currently ill-equipped to even begin to discuss.

I am developing a fondness for the non-scientific kinds of psychology, Freudian, Jungian, and beyond that do not pretend to fix an objective characterisation of the person and their activities. Don't get me wrong, we need rich stories about our individual selves. That is psychology, and we need it. It just isn't the kind of science it presents itself as, and it shouldn't aspire to being that kind of science. So theories of embodiment and enaction should stop casting themselves as fixer-uppers for cognitive psychology.

In the long run, we need to revisit what we think science is, and allow for a less positivist kind of account when we discuss the activities and values of the living. This essentially means taking the insights of mind and life seriously. Robert Rosen has suggested that we stop regarding biology as a special sub-domain of physics, and regard physics as a special sub-domain that arises from the activity of the living. I am wholly sympathetic to that notion.

Research Article **Open Access**

Genevieve Hayman

Mirror Neurons, Husserl, and Enactivism: An Analysis of Phenomenological Compatibility

DOI 10.1515/pipjp-2016-0003

Abstract: The potential for mirror neuron research to explain various aspects of social cognition has received considerable attention over the past two decades. Initially, mirror neuron research may seem in accordance with a phenomenological understanding of intersubjectivity, but the work of Dan Zahavi will be used to highlight significant incompatibilities between the two. Likewise, the enactivists Thomas Fuchs and Hanne De Jaegher identify significant issues with current interpretations of mirror neuron research and provide an alternative description of intersubjectivity. This article will assess whether the enactivists are able to provide a more phenomenologically consistent alternative to mirror neuron research alone, eventually determining that their enactive account overcomes Zahavi's incompatibilities. Consequently, Fuchs and De Jaegher should acknowledge their relation to Husserlian descriptions of empathy in their account, and mirror neuron research should be contextualised within a broader, phenomenologically-compatible framework, as that of the enactivists.

Keywords: mirror neurons; Husserl; empathy; social cognition; intersubjectivity; enactivism

Preface

Since the initial studies of F5 neurons in macaque monkeys in the late 1980s[1] and the identification of these neurons as 'mirror neurons' in the mid-1990s[2], the topics of social cognition and intersubjectivity have been dominated by discussion of mirror neurons. This focus stems from the possibility that mirror neuron research can provide the best explanation for the way we interact with each other. Emotional empathy, action understanding, and mind-reading are only a few of the topics that have been attributed to neurological activation in areas with these supposed mirror neurons, and many scholars have reacted strongly:

> The discovery of mirror neurons in the frontal lobes of monkeys, and their potential relevance to human brain evolution... is the single most important unreported (or at least, unpublicized) story of the decade. I predict that mirror neurons will do for psychology what DNA did for biology: they will provide a unifying framework and help explain a host of mental abilities that have hitherto remained mysterious and inaccessible to experiments (Ramachandran, 2000, para. 1).

Indeed, much significance has been attributed to mirror neuron research due to the supposed ability for this research to justify accounts of intersubjectivity as described by psychology as well as other fields, thus transcending the disciplinary divisions. Within philosophic practice in particular, intersubjectivity has been a recurring topic that has long been contemplated by various prominent academics. One such

1 See Rizzolatti et al. (1988).

2 See Gallese et al. (1996).

***Corresponding author: Genevieve Hayman,** University College Dublin

academic is the 19th century German philosopher Edmund Husserl who expanded on the role of empathy (*Einfühlung*) in the constitution of ourselves and our 'lifeworld' (*Lebenswelt*). Known as the founding father of phenomenology, Husserl's phenomenological work on empathy had a strong influence on later philosophers, such as Edith Stein and Maurice Merleau-Ponty. If a phenomenological account is meant to capture the essence of experience, as was established by Husserl as the purpose of his transcendental method,[3] then the claim that mirror neuron theories support a phenomenological description is indeed noteworthy. This article will thus highlight a few testimonies in which Husserl's accounts of intersubjectivity and empathy seem in agreement with the intersubjectivity described by many mirror neuron studies. However, shortly thereafter, we will turn to the work of Dan Zahavi to discover significant problems with this association. In response, we will shift to the enactivists Thomas Fuchs and Hanne De Jaegher for their own criticisms of mirror-neuron-based theories and explain why they believe their enactive approach to be superior to mirror neuron research alone in describing intersubjectivity. Finally, we will see if their alternative is indeed an improvement insofar as it is able to overcome the concerns highlighted by Zahavi, and thus, prove itself more consistent with Husserl's descriptions of intersubjectivity.[4] If this is determined to be the case, the implications of this article are twofold: 1) The enactivists can and should acknowledge Husserl's descriptions in their accounts, and 2) If one wants to provide explanations of intersubjectivity true to phenomenological descriptions, mirror neuron research must be contextualised within a more accommodating framework, such as that of Fuchs and De Jaegher's enactivist account.

I. A Brief Introduction to Mirror Neurons

According to Gallese, Keysers, and Rizzolatti (2004), the neurological system that comprises mirror neuron activity serves as "the fundamental mechanism at the basis of the experiential understanding of others' actions" (p. 396). Their reasoning is as follows: Because two individuals have similar brain-body systems, there can be a process of 'mirroring' or 'simulation' between areas of one individual's neuronal system and areas of another individual's neuronal system – these respective brain areas are referred to as 'mirror neurons'. The authors find support for their claim in experimental evidence, concluding that mirror neurons permit the "direct experiential grasp of the mind of others" (Gallese et al., 2004, p. 396).

The topic of mirror neurons has captured the attention of researchers because these groups of cells exhibit activation not only during the performance of a movement, but also during the observation of another performing that same movement. According to a prominent early study by Gallese et al. (1996), approximately 17 per cent of cells in the F5 area of macaque monkeys qualify as mirror neurons,[5] and about one-third are selective for one type of action. Specifically, mirror neurons have been shown to respond when either performing *or observing* goal-directed actions (e.g., picking up an object and placing it somewhere) (Rizzolatti et al., 1996).[6] As such, mirror neuron research has been linked to action understanding, such that understanding another's action need not require higher-level interpretation by a central conceptual system but persists as a direct result of the observation itself, or what Gallese et al. (2004) describe as "the 'penetration' of visual information into the experiential ('first person') motor knowledge of the observer" (p. 396).

Indeed, turning to the wealth of empirical data, one finds further evidence suggesting that the observation of another's goal-directed action can result in the activation of the same motor circuits in both the performer and the observer. Umilta et al. (2001), for example, conducted a study in which action-

3 For example, see Husserl, 1997, p. 118-119.

4 Husserl's writings on empathy spanned his career, and as such, his portrayal of intersubjectivity should not be understood as complete nor coherent. Indeed, many scholars identify drastic shifts in perspective from early works such as *Logische Untersuchungen* and *Ideen I*, which work to lay the foundation for his transcendental phenomenology, to *Die Krisis* in which he expands upon the constitution of a 'lifeworld'. However, this article will overlook these transitions and extract limited yet central ideas relevant to the collection of Husserl's writings on interpersonal experience.

5 Subsequent research has expanded the scope of mirror neurons to include other areas such as the parietal cortex (Gallese et al., 2002) as well as the primary motor (M1) and dorsal premotor (PMd) regions (Tkach et al., 2007).

6 Activity in these cells is not exhibited in cases of pantomime (lacking an object to grasp) nor in the observation of objects alone (lacking hand movement) (Rizzolatti et al., 1996).

specific mirror neuron activation was triggered even when the final part of the act (i.e., 'goal') was occluded. The experimental design involved four conditions:

A) The experimenter's hand moves towards and grasps an object in full view
B) The experimenter's hand moves towards and grasps an object behind a screen
C) The experimenter's hand moves towards and grasps an imaginary object (miming)
D) The experimenter's hand moves towards and grasps an imaginary object behind a screen (hidden miming)

Results revealed that the majority of scrutinised neurons showed activation in both A) and B) but not in C) or D). Therefore, Umilta et al. (2001) concluded that because mirror neuron activation need not depend on the *direct* observation of an action and its goal, the activation seen in these areas also indicates a level of action *understanding* necessary to infer the presence of an object behind a screen and discriminate against cases in which the occluded object is not present. In another example, Fogassi et al. (2005) found that activation in the inferior parietal lobe (IPL) was significantly distinct depending on the type of action observed (grasping-for-eating versus grasping-for-placing). Moreover, activations were selective *during the grasping action* (before completion of the task), indicating that the relevant IPL cells were sensitive not only to the specific action, but also to the specific *intention* of the act. In other words, these two studies seem to indicate that the ability of an observer to understand another's action and the intent of that action (i.e., to empathise) is attributable to similar mirror neuron activity for both observer and performer – the assumption being that similar neuronal activity entails similar experiences.

Of course, the relevance of these studies and their respective conclusions are debated within academic literature.[7] Because most of the empirical evidence arrives through experimentation with monkeys, there exists much debate over the presence of mirror neurons in humans. Indeed, most of the scientific evidence for humans employs indirect methodologies such as fMRI studies. Nevertheless, the majority of research indicates similar cortical activation of 'mirror neurons' in humans as well as in monkeys.[8] While this article will not attempt to resolve this dispute, it is important to note that mirror neuron activity is a discovery that has redirected the ways which the scientific community discusses and understands interpersonal interaction, observation, and empathy. Thus, the rest of this article will take for granted the debated scientific validity of mirror neuron theories and their controversial applicability to humans, instead focusing on the philosophical implications of mirror neuron research on intersubjectivity.

II. Husserl and Mirror Neurons

At first glance, the aforementioned mirror neuron research seems quite in alignment with Husserlian empathy:

> In fact, Husserl even writes, and this does sound remarkably like formulations found in Gallese, when I perceive the movement of the foreign body, it is *as if* I were over there, as if I were moving my limbs (Hua 15/642, 4/164). When I see the foreign hand, I feel my own hand. If I see the other hand move, I am inclined to move my own hand. My own kinesthetic system is affected by my perception of his moving body and by my anticipation of his future movements (Hua 14/527, 15/642) (Zahavi, 2011, p. 239-240).

Zahavi is far from the only philosopher to recognise this similarity. Others include Matthew Ratcliffe (2009) who contends that important parallels can be drawn between the intersubjectivity described by Husserl as a pre-reflective and non-inferential 'analogising apperception', and mirror neuron research, which claims a direct link between observing another's actions and one's own experience without the need for higher-level processing. He cites the work of Gallagher (2001) in claiming that "neuroscientific findings can provide support for Husserl and can also be integrated into the interpretation of phenomenological descriptions by clarifying the kind of relation described and showing how it need not be something

7 See Hickok (2009) and Dinstein, Thomas, Behrmann & Heeger (2008).
8 See Kilner & Lemon (2013).

mysterious or even impossible" (Ratcliffe, 2009, p. 339). Additionally, he goes on to say how this support is reciprocated, whereby Husserlian phenomenology can justify the neuroscientific research by providing "a phenomenological framework in which the role of mirror neurons can be conceptualized, interpreted, and explained" (Ratcliffe, 2009, p. 336). Even Evan Thompson, who has conducted much work under the banner of enactivism, draws a similar connection. With regard to mirror neuron research, he states: "...this kind of non-inferential bodily pairing of self and other is one of the hallmarks of the phenomenological analysis of empathy. Indeed, the mirror neuron findings support Husserl's position that our empathetic experience of another depends on one's 'coupling' or 'pairing' with the other...rather than some kind of affective fusion, as some of Husserl's contemporaries held" (Thompson, 2001, p. 9).

However, we will now turn to a more comprehensive view of Zahavi's work to see if the relation between mirror neurons and Husserl is really as compatible as it seems. In his 2011 paper, 'Empathy and mirroring: Husserl and Gallese', Zahavi analyses whether mirror neuron accounts of empathetic action understanding are actually in accordance with Husserlian phenomenology, and he concludes that "unfortunately, the question is too complex to really allow for a simple yes or no answer" (p. 245). We have already seen that Zahavi recognises similarities between mirror neuron theories of intersubjectivity and empathy as described by Husserl. However, his ambivalence towards this relationship arrives through the acknowledgment of equally important distinctions between the two.

First, according to Zahavi, the explanatory scope of mirror neuron activity must be evaluated: To what extent can mirror neuron research explain intersubjectivity?[9] This is particularly important for any comparison with a Husserlian description of empathy because, for Husserl, interpersonal understanding persisted at a variety of levels:

> In a manuscript from 1931-32, he operates with even more levels. The first level of empathy is the appresentation of the foreign lived body as sensing and perceiving. The second level is the appresentation of the other as physically acting, say, moving, pushing, or carrying something. The third level goes beyond this and sees, say, the running of the other in the forest as flight, the hiding behind a stone as protection from missiles, etc. (Hua 15/435). On a few occasions, Husserl goes even further and also speaks of the kinds of empathy involved in apprehending the unity of a normal community and in appropriating foreign traditions (Hua 15/436, HuaM 8/372-373) (Zahavi, 2011, p. 242).

Even if mirror neuron research indicates the passive and involuntary pairing of self with other, this should not be misinterpreted as portraying the full extent of interpersonal understanding. Indeed, the level of action understanding claimed in the aforementioned mirror neuron studies would be insufficient in capturing the entire scope of possible meanings for any given action. After all, goal-oriented actions are embedded within a framework of meaning that extends far beyond the physical movement itself. Zahavi (2011) makes the point that the givenness of the other is not simply about the actions and experiences of the other, but rather, "the other is given as intentional, as directed at the same world as I, and the other's world, and the objects that are there for him, is given along with the other (Hua 14/140, 14/287, 13/411, 4/168, 1/154)" (p. 243). As a consequence of this robust understanding of intersubjectivity, the self-other relation works to constitute not only the other and her experiences, but also the lived world more generally, and it is for this reason that "the plausibility of the mirror neuron hypothesis increases in reverse proportion to its alleged explanatory scope" (Zahavi, 2011, p. 247). In other words, the more heavily we rely on descriptions of mirror neurons to explain various aspects of social cognition, the less likely it is that mirror neuron research can sufficiently support the complexity required to bear relevance to Husserl's multi-layered descriptions of empathy and the role intersubjectivity plays in our understanding of the world.

Zahavi's (2011) second criticism of mirror neuron accounts is their frequent affiliation with simulation theory (p. 247), which can be described as the "attribution of mental states via emulation or replication of states of the other" (Jensen & Moran, 2012, p. 2). According to Zahavi (2011) this understanding of intersubjectivity implies a form of projectivism whereby we assume one cannot access the experience of another, so we must rely on our own internal simulations to transform our perceptual input into what

9 In fact, the explanatory scope of mirror neuron research comprises a large part of the debate regarding the validity of many mirror-neuron-based accounts of social cognition.

we ascribe as the 'mental states' of others (p. 247). Though most mirror neuron accounts contrast with an understanding of simulation theory in which an agent *deliberately* or *rationally* places herself in another's shoes, mirror neurons are often described within scientific literature as facilitating a direct and automatic simulation of another's actions and intentions. This is clearly seen in sections of Gallese (2009) in which he claims that mirror neurons provide "a parsimonious solution to the problem of translating the results of the visual analysis of an observed movement – in principle, devoid of meaning for the observer – into something that the observer is able to understand" (p. 520 - 521). Essentially, even if we inadvertently convert visual observations into internal simulations that are projected or attributed onto the other, Zahavi maintains that such an understanding would be inconsistent with Husserlian phenomenology.

Thirdly, even if one is to describe the implications of mirror neuron activation such that simulation and projectionism are not involved, Zahavi claims that the concept of mirroring itself is not in line with Husserlian thought, and he gives several reasons. One is that to explain the self-other relationship as 'mirroring' would be too static of a concept. It would not permit the "dynamic and dialectical intertwinement between self and other," which contrastingly, Husserl's texts promote (Zahavi, 2011, p. 248). But another reason Zahavi provides is that theories involving mirror neurons often fail to respect the other as an other, while for Husserl, his phenomenology emphasises the alterity of the other. Indeed, even Gallese (2009) concedes this idea when he states that "the 'objectual other' becomes 'another self,' a like-me, who nevertheless preserves his or her alterity character" (p. 527) as well as when he later admits that the mirror metaphor is potentially misleading as it implies a direct match between the observer and the 'other' being observed (p. 531). Nonetheless, it remains to be seen how sufficient appreciation for 'the other' can stand in accordance with the majority of mirror-neuron-based descriptions of intersubjectivity and social cognition found within academic literature, including that of Gallese.

Lastly, Zahavi (2011) highlights the difference between personal and subpersonal accounts of empathy, with Husserlian phenomenology adhering to the former and mirror neuron theories to the latter (p.248-249). In other words, Zahavi points out that the two actually describe different things – Husserl describes empathy at the level of the person or agent, whereas mirror neuron research accounts for empathy at the level of neurological activity. Any link from the one to the other is thus difficult to draw without relying on claims such as isomorphism or symbolic representation[10], which would be controversial and require much additional justification.

III. An Enactive Criticism of Mirror-Neuron-Based Theories

In their paper, 'Enactive intersubjectivity: Participatory sense-making and mutual incorporation,' Thomas Fuchs and Hanne De Jaegher (2009) describe this mirror neuron research as favouring "a third-person paradigm of social cognition as a passive observation of others' behaviour, based upon an inner modelling process in the individual brain" (p. 466). The authors see it as problematic that there exists no interaction between the two individuals beyond passive observation and direct yet internal simulation from the observer, and moreover, that a framework is presupposed in which interaction is first described at the level of (neuronal) simulation, which is then used to explain or even predict the interaction itself. In contrast, Fuchs and De Jaegher (2009) propose that, "social understanding is not realised by 'snapshot' activities of one individual's theorizing or simulating but *arises in the moment-to-moment interaction of two subjects*" (p. 466). This moment-to-moment interaction is a dynamical and embodied process consistent with an enactivist approach[11] and may include interactions such as the coordination of gestures or vocal expressions, bodily resonance, and affect attunement. Furthermore, the authors claim that their understanding of social cognition relies heavily on Merleau-Ponty's concept of 'mutual incorporation'. More will be said on this in the coming sections, but it is important to currently acknowledge that Fuchs

10 See Marr (1982), for example.

11 Of course, this is but one of many descriptions of social cognition that fall under the varied label of enactivism. Other accounts should certainly be considered but are beyond the scope of this particular article. See Di Paolo, Rohde, & De Jaegher (2010) for a discussion of the varieties of enactivist accounts of social cognition.

and De Jaegher (2009) ultimately describe their approach as appreciating both the dynamic coupling of two agents and a phenomenological perspective:

> Combining a dynamical agentive systems perspective with a phenomenological perspective will allow us to link two sides of the same process – the interaction. The dynamical agentive systems approach observes and describes the interaction as a coordination process between intentional and embodied agents...The *phenomenological approach* takes an immersive perspective, starting from a first- and second-person take on the same process and describing the experience of the mutual engagement in phenomenological terms (p. 466-467).

Again, one may argue that mirror neuron research describes (or is at least consistent with) a phenomenological perspective. However, Fuchs and De Jaegher's main issue with mirror neuron accounts is that they provide subpersonal understandings of intersubjectivity, and as such, there is no subject to interpret the simulation or mirroring. The matching cortical activations described in the scientific literature are the direct result of action observation, and thus, do not require higher-level processing (indeed, that seems to be their primary appeal). Yet this renders such cortical activation no more than a description of neurological activity that is impartial at the subject level. After all, the core of the claim is that the same mirror neuron activation is observed whether performing an act or watching another perform that act, and as such, mirror neurons could not then discriminate between one agent or another. In fact, Fuchs and De Jaegher hold that mirror neurons do not actually 'simulate' at all since, in contrast, that would require the intentionality of a subject. It is in this sense that the authors believe mirror neuron research, while perhaps relevant, cannot do the heavy lifting when it comes to explaining intersubjectivity. Indeed, "there are no interacting minds or brains but only interacting living bodies or persons" (Fuchs & De Jaegher, 2009, p. 468). It is because of their less reductive understanding of intersubjectivity that Fuchs and De Jaegher believe their account to be more in alignment with phenomenology than mirror neuron accounts alone.

IV. An Enactive Account

Now we shall return to the idea of 'mutual incorporation', which Fuchs and De Jaegher (2009) describe as the reciprocal interaction of two embodied agents such that "their body schemas and body experiences expand and, in a certain way, incorporate the perceived body of the other" (p. 472). In order to begin a discussion of mutual incorporation, one should first develop an understanding of 'unidirectional incorporation', which describes the way in which objects or instruments become extensions of one's body schema. Unidirectional incorporation is an idea attributed to Merleau-Ponty who provides several notable examples, including the blind man whose stick "has ceased to be an object for him, and is no longer perceived for itself" (Merleau-Ponty, 1962, p. 165). In this example, the hand of the blind man feels *through* the stick and is guided by the ways in which that stick (of a certain shape and dimension) interacts with the environment as he moves it. There is thus an intimate connection between sensation and movement such that organism and environment co-constitute each other.

With this unidirectional incorporation comes 'operative intentionality'[12], which Fuchs and De Jaegher (2009) interpret as "the prereflective meaningful connection that the body establishes with its environment, based on the inherent connection of perception and action" (p. 475). This is a kind of Merleau-Pontean sensorimotor coherence – for example, the way in which a tennis player's body coordinates with the incoming ball, allowing her arm to respond to the call of the ball, or as Merleau-Ponty might say, the being-towards-the-ball through the body. The authors claim that such connection can be attained not only with inanimate objects, but also with other individuals, so that the observer becomes fascinated by the performer, and the observer's lived body can be described as 'being-towards' the other. Like the blind man's stick, the other is subsumed into the observer's lived body as distinctions of embodiment disintegrate and

12 "The concept of operative intentionality (*fungierende Intentionalität*) found only in the form of mere hints in Husserl's later writings has been made much of by later phenomenologists. In fact, one may say that all the above concepts of intentionality – the concepts of passive synthesis, genetic constitution, horizon intentionality, anonymous intentionality and unconscious intentionality – are brought together in the last papers under the title 'operative intentionality'" (Mohanty, 2005, p. 15).

'operative intentionality' shifts towards that of the other: "For a moment, we might not even distinguish his movements from our own any more, and the ambiguity of incorporation gets lost" (Fuchs & De Jaegher, 2009, p. 474).

Finally, this idea can be extended to 'mutual incorporation' by which, as the term suggests, there is "reciprocal interaction of two agents in which each lived body reaches out to embody the other" (Fuchs & De Jaegher, 2009, p. 474). In this way, the coordination between the two embodied subjects (movements, utterances, gestures, gazes, etc.) draws together each individual operative intentionality in such a way that meaning is co-created through this joint interaction that would otherwise cease to exist for each individual.[13] As such, the intersubjective relation is not seen as a simulation nor as an individualised or mentalised process, but as the result of the coordination between two lived bodies, where both are mutually perceiving and perceived as well as acting and being acted upon (Fuchs & De Jaegher, 2009, p. 477).

IV. Husserl and Enactivism

Thus far, we have established that the conclusions derived from mirror neuron research may not be as compatible with Husserlian descriptions as initially thought. We have also considered Fuchs and De Jaegher's position as to why an enactivist account offers a more thorough, and specifically, a more phenomenological understanding of intersubjectivity than mirror neurons. Now, in this final section, we will assess whether the enactive account is indeed more consistent with phenomenology, at least according to a Husserlian phenomenological tradition. That is, we will see if enactive intersubjectivity is able to overcome Zahavi's four points of criticism where mirror neuron accounts fail to adhere to a Husserlian understanding of empathy.

To Zahavi's first criticism regarding the explanatory scope of mirror neuron research, this concern is exactly in line with the enactivist approach. It seems reasonable to assume that neither Husserlian scholars nor the aforementioned enactivists would target their primary criticisms at the empirical results themselves, but both would surely question the interpretation of such results in explaining intersubjectivity. To quote Fuchs and De Jaegher (2009), "...such explanations single out *one section only* of the whole circle of organism-environment interaction. They fail to address social interaction as a structured and structuring process which in turn influences brain functions" (p. 469). While the authors maintain that the link between perception and action demonstrated in mirror neuron research may be important for social understanding, they acknowledge that a mirror neuron system could only function within the context of embodied and meaningful experiences and interactions. As such, this embeddedness must also be given explanatory value. Likewise, Husserl appreciated the complexity of interpersonal understanding and acknowledged that various empathetic levels are at play when shaping our perception of the other, and accordingly, ourselves and our world. Of course, Husserl never had the opportunity to write on mirror neurons, but it seems one could safely assume mirror neuron studies would be bracketed in his epoché along with the rest of the scientific data of his time. That is, for Husserl as well as for the enactivists, the phenomenological description is not a mere consequence of scientifically-discovered biological mechanisms but is the appreciation of the embodied being as situated within a framework in which meaning is constituted not only at the level of another's actions, but also at the level of one's relations to the world as well.[14]

Zahavi's second criticism results from the resemblance of mirror neuron system descriptions to

13 For example, shared laughter.

14 There is, however, an important distinction between the accounts whereby according to the participatory sense-making of enactive intersubjectivity, meanings are "emergent products of interaction, and in many situations, they can be viewed as distributed phenomena rather than as individual, private mental acts or properties" (Fuchs & De Jaegher, 2009, p. 480). In contrast, Husserl maintains that one's experience of the other works to constitute one's experience of herself and her world in such a way that Husserl is sometimes accused of being too idealistic and solipsistic. While this is an important concern (and will be acknowledged in the concluding section of this article), the comparison discussed in this first point holds insofar as both Husserlian and enactive descriptions of intersubjectivity require an appreciation of the worldly embeddedness of the self-other relation and a broader explanatory scope than mirror neuron accounts alone can provide.

simulation theory or even a form of projectionism. Fuchs and De Jaegher (2009) address this idea in a footnote, stating, "Although there may be an element of imitation, fascination does not imply simulation" (p. 474). They refer to an example is which an observer watches an acrobat[15] with *fascination*: "Our lived body reaches toward and 'conjoins' with the acrobat's swinging movements – we may even be prompted to co-movements" (Fuchs & De Jaegher, 2009, p. 474). Still, this is not to say that observing the acrobat leads to the ascription of a mental state to the acrobat from the observer. For the enactivist, it is a process by which "the centre of the 'operative intentionality' of our body shifts towards that of the other" (Fuchs & De Jaegher, 2009, p. 474), and the resulting empathy is not so much an internal simulation or projection, but is the result of an entwinement between self and other.

Returning to Husserl, we see a similar idea, where empathy is not a reproduction of the other nor is one's understanding of the other's experience contingent upon an analogous experience in the observer:

> "[Every I] regards the [other] lived bodies as 'bearers' of I-subjects. But it 'sees' the other I's not in the sense that it sees itself or experientially finds itself. Rather it posits them in the manner of 'empathy;' hence other lived experiences and other character dispositions are 'found' too; but they are not given or had in the sense of one's own" (Husserl, 2006, p. 5).

Although Husserlian phenomenology emphasises that my own embodied self-experience is necessary for the apperception of the other, it is through the other that the apprehension of my own body as an object is acquired (Zahavi, 2011, p. 245), as well as the lived-world being for the other. Thus, "as Husserl proceeds to point out, I am not what I am for myself, independently of the other, nor is the other independent of me. Everybody is for himself and at the same time for the other in an inseparable being-for-one-another" (Zahavi, 2011, p. 245). This seems consistent with the enactivist description, where "interactional social understanding is not an inner modelling in a detached observer, but on the contrary, the other's body reaches out to my own, and my own reaches out to the other" (Fuchs & De Jaegher, 2009, p. 475). In essence, both Husserl and the enactivists reach the conclusion that the intersubjective relation is not one of "ineffective mirroring (*kraftlose Spiegelung*)," but rather, "the being of self and other are constitutively intertwined (Hua 15/191)" (Zahavi, 2011, p. 245). This intertwinement, consistent with Merleau-Ponty[16], is also, therefore, rooted in Husserlian empathy.

To Zahavi's third criticism, of which there are two parts, both are easily in line with enactive intersubjectivity. The first criticism is that 'mirroring' is too static of a concept. Fuchs and De Jaegher address this through their incorporation of dynamical systems theory, which is essentially a mathematical tool used to describe change in real time. Thus, the assumption that the intersubjective relationship continually unfolds in real time is implicit in their use of dynamical systems. In this case, unfolding in real time means that the changes that occur are not simply unidirectional cause-and-effect mechanisms but involve reciprocal and constantly evolving interactions between two beings. According to Fuchs and De Jaegher (2009), this interaction includes behaviour as well as the retention of past experiences and the protention of future experiences (p. 476).[17] Ultimately, enactivism agrees with Husserl's demand for a dynamic description, and Fuchs and De Jaegher's account portrays the intersubjective relationship in anything but static terms.

In the second part of the criticism, Zahavi presents the issue of maintaining 'otherness.' He claims that Husserl had greater respect for the preservation of 'the other' than is seen with many mirror neuron accounts. However, this difficulty is also overcome in the enactivists' model. Indeed, if we look at enactive intersubjectivity, we see that 'the other' is a necessary foundation for their claims of mutual incorporation: "Mutual incorporation implies a component of autonomy and otherness that is absent in unidirectional incorporation. The experience of even slight mismatches or unforeseen reactions suffices to establish a

15 In *Grundlegung der Ästhetik*, Theodore Lipps (1903) uses the acrobat as an example of empathy through internal imitation, though unlike Fuchs and De Jaegher (2009), Lipps is often associated with simulation theory (p. 474). See Zahavi (2010) for further discussion on Lipps and phenomenology.

16 See Merleau-Ponty (1968).

17 These terms can be traced to Husserlian texts, with retention and protention comprising significant aspects of Husserl's writings on temporality. See Husserl (1991).

difference between self and other" (Fuchs & De Jaegher, 2009, p. 475). The authors describe how mutual incorporation partially decentres one's embodied operational intentionality. Notice, they claim only a partial decentralisation, for they believe there to be a continuous oscillation between self and other, between activity and receptivity. "In order to understand the other *as other*, empathy has to be balanced by alterity" (Fuchs & De Jaegher, 2009, p. 476). We can now see that the maintenance of 'otherness', while perhaps overlooked by many mirror neuron descriptions, plays an important role for the enactivist. In fact, mutual incorporation requires the idea of 'the other' as a point to oscillate to and from. This shows some similarity to Zahavi's interpretation of Husserl that the original givenness of my own lived body (*Urleib*) is not projected onto others, but rather, servers as a 'contrast foil' by which others can be experienced as others: "To put it differently, the other might be a self in his/her own right, but the other can only appear as another for me in relation to, and in contrast to, my own self-experience. But in this case, my self-experience doesn't constitute the model; rather it is that against which the other's difference can reveal itself" (Zahavi, 2011, p. 240).

Finally, regarding Zahavi's last criticism, it should now be clear that the enactivists' description of intersubjectivity has been at the level of the subject or person and has described the interaction of lived bodies not only as embodied beings, but as subjective beings situated in and engaging with an environment. Again, this does not mean that to adhere to an enactivist account, one must reject the science behind mirror neurons. It simply demands the limitation of the explanatory value of mirror neuron research. After all, brain functioning can certainly influence the dynamics of the enactivist model, but the central point is that a phenomenologically accurate account of intersubjectivity requires descriptions of dynamical interactions not only at the level of a brain in a body, but also as a person in the world.

Conclusion

One intent of this article is to assess whether an enactivist account of intersubjectivity is phenomenologically superior to mirror-neuron-based accounts insofar as it is able to overcome Zahavi's criticisms of the latter. I believe we can safely conclude that, yes, the enactivist account is able to overcome these issues, and thus, is more in accordance with phenomenology, at least when it comes to a Husserlian understanding of empathy. This leaves us with two conclusions:

First, Fuchs and De Jaegher's descriptions of intersubjectivity rely heavily on the work of Merleau-Ponty, but Husserl is neither mentioned nor listed as a reference in the authors' twenty-one-page publication.[18] What this article hopefully indicates is that the works of Husserl can and should be cited, for a great deal of Merleau-Ponty's descriptions of intersubjectivity can be traced back to Husserlian origins. Of course, there are reasons to exclude Husserl (e.g., his idealistic tendencies and subject-object distinctions), and his own phenomenology should not be equated with that of Merleau-Ponty. Nevertheless, intersubjectivity is a topic essential to Husserlian thought, and his collected works provide a strong foundation for defining phenomenological perspectives of empathy. This article has shown that much of the enactivists' account is compatible with Husserl's descriptions of intersubjectivity – indeed, far more so than accounts provided by mirror neuron research – so to exclude Husserl from the discussion is to do a disservice to their account. Even if fundamental differences exist between the two descriptions, the overlap between Husserl and enactivism seems sufficient and significant enough to mention and discuss.

Second, and more relevant to the scientific community, the enthusiasm observed for mirror neuron research should not result from the ability for such research to explain phenomenological descriptions of intersubjectivity. Indeed, not only does such research fail to explain these descriptions (such as those provided by Husserl in his writings on empathy), but this research struggles to be in accordance (or not in accordance) with a phenomenological approach at all, for it describes intersubjectivity at an entirely

18 Edith Stein, who expanded upon much of Husserl's work on empathy, was not mentioned either. These observations are particularly relevant to Fuchs and De Jaegher (2009) who claim a phenomenological perspective, which is why we have focused on this article. Other publications, such as De Jaegher and Di Paolo (2007), do not commit themselves so strongly to phenomenology, though similar parallels might be drawn between their enactive descriptions and Husserlian phenomenology.

different level of understanding. As such, if it is to be given explanatory value, mirror neuron research should resign any claims to provide a foundational understanding of social cognition, and instead, work to be contextualised within a broader framework – one that is able to accommodate phenomenological descriptions. From what we have seen, the enactivist account provided by Fuchs and De Jaegher should certainly be a candidate for this contextualisation, as it permits a vocabulary and understanding largely consistent with the phenomenological descriptions of Merleau-Ponty and Husserl. At the same time, their enactive account could accommodate the discovery of mirror neurons by showing the role of neuronal conditions in influencing the state space of the dynamical system, a system that facilitates descriptions at the level of the lived and situated body.[19] Moving forward, one should continue to scrutinise Fuchs and De Jaegher's account in phenomenological terms (i.e., assess its relationship to other phenomenological descriptions[20]) as well as explore alternative accounts to intersubjectivity that appreciate these various phenomenologies.[21] To be sure, determining the proper context with which to frame the continuously evolving mirror neuron research will prove to be a challenging endeavour, but the hope remains that this task will be pursued with the continued intellectual vigour as was seen with the advent of the discovery of mirror neurons.

References

Depraz, N. (2001). The Husserlian theory of intersubjectivity as alterology. Emergent theories and wisdom traditions in the light of genetic phenomenology. *Journal of Consciousness Studies*, 8(5-7), 169-178.

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, 6, 485–507.

Di Paolo, E. A., Rohde, M., & De Jaegher, H. (2010). Horizons for the enactive mind: Values, social interaction, and play. *Enaction: Towards a new paradigm for cognitive science*, 33-87.

Dinstein, I., Thomas, C., Behrmann, M., & Heeger, D. J. (2008). A mirror up to nature. *Current Biology*, *18*(1), R13-R18.

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding.*Science*, *308*(5722), 662 667.

Fuchs, T., & De Jaegher, H. (2009). Enactive intersubjectivity: Participatory sense-making and mutual incorporation. *Phenomenology and the cognitive sciences*, *8*(4), 465-486.

Gallagher, S. (2001). The practice of mind: Theory, simulation or interaction? In E. Thompson (Ed.), *Between ourselves: Second-person issues in the study of consciousness* (pp. 83-107). Exeter, UK: Imprint Academic.

Gallagher, S. (2012). Neurons, neonates and narrative: From empathetic resonance to empathetic understanding. In A. Foolin, U. M. Lüdtke, T. P. Racine, & J. Zlatev (Eds.), *Moving ourselves, moving others: Motion and emotion in intersubjectivity, consciousness and language,* (pp. 167-196). Amsterdam: John Benjamins Publishing Co.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*(2), 593-610.

Gallese, V., Fogassi, L., Fadiga, L., & Rizzolatti, G. (2002). Action representation and the inferior parietal lobule. *Common mechanisms in perception and action: attention and performance*, *19*, 247-266.

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in cognitive sciences*, *8*(9), 396-403.

Gallese, V. (2009). Mirror neurons, embodied simulation, and the neural basis of social identification. *Psychoanalytic dialogues*, *19*(5), 519-536.

Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of cognitive neuroscience*, *21*(7), 1229-1243.

Husserl, E. (1991). *On the phenomenology of the consciousness of internal time (1893–1917)* (Vol. 4) (J. B. Brough, trans.). Dordrecht, London: Kluwer Academic Publishers.

Husserl, E. (1997). *Thing and space: Lectures of 1907*. Collected Works VII (R. Rojcewicz,trans.). Dordrecht: Kluwer.

19 Such an account would be in accordance with Thompson & Varela (2001), however, one should be careful regarding claims to naturalise phenomenology – Fuchs and De Jaegher (2009) do not subscribe to this endeavor (p. 470).

20 Zahavi (2001) might be useful in this regard, for he discusses intersubjectivity in the works of Scheler, Heidegger, and Sartre as well as Husserl and Merleau-Ponty.

21 For example, Depraz (2001) discusses how empathy as described by Husserl and empirical second-person accounts can influence each other. Likewise, Gallagher (2012) has provided an overview of empathetic views "consistent with phenomenological and scientific evidence," and has advocated the inclusion of narrative competency in accounting for intersubjectivity (p. 190-191). On the other hand, Thompson (2001) has continued Francisco Varela's push for a contemplative 'science of interbeing' as a way of integrating cognitive science and phenomenology as well as contemplative Eastern traditions.

Husserl, E. (2006). *The basic problems of phenomenology: From the lectures, winter semester, 1910-1911* (Vol. 12) (I. Farin & J. G. Heart, trans.). R. Bernet (Ed.). Dordrecht, Netherlands: Springer Science & Business Media.

Jensen, R. T., & Moran, D. (2012). Introduction: intersubjectivity and empathy. *Phenomenology and the cognitive sciences*, *11*(2), 125-133.

Kilner, J. M. & Lemon, R. N. (2013). What we know currently about mirror neurons. *Current biology*, 23, R1057-R1062.

Lipps, T. (1903). *Grundlegung der Aesthetik*. Bamburg: Engelmann.

Marr, D. (1982). *Vision*: *A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman and Company.

Merleau-Ponty, M. (1962). Phenomenology of perception. Translated by Colin Smith, London: Routledge & Kegan Paul. (Original work published 1945).

Merleau-Ponty, M. (1968). *The visible and the invisible: Followed by working notes*, C. Lefort (Ed.), A. Lingis (trans.). Evanston, IL: Northwestern University Press.

Mohanty, J. (2005). Husserl's concept of intentionality. In R. Bernet, D. Welton, & G. Zavota (Eds.) *Edmund Husserl: Critical assessments of leading philosophers: The nexus of phenomena: intentionality, perception, and temporality.* (Vol. 3). London: Routledge.

Ramachandran, V. S. (2000). Mirror neurons and imitation learning as the driving force behind "the great leap forward" in human evolution. *Edge.* Retrieved from*: http://www. edge.org/3rd_culture/ramachandran/ramachandran_p1.html*.

Ratcliffe, M. (2009). Phenomenology, neuroscience, and intersubjectivity. In H. L. Dreyfus & M. A. Wrathall (Eds.), *A companion to phenomenology and existentialism* (pp. 329-345). Sussex: Wiley-Blackwell Publishing Ltd.

Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., & Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey. *Experimental brain research*, *71*(3), 491-507.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive brain research*, *3*(2), 131-141.

Thompson, E. (2001). Empathy and consciousness. In E. Thompson (Ed.) *Between ourselves: Second-person issues in the study of consciousness* (pp. 1-32). UK: Imprint Academic.

Thompson, E., & Varela, F. J. (2001). Radical embodiment: neural dynamics and consciousness. Trends in Cognitive Sciences, 5(10), 418-425.

Tkach, D., Reimer, J., & Hatsopoulos, N. G. (2007). Congruent activity during action and action observation in motor cortex. *The Journal of Neuroscience*,*27*(48), 13241-13250.

Umilta, M. A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., & Rizzolatti, G. (2001). I know what you are doing: a neurophysiological study.*Neuron*, 31(1), 155-165.

Zahavi, D. (2001). Beyond Empathy: Phenomenological approaches to intersubjectivity. *Journal of Consciousness Studies* 8 (5-7), 151-167.

Zahavi, D. (2010). Empathy, embodiment and interpersonal understanding: From Lipps to Schutz. *Inquiry*, *53*(3), 285-306.

Zahavi, D. (2011). Empathy and mirroring: Husserl and Gallese. In *Life, Subjectivity & Art* (pp. 217-254). The Netherlands: Springer.

Research Article **Open Access**

Ane Engelstad

False Consciousness and the Socially Extended Mind

DOI 10.1515/pipjp-2016-0004

Abstract: In this paper I present a problem for the Marxist idea of false consciousness, namely how it is vulnerable to accusations of dogmatism. I will argue that the concept must be further developed if it is to provide a plausible tool for systematic social analysis. In the second half of the paper I will show how this could be done if the account of false consciousness incorporates Shaun Gallagher's theory of the socially extended mind. This is a theory that explores how the mind expands towards external objects and systems. I will conclude that it helps to reinstate false consciousness as a reliable tool for the analysis of cognitive dynamics within power structures.

Keywords: False consciousness; social mind extension; enactive cognition; externalism; social analysis

Introduction

In this paper I will present a problem for the Marxist notion of "false consciousness". This concept concerns the effects of ideology on individual consciousness, and particularly the ways in which human beings within a society incorporate and reproduce power structures that are not necessarily to their personal benefit. First, I show how false consciousness poses problems for its own systematic framework, and for presenting an account of cognition and knowledge as it actually manifests itself within a set of power structures. This issue must be resolved if false consciousness is to be retained as a useful tool for the analysis of the effect of material power structures on the individual mind. I then go on to briefly elaborate how a possible solution should be approached if false consciousness is to retain its value as a concept. In the final part of this paper, I develop my own solution based on this approach by elaborating on false consciousness as a concept in light of Shaun Gallagher's theory of the socially extended mind, which explores how the mind expands towards external objects and systems. I conclude that this elaboration of false consciousness is capable of presenting an account that fulfils the criteria for a successful solution to the problem stated at the beginning of this paper. [1]

The Problem With False Consciousness

The Origin of False Consciousness

The idea of false consciousness emerged as a reply to Wilhelm Reich's puzzlement with the fact that people do not actively resist exploitation (1933, 53). Although this worry dates back to Plato (book V, 459c-d), it became of particular concern to post-Marxist theory, as capitalism seemed to persevere even in times of economic crisis, and the experiments of social change of the early 20th century turned out to be repressive

1 Many many thanks to Dr. Sarah Fine for invaluable input, help and support during and after the writing process.

***Corresponding author: Ane Engelstad,** University College London (UCL)

rather than emancipatory (Benhabib 1994, 66). Marx never used the term "false consciousness" as such, but the idea is rooted in his materialist ontology (Eagleton 1991, 70). Marx argues that the material foundations of society, i.e. the available material goods, how they are distributed, and who they are controlled by, provide the foundation of social practice, and that social practice is crucial to idea formation. He rejects the thought that ideas exist over and above the material realm of human agency (Balibar 1995, 71). This inevitably means that ideas developed in a society serve to uphold the same power structure as the system of material distribution. Following on from this line of reasoning, the idea of false consciousness is concerned with how this feedback loop between idea production, belief systems and power structures affects individual minds (Marcuse 1964, 145). A belief is subject to false consciousness if, irrespective of the knowledge of the believer, it serves to uphold a specific distributional power structure which operates outside our direct control. As a result, the theory of false consciousness provides an answer to Reich's question; people are not always in the epistemic position to directly perceive what is in their own interest, thus the exploited do not realise that it is the current system that represses them.

It is Important to notice that for a belief to be an instance of false consciousness it does not *necessarily* need to be propositionally false; it is a sufficient condition that the beliefs are held because this would be instrumental for the reinforcement of a specific power structure, and that this is detached from the agent's personal reasons for holding this belief. The truth condition of the proposition is irrelevant. Coincidentally, the content of the belief is more *likely* to be false as it is not purposefully formed from evidence about the world *as it is*. The dynamics by which we come to form beliefs pervasively follow the logic of structures that aim to reinforce themselves.

As an illustration, this does not necessarily mean that all forms of scientific discovery must be undermined *as such*, but Marx questions *why* some forms of research and not others are prioritised, *how* the findings of the research interrelates with a prevailing belief system, and *what* kind of function this belief system has within the power structures of society. For instance, the belief that milk is healthy might contain a true proposition, but one might question why we hold this belief specifically, instead of beliefs about other things with similar health benefits. If it turns out that we ultimately believe that milk is healthy because the dairy industry benefits from us holding this belief, this must be an instance of false consciousness. Moreover, whether we hold the belief on these grounds is largely obscure to us, due to our socially situated perspective as epistemic agents. This perspective is skewed by the fact that the very mechanism of idea production, the way we access evidence and form beliefs about the world, follows a specific power dynamic.

At this point it is important to notice that the person who is subject to false consciousness is not simply making epistemic mistakes, or is intrinsically less skilled at making sense of epistemic evidence. The distortion is externally imposed.

Moreover, false consciousness is not a case of intentional knowledge control instigated by the ruling classes, as is the case with propaganda. This is because false consciousness is a by-product of the fact that idea production is inseparable from physical production in a given society. If material production serves a hierarchy of unjust material distribution, so does idea production.

Issues with False Consciousness

However, false consciousness is not an unproblematic concept. It has been criticised for not recognising the actual complexity and multiplicity of culture and knowledge *within* societies where capitalism is the pervasive mode of distribution. It seems incommensurable with the existence of dissidence and the fragmentation of the actual political landscape (Eagleton 1991, xxi). It cannot explain the existence of political activism that aims to counteract the effects of a given power structure.

In the remainder of this essay I will therefore aim to refine the concept of false consciousness such that it manages to account for the actual diversity that takes place within a society, while explaining why people so often are complicit in their own exploitation and oppression. The aim is to do this within a Marxist framework, which provides a systematic account capable of explaining not only singular cases of deception, but also the broader social tendency. It is important to salvage the theory of false consciousness as a tool

for social analysis, as the dynamics it highlights have been repeatedly demonstrated in sociological studies. For instance, Pierre Bourdieu found that

> there are many things people accept without knowing. [...] When you ask a sample of individuals what are the main social factors of achievement at school, the further you go down the social scale the more they believe in natural talents or gifts – the more they believe that those who are successful are naturally endowed with intellectual capacities. And the more they accept their own exclusion, the more they believe they are stupid (Bourdieu and Eagleton 1994, 269).

This account of class identity as mirrored in belief in innatism seems to fit the exact idea of what false consciousness *is*: a mechanism that unintentionally reinforces an existing social hierarchy at the level of ideas and self-understanding. This is not a case of false consciousness *merely* because the belief at hand is factually false, but because it shows how belief changes with social perspective in a way that serves to uphold the existing power structure, irrespective of the truth condition of the belief at hand. The concept of false consciousness should therefore not be discarded as a whole; rather what is needed in order to answer the critics is a more fine-grained conception of its mechanisms. In this light, the Marxist account of idea production must be further developed.

Praxis as grounding Justified Beliefs

Given that false consciousness is a belief held for the purpose of reinforcing a given power structure, it is necessary to discuss the account of *right* grounds for knowledge that follows from this view. How are beliefs produced in a way that reflects how the world actually unfolds itself for us?

Marx's materialism is at the core of his account of how the external world should be understood to impact knowledge acquisition, and it explains how idea production parallels material production. Refuting Hegel's idealism, Marx argues that everything is essentially material, meaning that the ontologically foundational is not the idea, but the subject having the idea (Marx 1843, pt.2a). This subject is a foundationally material being that acquires knowledge about the external world through its existence at the level of objects. This does not mean that subjectivity itself is essentially material, but that it supervenes on the physical world (Marx 1845, 144). An account of adequate knowledge is therefore found in the Marxist concept of *praxis*, namely the unification of theory and practice. Theory is to be understood as any belief informing the decision about which action to perform. Practice is the action itself. Only beliefs informed by a practice are appropriately justified (Marx 1844: ch. 6, XXIII). The concept of praxis is thus founded on the idea that human beings in equal part mould and are moulded by the external world. This is what allows propositions about the external world to have meaning in a way that reflects the way the world *is* for us. The mechanisms behind false consciousness are based on an inadequate connection between theory and practice. In false consciousness, beliefs are held in order to prompt a desired action, namely actions that comply with and reinforce a material power structure. False consciousness is therefore the reverse of appropriate knowledge acquisition, which is the formation of a belief based on the grounds of autonomous activity.

The implication of such a conception of knowledge in a capitalist system, where labour division is the key to economic efficiency, is that the working classes are effectively expelled from the realm of general knowledge by only being allowed to take part in minute operations of the production of the whole product. Their epistemic perspective is physically restricted. This is what Marx calls alienation from the means of production (Blunden 1999, Al). Moreover, false consciousness will be perpetuated in any society that distinguishes between manual and mental labour, as theory and practice are efficiently kept apart (Marx and Engels 1846, 72). Religious, legal and bureaucratic institutions are at the centre of this mental production as well as being founded on the pre-existing set of material power structures (Marx 1846, 4).

The issue at hand is therefore how a theory of false consciousness is compatible with how we

i. account for the fragmentation of epistemic stances within social groups subject to the same power dynamics, and thus
ii. also account for activism towards social change that aims to counteract the material dynamics of capitalist society.

False Consciousness From the Perspective of Philosophy of Mind

Requirements for a Solution

A more nuanced account of the role that external power structures play in individual idea production, and how this individual idea production helps reinforce these external power structures, is required in order to resolve the above issue. Moreover, this solution must explain how false consciousness is not merely imposed on the individual, but automatically endorsed upon encounters with the external world. This must be investigated if the pervasiveness and efficacy of the concept is to be accounted for. Finally, although false consciousness is to come from within the subject itself, a solution must show how this is not inherent to human epistemic abilities and access to self-knowledge, such that emancipation can be accounted for as genuinely possible.

I will examine whether a framework from philosophy of mind could provide an additional structure that accommodates all these requirements, as it considers the same explanatory gap between mind and world as the theory of false consciousness. While false consciousness focuses on the ways in which the world imposes itself on to the mind, philosophy of mind is preoccupied with how the mind relates to the world. Although these are two different domains of philosophical discourse, it is not problematic to say that they concern the same relation as seen from different perspectives, and can therefore inform each other. This provides me with the tools to examine the idea of false consciousness both on the grand (political) scale where ideology is produced, and the implication of ideology on the smaller scale of individual minds. This will allow me to pinpoint the level of cognition at which false consciousness operates, to what extent it is a necessary part of the human psyche, and what impact it could have on epistemic capacities and self-knowledge.

The kind of theory within philosophy of mind that could sufficiently refine the concept of false consciousness would be a theory that accounts for the ways the external world plays a role in determining mental states, and how these mental states again could serve the purpose of reinforcing external power structures. Traditional externalism about mental content, functional mind extension, and enactive mind extension are all views that could fit this description. I will show that only enactive mind extension is capable of providing a sufficiently fine-grained account of false consciousness such that multiplicity and political counterculture could be accounted for.

Content Externalism

Content externalism aims to show how external, environmental factors directly determine our mental content. Hilary Putnam's "Twin Earth" example demonstrates this. He imagines earth and the planet twin earth, which are identical in most respects, including language. The major differences in linguistic meaning depend on the material peculiarities of the planets themselves. For instance, what is referred to as water is identical in appearance on both planets, but consists of H_2O molecules on earth, and an XYZ molecular structure on twin earth. The molecular structure of water is not generally known in either place. Thus, when people speak of water it is irrelevant whether they know about the relevant molecular structure. Irrespective of this, Putnam argues that when two psychologically identical people, one on each planet, talk about water, they talk about essentially different things. The meaning of water as it functions in our mental content is entirely contingent on what "water" essentially is. This has nothing to do with the epistemic status of the speaker of the word (Putnam 1975, 584).

Moreover, meaning is *socially* contingent, and depends on the division of linguistic labour within a society where some people are scientists and decide the meaning of things like water, gold, beech or arthritis, whereas other people merely need to refer back to these definitions in order for their words to have meaning. This is why it is possible for a person to talk meaningfully about water without knowing what water *essentially* means. It is the social environment, not internal cognitive structures or experiences, which determines the meaning of any given proposition (Burge 1979, 600).

Externalism is therefore capable of giving some account of false consciousness based on this theory of content production. The linguistic division of labour could be such that meaning is generally dictated to suit a specific power structure. This means that all beliefs would be formed on false grounds, since all meaning serves to reinforce this structure. It therefore explains how our beliefs become instances of false consciousness without being consciously created this way, as the agent's intentions do not affect the way words acquire meaning. Bourdieu's example of belief in innatism can be analysed on these terms; being far down on the socio-economic scale is *defined* as equivalent to being less talented on this account. The meaning of the belief has been created in order to make people compliant to the existing social order.

However, there are several problems with this version of false consciousness. First, content externalism would enforce that the belief in the correspondence between social position and innate abilities would be accepted overall in the linguistic community, since this is the meaning to which "lower social standing" is generally attributed. This is not the case according to Bourdieu's study; the belief in this correspondence was only held at the lower socio-economic levels of a society. Secondly, externalism does not account for false consciousness in terms of general power structures, but in a deliberate dissemination of false consciousness by agents who control the determination of meaning within a society. Thus false consciousness bears closer resemblance to a propaganda mechanism than a default by-product of a given material power structure. Thirdly, content externalism does not provide an inadequate extension of the theory of false consciousness, as it is too crude to explain how belief could ever be accessible on the *right* grounds. I therefore reject content externalism as a plausible extension of the theory of false consciousness. I will therefore look towards theories that claim that not only meaning, but the mind itself, is determined outside the confines of the brain.

The Extended Mind

Such an account is mainly attributed Andy Clark and David Chalmers, who famously claim that in certain interactions with the external world "the human organism is linked with an external entity in a two-way interaction, creating a *coupled system* that can be seen as a cognitive system in its own right" (Clark & Chalmers 1998, 8). The essence of this argument is summed up by what Clark and Chalmers call the parity principle. If a process in the world works in a way that we would count as a cognitive process if it happened in the head, this external process should also count as a cognitive process (Clark & Chalmers 2002, 644). This principle is illustrated through the example of Inga who has normal mental capacities and Otto who has severe Alzheimer's disease but writes down all the things he knows and learns in a notebook. For Otto, the notebook serves as an external memory, responsible for founding beliefs which he may or may not choose to act upon, in much the same way as the latent memory stored in a mind like Inga's. When Otto and Inga both want to go to the Museum of Modern Art, their cognitive processes seem identical, with the single difference that Otto consults his notebook in order to know how to get from A to B, whereas Inga retrieves the same information from memory (Clark and Chalmers 2002, 647). They should therefore be treated as the same cognitive process.

With the extended mind view, cognition can happen in a way such that the external becomes a part of this cognitive process. This means that it is possible for the individual agent to produce beliefs by incorporating external factors in the cognitive process, such that these beliefs still have the same status as beliefs that the individual would form without any external influence whatsoever.

It is here possible to account for false consciousness as a result of this direct coupling with the external world. If the parity principle holds, this means that belief can be produced through the coupling of the mind not only with innocuous objects which do not alter the content of the belief itself, which is the case with the notebook, beliefs could also couple with external ideological structures. The coupling of internal and external structures means that the external structures could guide idea production such that the ideas produced could be made to reinforce a given power structure. Bourdieu's example can thus be explained through the coupling of the individual's cognitive structures with external power structures, for instance, the coupling with various social norms and institutions that function as cornerstones of a given social structure. The belief of the individual would depend on the social structures encountered.

However, this account of false consciousness is yet too vague to provide a sufficient elaboration of false consciousness. A further specification can be developed based on the responses given to a worry about the extended mind view raised by Rob Rupert. The worry, known as the "cognitive bloat", accuses extended mind theories of inflating the mental to the border of the insignificant (2004). Cognition could happen so far from the individual mind that it seems absurd to talk about these cognitive processes actually adhering to any specific mind. This is a worry that must be addressed by advocates of the extended mind view. If all cognition at all times could be a part of an external structure and happen at a level remote from the individual agency that it couples with, it becomes difficult to discern how emancipation from false consciousness could come from within any individual agent. It leads to a view of individuals as just a part of a larger cognitive network where individual agency is reduced, to the extent that it threatens any possibility of emancipation from this social structure.

A solution to the cognitive bloat must therefore be to qualify how extension to the external world could only happen if there is a way the individual holds a sense of ownership over this extension. There are two alternative accounts that fill this condition; Andy Clark's functional vehicle view and Shaun Gallagher's enactive view. I will examine both, and conclude that only Gallagher's enactive view is capable of presenting a potential solution to the issue at hand.

The Functional Vehicle View

Andy Clark's functional vehicle view presents four restrictions to the original extended mind thesis. The mind can only extend to external objects if it is reliably invoked, automatically endorsed, accessible when required, and has been consciously endorsed in the past (Clark 2008, 79). Thus the mind can only extend to vehicles of belief, since this is the sole aspect of the mind that fits all the criteria. A vehicle is a material object that plays a role in enabling a system to have a mental state. It could be any kind of internal or external function (Clark 2008, 76).

The mind therefore cannot extend its own structures to external ideological structures because a structure cannot be consciously endorsed. Otto's notebook, on the other hand, is such a vehicle, since it is reliably invoked and accessible when required (he carries it with him at all times), while being automatically endorsed because he knows that at some point, he personally acquired these beliefs. With this view, *belief* is not produced through coupling with the external world, rather the external plays the role solely as a vehicle for already existent cognitive content.

This leads to a problem for an account of false consciousness in terms of mind extension. Since the mind only extends to vehicles, it simply does not account for the production of belief at all, leaving it incapable to account for the mechanisms behind false consciousness. The only way in which false consciousness could be explained on this view would be through the distribution of vehicles within a given material power structure. This could for instance lead to a lack of access to research equipment, or a lack of notebooks for people with memory loss for that matter, meaning that access to certain types of belief would be restricted. But this would only diminish the set of possible beliefs in a society; it would not account for whether and when the beliefs actually held are instances of false consciousness.

The Enactively Extended Mind

Shaun Gallagher's enactive view of the extended mind incorporates the foundation of Clark and Chalmers' original extended mind view, namely the parity principle. Gallagher claims that if this principle is the foundation of the extended mind view, the restriction of the theory to vehicles only must be *ad hoc* (Gallagher 2013, 2). Vehicles are only a subset of all cognitive structures that would work on the parity principle, meaning that a restriction of the extended mind to the vehicle view would exclude possible cognitive coupling with external structures that *must* be invoked for central concepts in a society to acquire meaning. This is altogether counterintuitive, granted the acceptance of the parity principle.

To illustrate his point, Gallagher introduces the example of the legal system, which is what he calls a "mental institution". It can be argued that such institutions could play a similar role in cognition to, for instance, a calculator, which is generally accepted on Clark and Chalmers' view. When engaging with the legal system it provides external structures that take the role of cognitive structures, shaping how to think about a subject. When the mind extends to the legal system, it extends to a whole set of rigid social practices and norms, encompassing books, buildings, behaviour, people, and a network of specific cognitive procedures. When engaging with a legal issue, there is simply no way that it can be understood in cognition without the use of the structures provided by a specific legal system in a specific society.

Aiming to retain the force of the parity principle, Gallagher therefore introduces a different kind of restriction to his theory as a response to the cognitive bloat. For cognition to truly belong to an individual it is necessary to look at the *mechanisms* of the given account of ownership. As a result, Gallagher adapts a Lockean theory of ownership through labour as a restriction as to when and how the individual mind extends. Although Locke's account originally is a theory of physical labour, Gallagher understands idea production as the kind of labour that would be applicable (Gallagher 2013, 7). The Lockean theory of labour states that when a person X is mixing his labour with an external object in order to transform it, the object transformed by X's labour becomes a part of X's property (Locke 1690, §27). The mind therefore only extends to the external world if it is *actively* engaged with the individual's production of ideas. The cognition happening outside the individual's head only truly belongs to the individual if mental labour is applied in order to produce ideas. For example, the mind only extends to the legal system in cases where legal structures play a role in the formation of a new idea.

If someone stole my bike, my mind would extend to the legal system insofar as I would create the idea of *theft* in relation to my acknowledging that someone else took my bike without my consent. This idea of theft has meaning in virtue of instigating the appropriate set of behaviours that the concept entails: reporting to the police; instigating an official investigation; and a trial against the thief, eventually resulting in fines or imprisonment. In a society without a legal system, the fact that my bike was taken without consent does not mean anything other than just that. The concept of theft does not have any specific meaning because it does not entail an appropriate set of behaviours that it could instigate. This illustrates Gallagher's argument that certain concepts would be meaningless if they did not invoke these cognitive external structures.

Therefore, for the Lockean theory of ownership to be united with a theory of mind extension, Gallagher suggests that the parity principle should not hold functionally, but enactively (Gallagher 2013, 2). Enactive knowledge is knowledge acquired through action, and is therefore shaped by the external and internal structures that guide this action (Wilson & Foglia 2011). Consequently, as with the example of bike theft, when the mind extends to the external world, the external world becomes a part of our cognitive structures that shape how the external world is engaged with, and therefore also the ideas produced as a result of this engagement. A belief therefore acquires meaning from the actions that shaped it, and the actions it could lead to, as resonated in the concept of *praxis*.

Enactive Mind Extension and False Consciousness

It is here possible to show how enactive mind extension could provide an elaboration of false consciousness. The organisation of the external world provides structures that the mind could couple with, and therefore guides both action and production of new ideas. Since all external structures are restrictive in some sense, they delineate both the active space of the agent, and consequently the enactive space of cognition. Moreover, actively coupling with external structures when producing ideas means that these structures control idea production directly. In virtue of this, the structures are enabled to take part in idea production for the sake of instigating actions that comply with these structures. Since the ideas produced will only have meaning based on the set of actions they result in, this account also provides an idea of how false consciousness directly reinforces a given social hierarchy.

Bourdieu's example of a belief in innatism as an instance of false consciousness can here be explained along similar lines to the broader account of the extended mind and false consciousness as given before

the introduction of cognitive bloat as a worry, while allowing for physical restriction itself (through lack of resources) to play a role in the given epistemic space of a social group.

This is a subset of classic false consciousness commitment claiming that liberal society prompts us to believe that opportunity is available to anyone who tries hard enough, and that any failure is the fault of the individual. Other such commitments that have been attributed to a liberal outlook can also easily be accounted for by the enactivist. For instance, the belief that capitalism is a natural force can be explained by the fact that we engage with social structures that treat capitalism *as if* it was a natural force that should be left entirely to its own dynamic development. The way we engage with social and material objects form the beliefs we hold about them. Often, these beliefs are held on the wrong grounds simply because we are unclear about the origin of this behaviour. Although capitalism is man-made, once it is given priority over other forms of social organization, the mere fact that it is deregulated makes us relate to it in our agencies *as if* it was a natural force. In return, the belief that capitalism is a natural force reinforces our treatment of it as if it was one. The enactivist account thus perfectly coheres with the Marxist concept of praxis as virtuous belief formation; our belief in capitalism as a natural force is held in order to prompt a desired action. It is not a belief formed on the basis of autonomous agency. It is therefore held on the wrong grounds, and must be a case of false consciousness.

Another important feature of the enactivist account is that it provides a general theory of human cognition, while also accounting for how this cognition can be manipulated in virtue of its own faculties. For this theory to be able to produce a plausible conceptual extension of false consciousness, it is important to remember that all ideological structures are structures that could prompt enactive mind extension, whereas not all structures that could be coupled with enactively are necessarily ideological. This means that enactive mind extension as an account of human cognition is not a problem for false consciousness *per se* – the problem exists at the level of unfavourable, external power structures. This account therefore addresses the crux of false consciousness, which should not be seen as an account of how ideology forces itself on people, but "how can there be an animal which represses itself" (Brown 1959, 9).

The enactive mind view is thus the most apt elaboration of the theory of false consciousness, as it provides a specific account of how and when extension towards the external world could result in a case of false consciousness. The mind couples with external structures insofar as they provide a way to act and behave, and belief is based on this. For instance, in the case of building a fence, the mind does not extend to the fence insofar as it is a fence, although it *could* extend to the process of building a fence insofar as its function; how and where it is built, comply to a specific social practice or norm (Gallagher 2013, 8). Therefore, every aspect of the external world can stand in relation to a given social power structure, and all such norms and conditions can become part of the individual mind through the practices that they dictate.

As a result, this account of false consciousness is strong enough to sufficiently account for the pervasiveness of the concept. However, does this entail that the possibility for emancipation must be excluded? It seems intuitively desirable to assert that beliefs that invoke the legal system *must* result in cases of false consciousness, in virtue of its nature as a social institution, and therefore as the centre of ideological production as defined by Marx. Moreover, it also seems desirable to distinguish this kind of false consciousness from the false consciousness that *could* result from the process of building a fence. How can the extended mind theory provide a differentiation between these cases, and does this provide an apt explanation of political dissent?

The Solution

The Range of False Consciousness

Based on the desired distinction between false consciousness as the necessary result of mind extending to institutions, and the false consciousness that possibly arises from the process of building a fence, it seems appropriate to introduce a scale from non-ideological idea production to strict ideological idea production where the ideas created gain meaning because they instigate a specific action.

The clearest distinction between the process of building a fence and the process of engaging with the legal system is how they relate to the production of meaning. Whereas Gallagher argues that words like "theft" would be meaningless if they did not refer to the various social practices that were entailed by the word, the same would not be the case for concepts such as "fence". A fence constructed without regards to specific social norms could still meaningfully be a fence. Although most fences stand in a clear relation to the society they were built in, such as garden fences, which primarily serve as a symbolic delineation of property, this does not mean that "fence" as a concept would be meaningless to a hermit who builds a fence around her shelter in order to keep wild animals from coming in. However, on this account, "fence" only has meaning *as it is created by, and organises someone's agency*. This is what makes the difference between "fence" and a random stack of rocks to the hermit, and this is also why merely decorative garden fences have a specific meaning to us, according to Gallagher. They play a role in how we understand, organise and structure our lives, be it practically or symbolically.

"Theft", on the other hand, would be a meaningless concept to this hermit, since it will not invoke any specific action or behaviour outside a given social structure. This example illustrates a crucial distinction between necessary production of false consciousness and at least a possible access to meaning outside the social hierarchy. Given an enactive account of meaning as dependent on the actions a belief invokes, this distinction can be explored further.

The external cognitive structures that the mind couples with dictate the way in which the external world gains meaning, in virtue of shaping how the external world is engaged with. This also means that autonomous agency must be agency where internal cognitive structures are not completely dominated by the external structures that the mind couples with. For instance, mind extension towards the legal system does not allow for autonomous belief formation, because the legal structures are not open to alternative cognitive processes. On the other hand, the process of building a fence would only produces false consciousness insofar as the social norm is followed and reinforced. Obviously, the political significance of this belief formation might be minimal, and it is also possible to be fully aware of the fact that one is following the social norm for the sake of following the social norm, and still happily comply with it. This is not the issue at stake. The significant point, as stated in the example about the hermit, is that it is *possible* to develop a belief about fences without any reference to the socially normative, solely through the mere action of building one, while also retaining the account of enactive belief formation in conjunction with the theory of false consciousness.

The key question is therefore whether it is possible for an agent within a society to actually divorce a practice from a social norm, such that belief could be formed autonomously. It follows that I will need to explore how and when belief production that is not *necessarily* tied to external power structures could be cases of genuine belief, and contrast this with the strong case where meaning is necessarily given by such a power structure.

False Consciousness and Context Dependency

A weaker sense of false consciousness would be instances of idea production that generate meaning that does not *in all possible contexts* reinforce the ideological structure. Whether false consciousness is generated in the process of building a fence is contingent on social norms and situations. This contingency, present in the weaker sense of false consciousness, is context dependent. Some context dependency also exists in the strong sense of false consciousness, where the production of meaning itself is contingent on the kind of society it takes place in, but the weak account also allows for meaning to be context dependent *within* a society. The strictness of some social norms can depend on such small-scale contexts. For instance, the structure that leads to idea production in a family setting might lead to the production of ideas for the sake of the idea, and not for the sake of reinforcing either the familial structure, or a larger social hierarchy. The meaning of "fence" in a children's game might not lead to agency that reinforces ideological structures.

However, although the weak case allows us to hold beliefs for the right reasons, it does not explain our epistemic access to whether our ideas are held for these reasons. It might not be clear to us whether the

children's game reflects already existing social dynamics, and thus serves to reinforce these. Therefore, although this discussion of meaning provides an answer to *when* belief can be founded on the right grounds, it does not account for whether this allows for emancipation through access to genuine belief, since knowledge about this genuineness is largely inaccessible to the agent.

In order to solve this problem, it is necessary to deduce cases of enactive idea formation where the subject knowingly holds the produced belief on the right grounds. For instance, one such belief could be "I believe it is nice to sit in the cool shade of the tree in my garden". This is a belief that has been acquired enactively through the action of sitting under a tree in a garden, and it is a belief that also prompts further actions, like repeated sitting under the tree in the garden. It is a belief that suits only my personal tastes, and has been acquired for no specific reason other than that it is a place for me to sit. The fact that I have a garden, or the fact that I have a tree in it might be embedded in a row of social norms, but my belief about what it is *like* to sit under this tree is entirely subjective. It might be in someone's interest that people generally like to sit under shady trees in gardens, but my belief about my preferences for sitting under the tree in my garden can only have come about through the fact that I sat under the tree in the garden and realised that I liked it. This cannot be a case of false consciousness because it is a belief about the external world acquired through action, not a belief about the external world that is acquired *for the sake of* prompting this action.

This is a particularly isolated case, and might not in itself lead to a realisation about the unfairness of a social structure and therefore a desire and will for emancipation. However, such beliefs can create a building block for alternative enactive structures that are developed for no particular reason other than for the sake of the beliefs themselves. For instance, I might meet another person that has the same belief about the comfort of sitting in the shade of a tree in the garden. Based on this mutual interest we might be able to create a small social context founded on this appreciation and nothing else.

Through this social context, other norms and codes might arise as a result of our beliefs about the external world and the role it plays within this context. The beliefs that arise from this context are beliefs that *could* serve as instances of false consciousness, but not necessarily so. Despite this possibility, since these beliefs arise through the enactive structures of this smaller context, they hold no specific purpose. Moreover, the subject that holds this belief is aware of the fact that this belief is held on the right grounds, because the subject is aware that the context that gave rise to this belief was developed entirely from appropriately justified belief such as the belief about the pleasantness of sitting in the cool shade of a tree.

It is here possible to imagine that the budding alternative social structures that arise from this context could eventually be the source of enactively developed ideas critical to the larger power structures of a society. In their shared experience, the small community of tree lovers might, for instance, develop social values different to the ones they would have held had they merely been part of a larger community with a different kind of social hierarchy.

False Consciousness Sufficiently Elaborated

The possibility of knowingly creating small pockets of alternative contexts, founded on appropriately justified basic beliefs *within* a larger social power structure, could lead to idea production that within this context would not be a case of false consciousness such that this would be knowable to the subject that produced these ideas. This possibility is therefore capable of providing a solution to the issues at hand, as the enactive extended mind view provides an account of false consciousness as being pervasive and strict, at the same time as it creates a window for political counterculture.

This window is nevertheless small, as any closed social context embedded within a larger society will bear traces of these overarching power structures. To intentionally create a vacuum from the beliefs acquired and held in society in general seems in itself like an unlikely feat. Nevertheless, the aim at this point is merely to show how the possibility of emancipation does not have to be altogether excluded. Moreover, the existence of alternative enactive structures, even when they are only partly founded on genuine foundational beliefs, will provide an account of the actual plurality of beliefs and actions within

a capitalist system, at the same time as it explains why this plurality has not led to a great drive towards overturning a distributive system that works against the interest of most people.

This solution fulfils all the conditions stated at the beginning of the third part of this essay; it shows how external power structures play a role in individual idea production, and how the production of these ideas again reinforce the power structure from which they were developed. It accounts for how false consciousness arises in the same way as any other belief held by the individual, instead of being a belief externally imposed through, for instance, propaganda. Moreover, it provides an account of how false consciousness is not inherent to the human epistemic abilities as such, and is only the result of external power structures. The enactive extended mind theory allows for a more diverse picture of minds, ideas and ideology while retaining a theory of false consciousness as both strict and necessary on direct encounters with the ideological structure.

Conclusion

I have shown through this discussion that Shaun Gallagher provides an apt elaboration to the Marxist concept of false consciousness. Moreover, Gallagher's theory of the socially extended mind is capable of fulfilling the conditions needed in order to sufficiently fine grain the concept of false consciousness, retaining its analytic purpose while providing a potential account of social change. The force of this account is that false consciousness is accounted for as created by the thinking agent herself, without posing a threat to her epistemic capacities as such. False consciousness is maintained as strict but not immutable, and through retaining human agency at the core of creation, this means that human beings are, to a certain degree, in charge of change and development.

References

Balibar, Étienne (1995) *The Philosophy of Marx* (London: Verso, 2014)
Benhabib, Seyla (1994) *The Critique of Instrumental Reason* in *Mapping Ideology* ed. Slavoj Žižek (London: Verso 2012)
Benjamin, Walter (1936) *The Work of Art in the Age of Mechanical Reproduction*. (London: Penguin, 2008)
Berlin, Isaiah (1963) *Karl Marx: His Life and Environment* (London: Oxford University Press)
Blunden, Andy (1999) *The Encyclopedia of Marxism*, *http://www.marxists.org/glossary* (22.02.2014)
Bourdieu, Pierre, and Eagleton, Terry (1994) *Doxa and Common Life: An Interview* in *Mapping Ideology* ed. Slavoj Žižek (London: Verso 2012)
Breckmann, Warren (1999) *Marx, the Young Hegelians, and the Origins of Radical Social Theory* (Cambridge: Cambridge University Press)
Brown, Norman (1959) *Life Against Death: The Psychoanalytic Meaning of History* (Middletown: Wesleyan University Press)
Burge, Tyler (1979) *Individualism and the Mental* in *The Extended Mind* in *Philosophy of Mind - Classical and Contemporary Readings* (2002) ed. David J. Chalmers (Oxford: Oxford University Press)
Clark, A. (2008). *Supersizing The Mind: Reflections on Embodiment, Action, and Cognitive Extension* (Oxford: Oxford University Press)
Clark, Andy and Chalmers, David (1998) *The Extended Mind* in *Philosophy of Mind - Classical and Contemporary Readings* (2002) ed. David J. Chalmers (Oxford: Oxford University Press)
Eagleton, Terry (1991) *Ideology - An Introduction* (London: Verso, 2007)
Engels, Friedrich (1895) *Engels to J. Bloch In Königsberg*, trans. Brian Baggins (Marxist Internet Archive, 1999)
Gallagher, Shaun (2013) *The Socially Extended Mind* in *Cognitive Systems Research* vol. 25-26
Hook, Sidney (1950) *From Hegel to Marx* (New York: Columbia University Press, 1994)
Houlgate, Stephen (2005) *An Introduction to Hegel – Freedom, Truth and History* (Oxford: Blackwell Publishing)
Kolakowski, Leszek (1978) *Main Currents in Marxism* (London: Norton, 2005)
Locke, John (1690) *Second Treatise of Government* ed. C.B Macpherson (Cambridge: Hacket, 1980)
Marcuse, Herbert (1964) *One Dimensional Man* (Boston: Beacon Press, 1991)
Marx, Karl (1867) *Capital vol.1*, trans. Ben Fowkes (London: Harmondsworth, 1976)
Marx, Karl (1883) *Capital vol.3*, trans David Fernbach (London: Penguin, 1991)
Marx, Karl (1843) *Critique of Hegel's Philosophy of Right*, trans. Jolin, Annette & O'Malley, Joseph (1970) http://www.marxists.org/archive/marx/works/1843/critique-hpr/ (10.01.2014)

Marx, Karl (1844) *Economic and Philosophic Manuscripts of 1844*, trans. Mulligan, Martin (1959) http://www.marxists.org/archive/marx/works/1844/manuscripts/hegel (06.01.2014)
Marx, Karl (1843) *On the Jewish Question* in *The Marx–Engels Reader* ed. Tucker, Robert C. (New York: Norton, 1978)
Marx, Karl (1851) *The Eighteenth Brumaire of Louis Bonaparte* trans. Ben Fowkes in *Surveys from Exile* (London: Harmondsworth, 1973)
Marx, Karl (1844) *Third Economic and Philosophic Manuscript of 1844* trans. Martin Mulligan (Marxist Internet Archive, 2009)
Marx, Karl (1845) *Thesis on Feuerbach* in *The Marx–Engels Reader* ed. Tucker, Robert C. (New York: Norton, 1978)
Marx, Karl and Engels, Friedrich (1848) *The Communist Manifesto*, (London: The Collectors Library, 2004)
Marx, Karl and Engels, Friedrich (1844) *The Holy Family*, trans. Dixon, Richard & Dutts Clement (1956) http://www.marxists.org/archive/marx/works/1845/holy-family/ (10.01.2014)
Marx, Karl and Engels, Friedrich (1846) *The German Ideology* (New York: Prometheus Books, 1976)
Marx, Karl and Engels, Friedrich, *Selected Works*, vol.1, (London, 1962)
Plato (380 BC) *Republic* trans. G.M.A Grube (Indianapolis: Hackett, 1992)
Putnam, Hilary (1975) *The Meaning of Meaning* in *The Extended Mind* in *Philosophy of Mind - Classical and Contemporary Readings* (2002) ed. David J. Chalmers (Oxford: Oxford University Press)
Reich, Wilhelm (1933) *The Mass Psychology of Fascism,* trans. V.R. Carpagno (Harmondsworth: Penguin, 1975)
Rosen, Michael (1996) *On Voluntary Servitude*, (Oxford: Polity Press, 1996)
Rupert, R. (2004) *Challenges to the Hypothesis of Extended Cognition* in Journal of Philosophy, 101, 389–428
Wilson, Robert A. and Foglia, Lucia (2011) *Embodied Cognition* in *The Stanford Encyclopedia of Philosophy* ed. Edward N. Zalta

Research Article **Open Access**

Max Skipper Griffiths

Review of Carruthers' Massive Modularity Thesis

DOI 10.1515/pipjp-2016-0005

Abstract: According to Carruthers' (2006) massive modularity (MM) thesis, the central systems of the mind are widely encapsulated and operate via heuristics and approximation techniques similar to those found in computer science. It follows from this, he claims, that widely encapsulated central systems are feasibly tractable. I argue that insofar as Carruthers uses this weakened definition of encapsulation, his thesis faces a dilemma: either is a misnomer (Prinz, 2006) and therefore unrecognisable as a version of MM, or it isn't, and must put forward a convincing version of MM (Samuels, 2006). I claim that Carruthers' commitment to this claim about central systems meets this challenge by adopting an understanding of central systems whose information-frugal and processing-frugal operations allow for feasible tractability. I conclude that the CWT provides a plausible and distinctive account of MM.

Keywords: Massive Modularity Thesis (MM), Heuristics, Tractability, Carruthers

1. Introduction

In an elegant defence of Massive Modularity (MM), Carruthers (2006) sketches a revised account of the human mind. His thesis proceeds from two claims. The first claim is positive; it establishes that the input systems (e.g. those responsible for perception) and output systems (e.g. those responsible for motor movement) of the brain are modular in a strong sense.[1] The second claim is less positive; it claims that the central systems (e.g. those dealing with reasoning and decision making) are modular in a very weakened, "wide-scope" sense of the word. I'll call this the Central Wide-scope Thesis (CWT).

> *CWT*: The processes of central cognition are weakly wide-scope encapsulated *(WEM)*.

Carruthers' focus is on the second of these two claims (i.e. CWT). This claim is often objected to on the grounds that it necessitates a denial of the connection between encapsulation and modularity. On this point Prinz (2006) and Samuels (2006) see Carruthers' notion of module as a misnomer. I develop this objection into a novel challenge to defenders of CWT. I claim that those who hold CWT face a dilemma. The dilemma is:

> *The First Horn*: *WEM* cannot, plausibly, be described as encapsulation, and is therefore a misnomer.
>
> *The Second Horn*: *WEM* can, plausibly, be described as encapsulation, but must put forward a convincing account of modularity.

1 In what follows I shall use the word 'system' interchangeably with the terms module and mechanism.

***Corresponding author: Max Skipper Griffiths,** Univeristy of Bristol

So, either CWT is not close enough to the accepted definition of encapsulation and is therefore espousing a thesis based on a misnomer, or it is close enough to the accepted definition but must put forward a convincing modularity thesis.

So, insofar as Carruthers and others hold CWT, they face an explanatory burden: they must explain either why CWT is not a misnomer, or, if the misnomer challenge is met, why it is not an unconvincing account of MM. I claim Carruthers' commitment to this negative claim about central systems meets this challenge by adopting an understanding of central systems whose information-frugal and processing-frugal operations overcome tractability problems posed by the Computational Tractability Argument (CTA). I conclude that the CWT provides a plausible and distinctive account of MM.

My argument has five stages. In the first stage, (section 2) I state some preliminary claims. In the second stage (section 3) I explain and clarify Carruthers' thesis. In the third stage (section 4), I examine the two horns of the dilemma. Then, in the fourth stage (section 5) I look at objections and replies. I conclude in the fifth stage (section 6).

2. Preliminaries

As a preliminary, it is necessary to say something about how I shall view the position held by an advocate of massive modularity (MM).[2] The plausibility of MM, as I will understand it, rests on three main claims. According to the first claim the human mind is composed (largely or completely) of mental modules. I will call this the Module Thesis. According to the second claim the human mind has many of these modules. I will call this the Multiplicity Thesis. According to the third claim modularity is found in central regions of the human brain responsible for cognitive capacities such as reasoning and problem solving. I will call this the Central Thesis.[3] In what follows, I take these claims to represent the basic theoretical commitments that proponents of MM affirm. These claims may or may not be correct. I shall begin by assuming, however, that they are acceptable enough for the MM proponent to endorse.

I will also - following Carruthers (2006) - not take a position on the holistic arguments against MM. In particular, I will not defend against the challenge that central cognition is global: specifically, that central cognitive processes – paradigmatically, reasoning – are sensitive to the properties of the agent's entire set of background beliefs. Instead, I shall assume for the sake of argument that these arguments pose no threat to the various theses of mind as conceived by proponents of MM.

Another point worth mentioning is how I shall view Carruthers' position with regard to competing accounts. On this point I will remain agnostic. The Challenge, as I see it, is for Carruthers to put forward his thesis in a convincing way. So, insofar as Carruthers holds CWT, he offers a plausible alternative understanding of central system modules. The challenge, as I shall argue, is for him to present that alternative in a plausible way.

I put forward four conditions he must satisfy to meet this challenge. I propose four basic requirements.

Firstly, a convincing account must meet an informative condition. That is to say, it should provide a clear explanation of the underpinning features of its account of modularity.

Secondly, the thesis must meet a '*significance condition*'. It should therefore say something nontrivial about modularity – a point expressed by both those who hold a non-modular view about central cognition (e.g. Samuels (2006)), and those who hold that central cognition is modular (e.g. Sperber (1994, 2002)). To see this more clearly, consider how an uninteresting modularity hypothesis might, for example, espouse the view that the mind is made up of dissociable modules. This may be true, but it is a trivial point and hence uninteresting and insignificant (cf. Samuels, 2006).

2 Adherents of this approach, whose original proponents are noted for their work in Evolutionary Psychology (Cosmides and Tooby, 1994; Sperber, 1994; Pinker, 1997; Barrett, 2005; Barrett & Kurzban, 2006), claim that the mind is like a Swiss army knife: assembled of specialised tools (or devices) designed for specific tasks.

3 This is contrasted with those modularity theorists who claim – qua Fodor (1987) – that modularity is only found in peripheral regions of brain architecture.

Thirdly, the thesis must be close enough to the ordinary understanding of modularity to satisfy a '*consistency condition*'. As a result, it cannot simply ignore standardly accepted concepts or be at odds with background theory. It must not draw invalid conclusions from unsound premises.

Fourthly, an adequate account must meet the '*distinguishable condition*'.This condition must be met in order to prevent, for example, a non-modular account masquerading as an MM account. For example, whereas Fodor (2000) thinks that the central regions of the mind are non-modular, Sperber (2006) explicitly denies - he claims that the central regions are modular. Carruthers must situate himself in this debate in a way that doesn't assume both modular and non-modular architecture.

Consequently, I take it that Carruthers must, when clarifying his thesis, satisfy all of these conditions in order to meet the explanatory burden. However, as I will show, whilst he may meet some of these requirements it is by no means clear that he meets all of them.

3. Carruthers on Modularity

In this section I explain Carruthers' MM thesis. In particular, I focus on the key features of his revised Fodorian criteria.[4] Carruthers believes that brain architecture consists entirely of mental modules. Each module is responsible for a specific feature of cognition, including high-level tasks like problem solving and planning. There are, he claims, five distinctive features that characterise a module. Specifically, a module must satisfy five criteria:

1. Dissociability
2. Domain specificity
3. Mandatory operations
4. Localizability
5. Central inaccessibility

Note firstly that Carruthers' leaves encapsulation out of the above criteria. So understood, we might reasonably describe this as a unique feature of his criteria – a judgment of distinctiveness that is motivated by comparison with Fodor (1983). According to Fodor (1983) central systems are non-modular while peripheral systems are modular and, he claims, encapsulation is the most distinctive criterion of a module (Fodor, 1983, p. 72-73). I discuss why this is important in due course. But for now it will be useful to begin clarifying these concepts.[5]

Firstly, dissociability. This notion refers to idea that a cognitive system can become impaired through neurological damage with little or no effect on the operation of other systems.

Secondly, domain specificity. This notion refers to the determinate class of representations that a cognitive system can receive as input (Carruthers, 2006, p. 5). It is, therefore, this class of representations that activate a system to initiate it into operation. Along with the other four features of a cognitive mechanism, this notion places architectural constraints on the representations a cognitive system can calculate. Furthermore, in the course of its processing, the only information a module can access is what it receives as inputs together with a determinate body of information stored within its system, which affects its internal operations – paradigmatically, its 'subsidiary database'. In addition, and perhaps most importantly, these two features – input to a system and the processing database of a system – should be distinguished from one another (Sperber, 2002; Carruthers, 2003). I discuss the importance of this distinction in due course.

Thirdly, mandatory operations. To say that a cognitive system is mandatory in its operations is to claim that it is automatic – working below the level of conscious control (cf. Bargh and Chartrand, 1999).

Fourthly, neural localizability. According to this modular feature, a cognitive system is associated with

4 Fodor's (1983) criteria consists of nine conditions: mandatory operation, shallow outputs, fast operation, dissociability, fixed neural architecture, characteristic ontogenesis, informational encapsulation, inaccessibility, domain specificity.

5 As Carruthers writes, "if a thesis of massive modularity is to be remotely plausible, then by module we cannot mean 'Fodor-module'" (Carruthers, 2006, p. 12).

neural localizability to the extent that it is instantiated in neural circuitry – circuitry that may perform more than one cognitive function.

Finally, central inaccessibility. According to the standard definition, an inaccessible system is one whose internal operations and proprietary database are inaccessible to other systems held in the mind (Carruthers, 2006, p. 5).[6]

Overall these criteria seem to offer a promising foundation for the proponent of MM. Some of the details may of course be disputed, but for present purposes they serve simply to illustrate the general idea at the heart of Carruthers' thesis. Furthermore, the criteria do not (at this stage, at least) appear to violate my requirements for an adequate account of MM. That is, the criteria are well explained and descriptively adequate so it meets the instructive condition; it's consistent with background theory, so does not fail the consistency condition; and it is distinct from alternative accounts (e.g. Fodorian modularity) so it doesn't violate the distinguishable condition.

Carruthers' account centres on to two independently credible claims. First, he argues, the input and output systems of the mind are "narrowly" modular or 'encapsulated' with respect to the class of representations those systems can receive. Specifically, this class of representations determines the informational properties a module can access and internally process and, hence, the range of inputs it can compute. This information must, according to the standard definition, be "less than all of the information at the disposal of the organism whose cognitive faculty it is" (Fodor, 1983, p. 25). In contrast to (3) domain specificity, encapsulation concerns the class of representations that a system can use once it has been activated through a relevant class of domain specific representations.[7]

Low-level perception and phonology, neither of which draws on representational states like beliefs or goals, are paradigmatic examples in the literature of narrowly encapsulated modularity (*NEM)* – a claim accepted by nearly all modularity theorists (e.g. Fodor, 2000; Samuels, 2006; Sperber, 2002; see Prinz, 2006, for arguments against).[8] More precisely:

> *NEM*: a cognitive mechanism M is narrowly encapsulated if, in the course of its processing it cannot be affected by most of the information held in the mind.

The Müller-Lyer illusion, construed as a paradigm of peripheral modularity and an example of *NEM,* is a case in point.[9] Taking this example we can – plausibly – infer that the system responsible for visual length perception is unable to draw on the belief that, visual appearances notwithstanding, the two lines are of equal length. This system would be unable to utilize other information even if it were relevant for length perception. This leaves us with an archetypal input-module[10], which can access and process "sensorily-transduced" information, but can't access any of the stored information held in the mind (cf. Carruthers, 2006, 59).

For Carruthers, however, the modular status of the central systems require refinement. The refinement concerns the 'modular status' of the central systems of the brain. (His focus, therefore, is to build upon the Central Thesis outlined in section 2). As Carruthers sees it, these central systems are *WEM*; that is, they have a limited access to an unrestricted body of information. As a result, they can receive most of the information held in mind, but cannot process it all at once. I will refer to this widely encapsulated sense of modularity as *WEM* (Carruthers, 2006, p. 58).

---

6 This argument is problematic, however. One may object, as Carruthers does, that all mechanisms are completely accessible to other mechanisms. But in doing so he fails to acknowledge the possibility that *some* mechanisms may be able to access *some* of the information contained within other modules.

7 Four points about encapsulation. First, encapsulation is an enduring and permanent characteristic. Second, it is not a product of performance factors, such as e.g. motivation, knowledge, or attention. Third, it is cognitively impenetrable; that is, it cannot be changed through beliefs or intentions. Finally, the access relations are modal in the sense that they relate to the set of information a system can make use of in the course of its processing (Cf. Samuels, 2006).

8 Another example of NEM is how facial recognition system(s) only has access to information about previously encountered faces.

9 Paradigmatic examples such as length perception or phonological processing are given in the literature.

10 As argued for by Fodor (1983).

> *WEM*: a cognitive mechanism M is widely encapsulated if, in the course of its processing, it can access most of the information held in the mind, just not all at once.[11]

To summarise, then: whereas *NEM* cannot be *affected* by most of the information held in the mind, *WEM* can *access* most of that information, just not all at once. Specifically, *NEM* can, for instance, access and process "sensorily-transduced" information, but cannot be affected by most of the stored information held in the mind (cf. Carruthers, 2006, 59). And with respect to the abovementioned distinction between the input to a system and its processing data-base Carruthers would claim that within the processing data-base of a *NEM* there would be some determinate body of information that can affect the operations of its system. The implication being that most of the remaining information held in the mind cannot affect the operations of that system *(*except, of course, that information which is taken as input).

According to Carruthers, *NEM* cannot be a wide-ranging model for how we ought to understand modularity. This is because if we are aiming to characterise central system modularity, then we can no longer think of encapsulation as a matter of isolating the system from stored information. The reason is simple: central systems will often need to operate on stored information as input.

With this in mind, if central systems only require *WEM*, then Fodor's (1983) argument against central modularity, and in particular his claim in support of the conceptual connection between encapsulation and modularity is, following *WEM*, presented with a challenge (see Fodor, 1983, pp. 107-110). The challenge is to explain why, given *WEM*, we should not deny the following radical claim (RC):

> (RC) We can deny that *NEM* is required for modularity.

In light of this, we can – contrary to Fodor – deny the claim that the more encapsulated a central system is (with respect to the amount of information it can process), the more modular the system required to accomplish a task will be (i.e. the nearer to *NEM* it will be). According to Carruthers, either a system can access *no* stored information in executing its algorithms, or:

> [...] it can only access a limited data-base of information that is relevant to the execution of those algorithms (in which case the system is encapsulated to a degree inversely proportional to the size of the data-base) (Carruthers, 2006, p. 19).

This may of course be rejected. But it appears to be a good strategy for proponents of MM. Non-modular accounts (e.g. Fodor, 2000; Samuels, 2006), by contrast, cannot take this approach. They explain the information processing of central systems in terms of being strongly global. For instance, Fodor (2000) agrees with Carruthers' positive claim (outlined in the introduction) that the input and output systems of the mind are modular (the Central Thesis), but rejects CWT in favour of a non-modular approach (see Fodor, 1983, pp. 107–108). I will call this the Global Central Systems Argument (GCA). We can represent this as follows:

1. Central systems subserve global cognitive processes.
2. Global cognitive processes require less encapsulation; that is, they are inversely proportionate.
3. There exists a strong connection between encapsulation and modularity.
4. Hence, (1,2,3) there exists an inverse relationship between global cognitive processes and modularity (i.e. the more global the process, the less modular the system).
5. Hence, (4) Central systems process information globally.

'Global' here means, roughly speaking, those cognitive processes that have access to effectively all of our belief system i.e. beliefs, goals and intentions. This quick sketch of Fodor's argument may of course be disputed. I will assume for present purposes, however, that it will suffice to warrant acceptance since my

11 Carruthers' own formulation is as follows: "the system is such that it *can't* be affected by most of the information held in the mind in the course of its processing" (Carruthers, 2006, p. 19).

aim is not to assess the plausibility of this argument. Rather, I use it instead firstly to illustrate the difference between Carruthers (2006) and Fodor's (1983, 2000) approaches to modularity, and secondly, highlight the challenge it poses to a modular approach towards central systems.

One point to mention about this argument is that Carruthers' RC directly challenges its third premise. Another point, related to the first, concerns how if Carruthers' CWT is true (i.e. that the processes of central cognition are *WEM*) and his RC thereby sound, then it follows that Fodor's GCA is false. And so, according to Carruthers, we can plausibly claim that central systems are modular in the sense of *WEM*.

However, given the importance of *NEM* to Fodor's modularity thesis, and in the absence of strong supporting arguments for *WEM*, all this challenge represents, as I understand it, is that central cognition *might* be modular in a non-Fodorian sense – but, crucially, not that its systems are in fact *WEM*.

Let us take stock. As Carruthers' sees it, *WEM* entails that central systems can access most of the information held in mind, but cannot process it all at once. Hence, whilst this does not immediately entail the falsity of GCA it does threaten to challenge it.

But what is the explanatory ground for this class of *WEM* central systems? We can summarise the answer as Carruthers explaining the existence of central systems in terms of "search heuristics and stopping rules". Specifically, processing systems whose operations are "both information-frugal and processing-frugal" (and hence, encapsulated in the *WEM* sense). Which is to say, those central systems must only access a small sub-set of the total obtainable information while executing their tasks (Carruthers, 2006, pp. 16, 22). Carruthers' stresses this point:

> "[...] a module can be a system that *must* only consider a small sub-set of the information available. Whether it does this via encapsulation as traditionally understood (the narrow-scope variety), or via frugal search heuristics and stopping rules (wide-scope encapsulation), is inessential. The important thing is to be both information-frugal and processing-frugal" (Carruthers, 2006, p. 20).

He arrives at this conclusion via a Computational Tractability Argument (CTA). This argument is essential, I think, for forming his overall picture of brain architecture and function, and more specifically within that architecture, the state, structure and processes of a constituent module. As a consequence, the force of Carruthers' thesis turns on the soundness of this argument. We can represent it – CTA – as follows:

1. The mind is computationally realised.
2. All computational mental processes are suitably tractable.
3. Only processes that are at least widely encapsulated are tractable.
4. Hence, if all mental processes are to be feasibly tractable and only (mental) processes that are at least *WEM* allow for such tractability, then all of those processes must be at least *WEM*.
5. Hence, (4) if all of those mental processes are at least WEM, then the mind is massively modular.
6. Hence, (5) the mind is massively modular.[12]

Before clarifying CTA, it is necessary to point out that this argument assumes the classical computational theory of mind, according to which cognitive processes can be realised as algorithmically specifiable mental representations. This premise may be rejected. But I shall assume – following orthodoxy – that it is warranted. According to this assumption, cognitive processes can be realised in computational terms. It follows from this that cognitive capacities – such as, for example, visual perception, practical judgments and analogical reasoning – can be modelled in terms of input-output mapping with algorithms on which the systems of those cognitive capacities run. And as a consequence, if all else is equal, we should expect a system with a greater computational load, required to check many items of information, would demand a more complex algorithm. However, simply stipulating algorithms that demand computational resources (memory, processing power, etc.) beyond the level attributed to humans, would be inconsistent with real-time human performance. So, cognitive systems must be computationally tractable in the sense that mental processes are successfully completed in a finite amount of time, and do not require more time and resources,

12 Cf. Carruthers, pp. 13-14.

in the form of memory, information, and computational power, than human beings are reasonably understood to possess. This is called in-practice tractability. As a category of tractability – contrasted from in-principle tractability – this refers to the capacity of a system to undertake state changes at the same time as changes take place in the external environment.

It follows as a consequence of this that the computational theory of mind requires cognitive processes to be practically feasible. This means that the algorithms behind those cognitive processes must be tractable in two ways. First, the algorithms must be an appropriate level of complexity. Second, the computations of those algorithms should only draw on a limited set of information relevant for their computations (cf. Carruthers, 2006, p. 18). So, if a cognitive capacity, characterised by a specific function, cannot be computed within a realistic psychological and computational time limit (i.e. practically unfeasible), then it is said to be intractable.[13]

Proponents of MM claim that domain specific modules with the property of encapsulation and domain specificity are able to solve this problem. This, they claim, is because such modules only operate on the inputs of a restricted domain, thereby reducing the scope of information they must consider, select and revise in light of new information. And insofar as this is true, those modules process information faster than their non-modular, domain-general counterpart. By contrast, the non-modular view fails to overcome the challenge set because, with unconstrained access to the remainder of cognition a domain-general mechanism ostensibly fails to determine which operations, and which sets of information, are relevant for computing.

According to CTA, only at least *WEM* are feasibly tractable. Consequently, the mind must consist entirely of at least *WEM*. Hence, the mind must be MM. According to Carruthers, this argument presents a promising alternative to non-modular approaches (e.g. Fodor's GCA), which claim that central processes are global. This alternative approach views central systems as systems whose operations are both information-frugal and processing-frugal.

Let us take stock. So far we have seen that a system's algorithms are such that only a limited amount of information is ever consulted before a given task is finished. Assuming then that CTA is true, the following statement would be false: if cognitive processes are to be feasibly tractable, then the mind must be composed out of processing systems encapsulated in the sense of *NEM* (Carruthers, 2006, p. 13)). It follows from this, Carruthers claims, that the traditional strong argument for MM, which states the mind is composed entirely of *NEM-type* modules, would collapse too. But we still have the argument that if cognitive processes are to be tractably realized, then the mind must be *WEM*.

Hence, for Carruthers central systems do not engage in an exhaustive search considering all of the elements of information held in the mind – if they did, then that would render those systems intractable. Instead in order to ensure feasible tractability central systems must, together with the sort of approximation procedures found in computer science and Artificial Intelligence (AI), use "search heuristics and stopping rules" (Carruthers, 2006, p. 16)). This approach to human decision-making is not optimal. Rather, it employs a method that is *satisficing* – producing solutions that reach some threshold level for satisfactory decision-making.

With heuristics Carruthers' account can avoid time-consuming database searches by identifying a relevant subset of information (beliefs) relevant to a cognitive task. This procedure might not hit upon relevant solutions all of the time, but that is not a problem because human reasoning processes are mostly suboptimal. So, it follows that encapsulation is not the only way to ensure tractability. But what evidence is there for this?

That humans use heuristics in performing cognitive tasks is empirically supported. To see this, consider the Take the Best heuristic. This heuristic requires that a given system draw on additional information concerning the items of information related to the task in question. Crucially, though, it does not look at *all* the information related to those items. Instead, it "searches for the item of information concerning the two target items that has most often been found in the past to discriminate between items of that type along the

13 The roots of this intractability can be traced back to the well known 'frame problem', which Fodor describes as the problem "of putting a frame round the set of beliefs that may need to be revised in the light of specified newly available information" (Fodor, 1983, p. 112).

required dimension" (Carruthers, 2006, p. 14). Gigerenzer *et al.* (1999) have shown that this heuristic can perform at the same level as far more sophisticated processing algorithms, and all whilst being more frugal in the information that it uses and the requests that it places on the system's resources (Gigerenzer et al., 1999).

Real-world computational systems can search through an enormous database without incurring any significant computational cost. Internet search engines are a case in point, for they highlight the sense in which frugal, algorithmic processes enable central systems of the mind to calculate information fast enough to be practically feasible; then, Prinz claims that with heuristics "we can avoid exhaustive database searches even when a complete database is at our disposal" (Prinz, 2006, p.33).

This brings into focus the sense in which frugal, algorithmic processes enable central systems of the mind to calculate information fast enough to be practically feasible. And this feasibility is necessary because, all things equal, humans respond in real time to stimuli, and therefore, require central systems with the capacity to execute cognitive processes virtually immediately (i.e. within milliseconds).

Before turning to the discussion of CTA, it is worth developing the account a little further in the context of encapsulation. Carruthers explains that it is frugality, not encapsulation, that is necessary to ensure feasible tractability in central systems, which appears to explain why he omits encapsulation from the five criteria a module must satisfy. On his view, search heuristics and approximation techniques which access no more than a small sub-set of total information available solve the problem of intractability. Therefore, given that this modular property – encapsulation – is left out of Carruthers' list, it represents a prima facie case for saying that there are strands to CWT that are distinctive. In this sense CWT is distinguishable from non-modular accounts. So, at this stage, it appears to satisfy the *distinguishable* condition, previously outlined in section 2.

So if CTA is sound, then it strongly supports the core claim behind CWT: that central systems are *WEM*. As a result, if we are to accept the soundness of CTA we possess a sufficient, direct argument that would render non-modular accounts of central systems (e.g. Fodor's GCA) computationally intractable. And so, at the very least advocates of those non-modular central systems, such as Fodor (2000), face a prima facie challenge to explain the tenability of those non-modular systems. Carruthers supports this position with a bottleneck argument. It runs as follows. Unlike a multiplicity of modules, general-purpose systems face a tractability problem because by having access to effectively all of our belief system (i.e. beliefs, goals and intentions) those general-purpose systems entail a processing 'bottleneck' that would prohibit the formation of beliefs in real time. One may object to this argument (see Samuels, 2006). But this clearly falls outside the scope of the present examination.

Carruthers' thesis – and especially my brief outline of it – clearly requires further philosophical scrutiny. But, assuming that it withstands this scrutiny it can be thought to provide a promising response to Fodor's GCA. This is because it gives a central place to the practical utility of believing in search heuristics in explaining how central systems can be responsive to the huge range of information within the remainder of cognition. But there is a complication. One might think that with respect to information processing, this notion of *WEM* is counter-intuitively *unencapsulated* – after all encapsulation entail sharp constraints on the amount of information a module can receive does it not?

4. The Dilemma

In this section I outline, clarify and assess the dilemma facing the defender of CWT. My initial focus will be on the first horn: that CWT is a misnomer. I claim that whilst this argument does not immediately entail the falsity of CWT, it presents a strong prima facie challenge to it. The challenge is to explain why *WEM* is modular. Following this I then claim that, assuming CWT is able to withstand this challenge (i.e. can be plausibly described as modular) it falls on the second horn of the dilemma according to which an explanation must be given as to why it is not an unconvincing account of MM.

4.1 Misnomer

Given the foregoing discussion, defenders of Carruthers might be tempted to think that CWT provides a plausible explanation of central cognitive systems. On this view, not only is CWT a sound metaphysical thesis in that it offers a plausible account of the sense in which properties of the mind exist, but also, in telling us how we should engage in normative enquiry about those properties, it presents us with an epistemological claim; and furthermore with a semantic one, since it tells us about the relationship between a specific set of theoretical words and their meanings.

Despite its promise however, Carruthers' understanding of encapsulation (i.e. *NEM* and the non-standard *WEM*) is problematic. One obvious problem concerns the inaccuracy of the use of the term encapsulation. The idea here is that, insofar as *WEM* can "draw on most of the remainder of cognition, though not all at once," one may object that it cannot be reasonably understood as an instance of encapsulation. As such CWT threatens to miss the consistency condition to make it an adequate account of MM. Therefore, Carruthers commitment to the view that central systems are, via heuristics, *WEM* belies an understanding of central systems at odds with encapsulation, as it is standardly understood. This is the first horn of the dilemma. I will call it the Misnomer argument (MA); it establishes that if Carruthers' thesis works, and hence is true, then it follows that *WEM* is not, by definition, encapsulated. This argument – MA – can be represented as follows:

1. If Carruthers' thesis is true, then central modules are *WEM*.
2. If central modules are *WEM*, then most of the information in the human mind is accessible, but not all of it can be accessed at once.
3. But encapsulation implies restrictions on information into a module.
4. Hence, *WEM* does not have the distinctive property of encapsulation required to make it modular.
5. Hence, either Carruthers' thesis is false, or *WEM* is modular even though does not have the distinctive property of encapsulation required to make it modular.

In order to make this stand out, consider that to overcome the first horn of the dilemma Carruthers must somehow show that *WEM* (i.e. a limited access to an unrestricted body of information) can be legitimately described as an instance of encapsulation (and hence, modularity). This argument – appealed to by Samuels[14] - therefore reflects a warranted scepticism about the plausibility of CWT, questioning whether Carruthers is entitled to call *WEM* modular.

I do not think, however, that MA is a knockout argument against CWT. It fails to be a knockout argument because premise 4) is not straightforward. That is to say, it is not obviously clear that *WEM* is unrecognisable as an instance of encapsulation. If it is not, then the resulting argument fails to establish premise 5). Indeed, defenders of Carruthers may object to MA on the basis that insofar as central systems, as they conceive of them, cannot at any one time process all of the information held in the mind, then those systems meet the core condition required for encapsulation. This core condition, for Carruthers, is that those central systems "can't access more than a small sub-set of the total information available before completing their tasks" (Carruthers, 2006, p. 16). Specifically, if computational tractability requires *WEM,* then for each system and sub-system it must be the case that:

1. We are able to identify and differentiate its input from its processing database (assuming it has one).
2. The system's processing database must contain only a small sub-set of the total amount of information available.

If heuristics propose frugality without either (1) or (2), then it follows that CTA would threaten to undermine CWT. But given that the operations of *WEM* are (via heuristics) both information-frugal and processing-frugal, we can reasonably claim that those systems only access a small sub-set of the total accessible information (from the total set they can access) during their processing. And therefore *WEM* arguably meets

14 Though Samuels (2006) only mentions this as a brief remark, in contrast to the lengthy formulation and discussion that I have given it here.

the abovementioned requirements (1) and (2). The result is that *WEM* appears to be both encapsulated and tractable. And therefore there is good reason to believe that MA is a straw man and a false argument.

Against this, however, one might claim that you cannot incorporate heuristics into a modular architecture. Hence, on this view, by acknowledging the merits of heuristics, Carruthers allows that the argument for encapsulation vis-à-vis central cognition is unsuccessful since encapsulation is replaced with heuristics.

Two responses. The first is to mention that, like an encapsulated module a heuristic module can ease the cognitive load through adopting the satisficing heuristic of stopping a search when it has found an item of information that is acceptable for use in its current task. Furthermore, simple heuristics offer a way to regulate interactions among modules, thereby engendering limits on the information a module can access during the course of its processing. For example, the mind-reading system, which is required to query a large range of other systems for data relating to social norms, values and behaviours, could employ heuristics to acquire that data (cf. Carruthers, 2006, p. 11).

The second response is to say that simple heuristics can fit within modular architecture, as shown by the ecological rationality heuristic. According to the ecological rationality heuristic, there exists a range of different environments in which heuristics will function in an accurate and reliable way. We can think of these heuristics as having been selected (under evolutionary pressures) to solve problems within those environments. This requires that they hold within them, or 'encapsulate', a level of information that is sufficient for its immediate objectives – in such a way as to be described modular.

The Recognition heuristic, according to which when an agent only recognises one of two items they should assign the recognized item a higher value, is another case in point. This is because the only information concerned with the recognition calculation needs to be accessed. And so, the recognition heuristic merits the standardly understood label, 'module'.

However, one might object on behalf of MA that the problem here has to do with Carruthers explanation of CWT. The reason is this. If Carruthers does rely on this problematic understanding of *WEM*, then it seems strange that it is only briefly discussed near the end of the 2006 compared to his lengthy discussion of the supporting arguments from biology, task specificity and computational tractability. Furthermore, this is confounded by the fact that this is not a thesis either explained by or argued for in the writings of any other theorists of modularity. Carruthers is putting forward the non-standard view. It is he who has applied problematic vocabulary in the context of central systems. It follows, then, that it is he who must bear the explanatory burden that that necessitates.

One response to the first challenge is to concede that the CWT is vague and ill-defined but suggest that, properly understood, this concession does not undermine the account. And this is because lack of specificity is an objection that can (plausibly) be raised against other modularity architecture, since there are several ways in which we might describe the concept 'module' (see: Prinz, 2006).

Both the scope and precision of this defence are problematic, however. Simply pointing to the vagueness of other modular architectures is a weak form of counter argument. If ambiguity is a quality that can be proven to apply to these other accounts, then so be it; the proponent of this view can easily show this. However, this does not support the stronger claim that these rules should be substituted by the CWT.

One other means by which a defender of the CWT might answer this problem is to point out that Carruthers' thesis does not appear to violate my criteria for an adequate account of MM. The reasons are as follows. The CWT is well explained and descriptively adequate, so can be said to meet the *informative condition*; it is consistent with background theory on MM, heuristics, and computational theory, so does not fail the *consistency condition*; and finally, it is distinct from alternative accounts (e.g. Fodorian (1983)), so does not violate the *distinguishable condition*. It seems to follow that *WEM* can overcome the first horn of the dilemma (that it can, plausibly, be described as encapsulation). As for the second horn, I shall now assess whether it meets this challenge, that is to say, whether it is a convincing account of MM.

4.2 Convincing Account

In order to assess whether CWT is a convincing account, I shall now look at whether CWT meets the remarkability and consistency conditions required to satisfy the explanatory burden. I focus only on these two for I take it that the distinguishable and informative conditions have been met.

Remarkability

One objection to CWT is that it is an uninteresting account of MM. This objection runs as follows. *WEM* is simply what you get by denying exhaustive search; but given that no one thinks exhaustive search is characteristic of human cognition, then *WEM* is simply stating the obvious point that the mind makes use of heuristics; therefore, CWT is neither distinctive nor interesting (cf. Samuels, 2006, p. 45).

Note firstly that "interesting" here may either be referring to CWT as a general hypothesis or its form of modularity (number and degree of modules) and its function (i.e. to avoid intractability).

Let us begin with its form and function. We can distil the essence of Carruthers account in terms of modules processing the inputs of a small sub-set of the information held within cognition. This is not a unique picture of the mind. But the use of heuristics is. MM formulated with this notion is interesting not only because it overcomes tractability problems via a programme of heuristics consistent with encapsulation, but also because it claims there are a large number of such mechanisms (Samuels, 2006, p. 39). This is interesting not only because CWT argues that central cognition is modular in the teeth of evidence that central cognitive processes are global, but also – as shown already – is distinguishable from its modular and non-modular counterparts. Insofar as this is true, CWT provides an interesting account of cognitive architecture.

I also take it that as a general hypothesis Carruthers' MM thesis is distinctive and interesting. This is because it is sustained by, and consistent with, the understanding of the simple heuristics research program. And not only is employing heuristics to solve tractability problems interesting but the *way in which* it is argued for is, too. As such, we have good reason to describe the account as a convincing account, which answers the challenge of the second horn. I will now assess this claim, alongside questions concerning the consistency condition, by looking at two arguments given in support of WEM.

Consistency

At the outset of *Architecture of the Mind* Carruthers introduces three arguments for massive modularity – the argument from biology, the argument from task specificity, and the argument from computational tractability. Together with an analogy from AI, these arguments are essential for forming his overall picture of brain architecture and function, and more specifically within that architecture, the state, structure and processes of a constituent module. This, one can argue, shows just how convincing Carruthers account is.

There is reason for scepticism, however. This scepticism concerns how these arguments fail to marshal evidence in favour of the particulars of CWT. And given that the force of his thesis lies in the strength and range of this and the two other theoretical arguments for which he proposes a MM of brain architecture and function, a failure in any one of these arguments threatens to undermine the force of CWT. I shall consider One of these arguments, as well as the analogy from AI.

We can represent the argument from biology (AB) as follows (Cf. Carruthers, 2006, p. 25):

1. Biological systems are designed systems, constructed incrementally
2. Such systems, when complex, need a massively modular organisation
3. The brain is (i) a biological system, and (ii) complex
4. So the brain is massively modular in its organisation

In defence of premise (1) and (2) of his argument, Carruthers asserts that there exists a wide range of evidence from across biology that "complex functional systems are built up out of assemblies of sub-components". Following the proposition behind this statement – that functional complexity demands structural complexity – and serving as the seat of brain architecture and function, Carruthers concludes that the brain is massively modular in its organisation.

This is an intuitively plausible argument. But, as I shall now suggest, it still does not follow from AB – as Carruthers intends – that the mind is massively modular. The reason is the following. All that AB establishes is a strong case for the multiplicity thesis (discussed earlier in the preliminaries). All it shows, therefore, is that the mind is hierarchically organised into dissociable subsystem. It does not, however, follow from this that the central systems of the mind are massively modular (i.e. dissociable, domain specificity, and so forth). Furthermore, the claim that central cognition contains decomposable, dissociable subsystems is not inconsistent with non-modular accounts, which also espouse such views.

In light of this, we lack reason to believe that Carruthers can legitimately infer massive modularity of mind (in function) from massively modular organisation and structure – and AB misses its mark.

One means by which Carruthers might respond is to turn to the argument from AI. The argument from AI proceeds by identifying that during the last decade, researchers within AI have converged on modular architectures. This, so the argument goes, has been something they have converged on through trial and error, with the implication that successful design systems are structurally organised in a modular way. Indeed, many theorists believe that to ensure tractability, computational processes need to be divided into modular systems and subsystems (McDermott, 2001). Gigerenzer suggests "smart robots need to be ... equipped with special-purpose abilities without a centralized representation and computational control system" (Gigerenzer, 2001, p. 43). In light of this, we have reason to believe that human cognitive architecture has a modular organization. Whilst this argument may not be sufficient to claim MM is true, it provides further justification for believing that conclusion.

Non-modular systems have in practice turned out to be intractable while modular systems have not had any such problems. As a result, AI and robotics provide a strong case for MM, and in particular the need for modules. These modules – which Brooks (1999) calls "reactive behaviours" – are strongly encapsulated devices whose internal processes engender specific "predefined responses to environmental conditions" (Brooks, 1999, p. 90). However, the problem that emerges is whether the notion of "module" in this argument from AI bears any resemblance to, or in any relevant way describes, the modular architecture found in CWT. At first blush, it appears to describe a NEM-type module, not WEM. So the burden of proof rests with Carruthers to show why this is not the case.

Despite these shortcomings I believe that this still does not constitute a significant enough case against CWT to warrant the claim that it is unconvincing. Instead, I take it to simply weaken the force of CWT's argument to the extent that AB and the argument from AI are not sufficient to establish the truth of Carruthers' thesis. In the context of justification, this means that CWT is not completely truth-tropic (i.e. that it will *always*, in every relevant instance, yield true beliefs or conclusions).

5. Objections and Replies

I'll now consider two objections to my discussion of CWT:

Objection 1.

The objection concerns my failure to acknowledge the strength of objections to the MM thesis. Specifically, one strategy for undermining the MM thesis is via a developmental neurobiological argument. Consider, for example, Buller and Hardcastle's (2000) argument that there exists a discrepancy between recent neuroscientific evidence of neurobiological plasticity and the view of innateness held by proponents of MM, where innateness is understood in the Fodorian sense of developing according to "determined patterns under the impact of environmental releasers" (Fodor, 1983, p. 100). It is a consequence of this, they claim, that the massively modular position is undermined and suffices to warrant rejection.

One response to this challenge is to reject the strength of plasticity premise of the above argument. Specifically, there exists strong evidence that the cerebral cortex develops under the control of a particular set of genes whose specific, endogenous operations produce the neurons responsible for cognitive processes such as thought, perception and memory. In light of this evidence, the developmental plasticity premise of Buller and Hardcastle's argument is insufficient to preclude innate specification of brain architecture.

In this case, Buller and Hardcastle's argument is too quick. I claim that the soundness of my counter argument in favour of innateness renders a more balanced understanding of brain architecture, namely that the above discussion illustrates how the judgment of likelihood associated with the influence of innateness and developmental plasticity on brain architecture admits of degrees. Hence, for the most part the view of innate specificity and developmental plasticity of the brain will depend upon empirical considerations from cognitive science/neuroscience, and the strength of their net contribution, measured in degrees. Given these considerations, a module is more likely the product of the joint net effect of both innate and developmental processes.

Suppose for the sake of charity to Buller and Hardcastle that the above response fails to defuse the developmental neurobiological argument. It would still not follow that MM is false. This is because the proponent of MM need not take a strong stance on innateness at all, as with for e.g. Carruthers (2006) and Tooby and Cosmides (1992). Hence, given that Buller and Hardcastle falsely presuppose innate specification as a necessary condition of MM, when no such condition exists, and given that nothing in my argument turns on this, we can put this matter to one side.

Objection 2.

The second objection concerns what these theoretical models bring to bear on scientific research of cognitive architecture. The argument is as follows. To the reader not familiar with the literature, discussions within the modularity debate may appear to be philosophers trying to answer empirical psychological questions through a priori reasoning.

One means by which a defender of modularity theory might answer this problem is to claim that on a theoretical level modularity specifies theoretical, evolutionary explanations of mental phenomena and on an experimental level modularity allows for explanations that are plausible in relation to data collected from comparative psychology. Furthermore, the modularity framework from which its theories develop can be fallible (i.e., believing something that is false, e.g., every event having a cause) and defeasible (i.e., shown to be wrong by further evidence).

In addition, modularity theory not only provides explanations of empirical data such as, for example, the speed of cognitive processes and behavioural dissociations at the behavioural level; it also obtains evidence at the psychological and neurological level as with, for example, functional dissociations corresponding to neuropsychological case studies of language (Prinz, 2006). Modularity theory must therefore comport with what we have good reason to believe about human reasoning from detailed empirical inquiry, theoretical grounds or both.

It is possible to argue against this and claim that a priori arguments should not be used since questions about cognitive architecture are to a very large degree empirical. In this case however, the sceptical challenge is too quick. The concept of modularity is not undermined. This is because in response the defender of modularity theory can argue that insofar as philosophy embraces naturalism their inquiry is continuous with scientific inquiry. As such we should not view philosophy and the natural sciences as non-overlapping forms of inquiry but rather as complimentary disciplines. And insofar as empirical neurophysiological evidence is limited the former arguably supplements the latter.

Not only, then, have a priori arguments been used but they have been correctly used, beginning with empirically based premises (derived from a posteriori observable facts) and ending with logically warranted conclusions. It is, therefore, subject to empirical falsification. Furthermore, it is fairly normal in science to use what are in a sense "a priori" arguments to argue for empirical conclusions. These involve considerations of general plausibility, descriptive adequacy and so forth and allow us to determine how well an explanation coheres with our current system of beliefs rather than directly appealing to empirical

evidence. With this in mind, I think that the appeal to computational tractability in the modularity of mind debate is not solely philosophical rather than empirical. Rather, it seems to be a combination of both forms of inquiry.

6. Conclusion

I have presented a challenge to Carruthers' CWT. The challenge is that insofar as Carruthers holds CWT he faces an explanatory burden. The explanatory burden is to explain either why CWT is not a misnomer or, if the misnomer challenge is met, why it is not an unconvincing account of MM. I have shown that Carruthers successfully meets this challenge by adopting an understanding of central systems whose information-frugal and processing-frugal operations overcome tractability problems posed by CTA. I conclude that the CWT provides a plausible, distinctive and convincing account of MM.

References

Bargh, J. A. and Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, 54: 462–479.
Barrett, H. C. (2005). Enzymatic computation and cognitive modularity. *Mind & Language*, 20: 259–287.
Barrett, H. C. and Kurzban, R. (2006). Modularity in cognition: Framing the debate. *Psychological Review*, 113: 628–647.
Brooks, R. (1999). *Cambrian Intelligence: The early history of the new AI*. Cambridge, MA: Bradford books/MIT Press.
Buller, D. and Hardcastle, V. G. (2000). Evolutionary psychology, meet developmental neurobiology: Against promiscuous modularity. *Brain and Mind*, 1: 302–325.
Carruthers, P. 2005. The case for massively modular models of mind. In R. Stainton (ed.), *Contemporary Debates in Cognitive Science*. Blackwell.
Carruthers, P. (2006.) *The Architecture of the Mind*, Oxford: Oxford University Press.
Fodor, J. (1983). *The Modularity of Mind*. MIT Press.
Fodor, J. (2000). *The Mind doesn't Work that Way.* MIT Press.
Gigerenzer, G., Todd, P. M, & the ABC Research Group. (1999). *Simple heuristics that make us smart.* New York: Oxford University Press.
Gigerenzer & Selten. (2001). *Bounded Rationality: The adaptive toolbox*, MIT Press.
McDermott, D. (2001). *Mind and Mechanism*. MIT Press.
Pinker, S. (1997). *How the Mind Works*. Penguin Press.
Prinz, J. J. (2006). Is the mind really modular? In R. Stainto (ed.), *Contemporary Debates in Cognitive Science*, Oxford: Blackwell, pp. 22–36.
Samuels, R. (2006). Is the human mind massively modular? In R. Stainton (ed.), *Contemporary Debates in Cognitive Science*, Oxford: Blackwell, pp. 37–56.
Sperber, D. (1994). The modularity of thought and the epidemiology of representations. In L. A. Hirschfeld and S. A. Gelman (eds.), *Mapping the Mind*, Cambridge: Cambridge University Press, pp. 39–67.
Sperber, D. (2002). In defense of massive modularity. In I. Dupoux (ed.), *Language, Brain and Cognitive Development*. MIT Press.
Tooby, J. and Cosmides, L. (1992). The psychological foundations of culture. In J. Barkow, L. Cosmides and J. Tooby (eds.), *The Adapted Mind*, Oxford University Press.

Research Article **Open Access**

Tony Cheng

The Situational Structure of Primate Beliefs

DOI 10.1515/pipjp-2016-0006

Abstract: This paper develops the situational model of primate beliefs from the Prior-Lurz line of thought. There is a strong skepticism concerning primate beliefs in the analytic tradition which holds that beliefs have to be propositional and non-human animals do not have them (e.g., Davidson 1975, 1982). The response offered in this paper is twofold. First, two arguments against the propositional model as applied to other animals are put forward: an *a priori* argument from referential opacity and an empirical argument from varieties of working memory. Second, the Prior-Lurz situational model based on state of affairs as opposed to propositions is introduced and defended with two significant modifications. With this model of primate beliefs we can make progress in understanding how other primates can have certain mindreading capacity.

Keywords: beliefs; primates; mindreading; situational structure; state of affairs

1. Animal Mindreading and Animal Minds

Homo sapiens are minded beings. They are very likely mind-readers too, in the sense that they explain and predict (with confabulations) others' behaviors in terms of their grasps of others' mental states or episodes – doxastic states, motivational states, and so on. But what about other animals, in particular our primate relatives? More than three decades ago, Premack and Woodruff succinctly summarized relevant issues by asking, "Does the chimpanzee have a theory of mind" (1978)? In order to leave open various theoretical possibilities, we should not read too much into the word "theory" in this context, since it may turn out that some animals do read minds, but their relevant abilities do not consist in anything like a theory in some strict sense.

The answer to the question about primate *mindreading* conceptually presupposes the answer to the question concerning primate *minds*: what do primates' mental lives look like? Initially, one might naturally extend categories such as perception, belief, desire, emotion, and other familiar *human* mental categories to primates in understanding their minds. However, one salient difference between humans and other primates have to be taken into account – humans have language in the strict sense, while other primates do not. It is difficult to give a theoretical characterization of the distinctiveness of human languages that satisfies most people. Some philosophers appeal to compositional semantics, for example, but such an appeal is not without its opponents.[1] For our purposes a principled answer to this question is not needed; suffice it to say that no other animals can *talk* or *write* about abstract affairs, things not directly related to animal strives, and so on. Human languages significantly shape human minds, so when we attempt to extend our model to other animals this huge difference has to be put in the center of focus.[2]

1 See, for example, Stephen Schiffer (2003).

2 In this paper my concern is only belief state and I do not assume that the account can be generalized to other mental states. However, if the structure of primate beliefs is indeed very different from human beliefs it is reasonable to assume that other mental states of primates are also less abstract and intellectual, as in the case of beliefs. Similarly, although I talk about "other primates" in this paper most studies I will refer to are about chimpanzees; I do not assume that what I argue here can be extended to, say, orangutans and monkeys.

***Corresponding author: Tony Cheng,** University College London

Examples of human full-blown beliefs include "I believe *that* it is going to rain very soon." Some languages do not have a grammatical equivalence to "*that*-clause" in English, but that is a superficial difference that can be ignored for present purposes. Human beliefs typically involve an assertoric, sentential structure, no matter what the surface grammatical structures look like. Now it should be clear that even if other primates do have something like beliefs, those states would presumably be very different from the human variety, since unlike humans, they do not have a language in the strict sense. However, mainstream analytic philosophy seems to be oblivious to this. Here I side with Dale Jamieson that "it is downright harmful if, after assimilating thinking to having a thought, we go on to suppose that having a thought is the same as having a *propositional attitude*" (2009: 15, italics mine). Donald Davidson's ingenious arguments against animal beliefs and belief attributions come to mind (1975, 1982). The next section deals with, and ultimately rejects, this idea that beliefs have to be propositional. A positive model will then be suggested.

2. Against the Propositional and for the Situational

The general form of the Davidsonian argument starts with certain distinctive features of *human* beliefs and argue that since other animals lack the crucial precondition of that kind of belief they have *no* belief whatsoever. Here a leap from human beliefs to beliefs in general is made. To be sure, it is possible that at the end of the day Davidson is right in thinking that the only format of belief is *propositional*, linguistic belief; but it is methodologically vicious if philosophers stick to *a priori* arguments so much that they do not bother paying attention to empirical findings and relevant conceptual clarifications at all (e.g., the possibility of other formats of belief).[3] John McDowell (1996) extends this human-centred picture to perception and experience in general, which is perhaps even more controversial since experiences are supposed to be more primitive than beliefs. In his later writings, however, he seems to become sensitive to the fact that non-human animals have mental lives with very different formats. Evidence of this is his use of the term 'world-disclosing experience' to single out human experiences, leaving room for experiences in non-human animals with the Gadamarian distinction between world and environment, a detail that has no direct bearing to the present investigation. I shall not go into the details of these arguments here. In what follows I offer *positive* reasons for thinking that animals beliefs are non-propositional.

Again, human beliefs have the form "I believe that __." One interesting feature of human belief attribution is that it is subject to what we might call the *referential opacity* constraint. Suppose that Lois believes that Superman can fly. Unbeknownst to her, Superman is in fact Clark Kent. Most people agree that it is problematic to attribute Lois the belief that "Clark Kent can fly." We can perhaps agree that in some sense she does have that belief, but to do this we need to employ a very special reading of the sentence, i.e., the so-called *transparent* use, according to which "Clark Kent" functions merely as a reference-fixer. Exactly how to understand the situation and how to solve the puzzle are huge topics in philosophy of language, but we do not need them here. What we need is the idea that this opacity *does* happen in human beliefs. One reason for opacity is that humans can think of things with many different "modes of presentation" or "guises," but we of course cannot know every possible way of talking about specific objects. With another terminology, human beliefs are *fine-grained*, at least according to a natural view in philosophy of language. It is not clear if this also applies to other animals. The ways they represent things are anchored to salient factors, for example, whether there are more dominant conspecifics around.[4] This fits well with the account that will be developed later in this section. On that account, the objects of animal beliefs are *states of affairs*; compared to propositions that are composed by Fregean modes of presentation, states of affairs are *coarse-grained* in the sense that they are composed by objects and properties and not by modes of (re) presenting them (Lurz 2011: 134).

There are also empirical reasons for doubting that other primates' beliefs are propositional. Here I focus on recent works by Matsuzawa et al. on chimpanzees' working memory (2007, 2009). Their basic finding is that although chimpanzees might have worse memories in some respects they nevertheless "have an

3 For a more detailed discussion of a similar line of thought, see Robert Lurz (2011: 131-8).

4 Here I use the term 'represent' in a deflationary way; nothing hinges on it at this point, but more on this below.

extraordinary working memory capacity for numeral recollection – better even than that of human adults tested in the same apparatus following the same procedure" (2007: 1004). I would say their description of chimpanzees' relevant capacities is too moderate; in fact, they are *much* better than both adult and young humans, as demonstrated in the above mentioned study by Matsuzawa et al..[5] In this experiment the subjects are six chimpanzees. They are all trained to use Arabic numerals, though five of them had no experience of using numerals in any task. The initial task consists in touching the numerals with the 1-2-3-4-5-6-7-8-9 sequence. Once being touched, the numeral in question will disappear, but this task is not very challenging since the rest of the numerals will *not* disappear until being touched. Although not difficult, one interesting fact is that chimpanzees can do it very quickly, much quicker than average human beings.[6] It seems that they have different ways of retaining information. But what is really crucial is the later challenging task in which all other eight numerals (2-9) are *masked* after the first one (1) is touched. The subjects need to retain all the required information for several seconds, and when subjects start to perform the task they need to touch the screen by hand, an action which is distracting. But surprisingly enough, the chimps perform the task very well.

To see the difference between chimpanzees and humans in this respect, consider the famous Sperling Paradigm (1960) with humans as subjects.[7] At the beginning of the experiments, subjects look at a blank screen with a fixation point in the middle of the screen. An array of letters constituting a grid then comes in as stimuli for 15-500 milliseconds. After the stimuli disappear, there will be a delay with a blank screen; the period of delay depends on specific experimental settings for different purposes. Then crucially, a cue tone comes in signifying which row the subjects are supposed to report (high tone for the high row, etc.). Since there is a delay between the stimuli and the cue, one might expect that the cue will not have any significant effect, since it comes too late. Surprisingly, however, subjects are actually very good at reporting the given row accurately, even though the cue comes much later than the stimuli.

Although human subjects are good at performing the task, what they can reliably report are only four letters or so. Comparing to the chimpanzee cases, it should be clear that humans' iconic working memory is much poorer. Granted, there are many differences in the two paradigms, so direct comparison of the two cases is not available. Still, it should be safe to say that the ways humans and chimpanzees store information are very different, and this is a good reason for thinking that if chimpanzees have beliefs, the formats of their beliefs should be very different. Notice how badly humans perform in the Matsuzawa task. One might want to restrict the difference to the perceptual level only, but it is a rather incredible picture that humans and chimpanzees have very different perceptual states but nevertheless their beliefs may possibly share a similar format. To be sure, this is by no means a deductive argument. What philosophy can take from empirical studies in general, I believe, is only evidence gesturing in certain directions. Now, some might argue that what is provided above is not even gesturing in the direction I am heading to. To this I reply that by itself it might not. Again, when scientists set out to identify research questions and conduct experiments, what they have in mind is very different from philosophers. For example, they do not think in terms of propositionality. By themselves empirical studies cannot replace philosophical arguments. In the present case, one might say that what were introduced above are about working memory performances, period. I do agree that the experiments themselves are about working memory and not beliefs, but if we can agree that the latter crucially rest on the former, then those experiments can be seen as indirectly relevant to our purposes. Beliefs are something we need to acquire and sustain through time which crucially involves working as well as other varieties of memories. To have a full theory of the relations between working memory and belief, of course, is beyond the scope the current investigation.

On the other hand, there are many other cognitive tasks that are easy for humans but difficult for chimpanzees. Chimpanzees are, for example, not good at "generalized imitation" (Myowa-Yamakoshi et al. 2004), "cross-modal matching" (Hashiya and Kojima 2001), "number concepts" (Wood et al. 2008), and so on. The point about number concepts is especially relevant here. Given that chimpanzees are so good at the

5 The experiment is available at the following https://www.youtube.com/watch?v=aAIGVT3N7B0.

6 See the clip from 0:35 on.

7 Sperling experiment:http://www.nyu.edu/gsas/dept/philo/faculty/block/demos/Sperling320msec.mov.

task designed by Matsuzawa et al (2007), one might expect that they and their relative primates might be able to manage number concepts to some extent. However, as shown by Woods et al., rhesus monkeys are able to compare 1 and 2, 2 and 3, 3 and 4, but profoundly unable to do 4 and 5 / 3 and 8. This suggests that their grasp of numbers is not very good. Perhaps their working memory for abstract things is incomparable with their working memory for perceptually salient items. This is the main contention of the "trade-off theory" proposed by Matsuzawa:

> At a certain point in evolution, because of limitations on brain capacity, the human brain may have acquired new functions in parallel with losing others – such as acquiring language while losing visuo-spatial temporal storage ability. (Matsuzawa 2009: 97).

And not surprisingly, number concepts go with language. How to work out the details of this trade-off story is of course difficult and controversial, but the general line seems to be promising. If so, we have further reasons to believe that beliefs in other primates are non-propositional, since the components of propositions are abstract concepts.

One might object that the above two arguments, if successful, only show that *if* animals have beliefs, those beliefs are non-propositional. But this is a big "if." Maybe what animals have are only "belief-like" states, but they are not identical to beliefs. This objection seems to be a fair one, and in response I do not propose a direct argument to the effect that what animals have, according to the above arguments, are beliefs. Rather, I propose to deflate the objection by arguing that this worry, though genuine, is orthogonal to my purposes. What is more interesting and significant, I submit, is to understand the kind of mental states and episodes non-linguistic animals possess and how those mental states and episodes cause their behaviors. Whether those mental things deserve to be called "belief" seems to be a judgment call. If one insists that only propositional states are worthy of the name, then I can concede that what animals have as argued above are only belief-like. However, I do think that a rather liberal use of the term "belief," is more fruitful. The reason is that what is crucial for belief is the assertoric force and its capacity to represent the world as being a certain way. Being propositional is our way of doing it. It does not follow that it is the *only* way of doing it.

To say primate beliefs are non-propositional invites the question about its nature – in particular, the challenge for us to characterize those beliefs in *positive* terms. Robert Lurz takes up this challenge in chapter 4 of his recent book (2011). In what follows I first briefly summarize his conception and then try to further develop the model.

Lurz credits Arthur Prior (1971) as a precursor of the conception.[8] The Prior-Lurz model begins with the observation that belief relations do not have to be analyzed as

(a) S/believes/that p,

but can be analyzed instead as

(b) S/believe that/p (2011: 133).

One initial worry is that the proposed analysis might get the structure of "that-clause" wrong. However, as I suggested at the beginning of the paper, we should not be too obsessed with the surface grammatical structure. What is important for the Prior-Lurz model is that the relatum on the right hand side is not an abstract proposition. For example, Lurz appeals to the notion of "states of affairs," possible states of the world (ibid.: 133). Although I follow the general line of the conception, I do not endorse the claim that states of affairs are "abstract entities" (ibid.). For one thing, the main virtue of states of affairs is that they are more concrete, hence it makes more sense to think that other animals can bear belief relations to them, at least on the face of it. As Lurz says, states of affairs consist of "individuals, properties, and relations." Given this, it is not clear to me that they are not concrete. To be sure, relations might themselves be abstract, but when animals perceive and think about certain relations between entities

8 It should be noted that in chapter 4 of his book Lurz explicitly states that this model is not an account about what animal beliefs are; rather, it is about what it is for animals to attribute beliefs to others. So this paper can be seen as an extension of the Prior-Lurz model. How plausible this extension depends on the following argumentations.

and their properties, those relations are not perceived or thought as abstract. So it is more reasonable to hold that states of affairs are not abstract.

A worry about states of affairs is that their individuation conditions are not clear. For example, how many states of affairs are there in a given room? Are "that box is on the table" and "that table is under the box" different states of affairs? This is the general worry underlying the Fregean slingshot argument (1892). Or as Davidson puts it, if we really want to talk about "facts" (i.e., states of affairs that obtain), we can only talk about the "Great Fact" (Davidson 1969). However, the Prior-Lurz model does not commit to any specific ontology of facts or states of affairs. The notion of states of affairs can be defined relative to contexts, including the perceiving/thinking animal. For example, we can say that a certain chimpanzee at time T is in a belief relation with a state of affairs that consists of a dominant conspecific and its distance to certain foods. Of course, theoretically there is no principled way to single out the state of affairs in question, but that is not relevant, since only a few ways, or perhaps one way can be used to explain and predict that chimpanzee's subsequent behavior, and *that* is what justifies our belief attributions in question. Other idiosyncratic ways of individuating states of affairs are only in the eyes of theorists, not on the chimpanzee's mind. This partially hinges on Daniel Dennett's intentional stance idea (1988), and some might find this assumption problematic. For the purposes here, we do not endorse Dennett's theory, which has robust ontological import. Yet, I would recommend the Dennettian stance as at least a methodological guidance for belief attribution.

What is the main merit of the Prior-Lurz account? The crucial difference between propositions and states of affairs lies in their status and structures. States of affairs, unlike propositions, do not have sentential structures, and that is why we should use them to capture beliefs in non-linguistic animals, especially other primates for the present purposes. To give it a label, states of affairs have a *situational* or *circumstantial* structure. Non-linguistic animals do not think in abstract terms, since they lack the relevant apparatus. However, they are sensitive to environmental factors so that they can survive. An analogy would be visual thinking in mathematics (e.g., Giaquinto 2011). Some procedures in mathematics are so abstract that even sophisticated and well-trained humans find them difficult. However, if they visualize some items and their relations, it becomes easier to comprehend the procedure. To be sure, visuo-spatial thinking does not *justify* certain proofs in mathematics, but it is important for *thinking*. Similarly, as we have seen in the studies conducted by Matsuzawa et al., chimpanzees have strong visual working memory, so their belief states are likely to be more visual than abstract, and the relata of their beliefs are situational states of affairs, not abstract propositions.

Another possible difference between the view suggested in this article and Lurz's version is that while on his account primate beliefs are non-representational (Lurz 2011: 134) I am hesitant about this point. Lurz is right that "states of affairs, unlike propositions, are not by their nature representational entities" (ibid.). However, we have not ruled out the possibility that primates have something representational in mind and those items *represent* states of affairs. This representational account might have the virtue of making sense of the fact that non-linguistic animals also make *mistakes* in their beliefs. Indeed, even in some propositional accounts (for a prominent example see Tye 2000), propositions are thought as being *represented*, as opposed to things *representing*. That is, humans have a belief-relation to propositions. If so, then propositions and states of affairs are both things being represented. The crucial difference between them lies in their *structures*. It is fine to have different versions of non-propositional accounts, and here I do not argue for my variant that primate beliefs are representational. The point to emphasize is that the main disagreement between the propositional account and the account based on states of affairs concerns *structures*.

Another concern about the representational account is that it over-intellectualizes non-linguistic animals. A thrust of the Davidsonian argument is that non-linguistic animals cannot have concepts of belief and objective truth. The merit of the Prior-Lurz account is precisely to avoid this over-intellectualization. Now if we think that primate beliefs are representational (though non-propositional), don't we fall prey to Davidsonian objections again? I suggest that we go back to Quine's picture, in particular his use of "assent" and "dissent" (Quine 1960): these two notions are purely behavioral and involve only observable behaviors. This pair of notions is what will help us start making progress in radical translation. Davidson's criticism

of this maneuver is that Quine's position cannot give us a systematic theory of language. Davidson's alternative is to use the idea of "holding true" to replace "assent," so that the concept of truth comes into play, and in turn brings in Tarski's recursive story (1967). However, this objection to Quine does not apply in the present context, since our subjects here are *non-linguistic* animals. Since they do not have linguistic abilities, the Davidson-Tarski theory does not apply to them. Quine's story has the merit that it does not over-intellectualize non-linguistic animals. One way to understand Davidson's objection to Quine is that Quine's picture is too primitive (i.e., "assent" as opposed to "holding true") to capture human discursive language; but if so by the same token Davidson's picture is too advanced or sophisticated to capture non-linguistic animals' beliefs.

I shall end this section by considering a general worry of the Prior-Lurz picture. Since the situational structure is very different from the propositional one, one might be skeptical about whether we are doing *belief*-attribution at all when we operate with the situational structure. One initial response is that the attribution helps us explain and predict animals' behaviors, but this is not enough since other mental states, in particular perceptual experiences, can explain and predict behaviors too. My reply to this concern is that we can see the distinction between the occurrent and the dispositional in animals. It is true that in the study of Matsuzawa et al. quoted above, chimpanzees' behaviors are immediate after they see the arrays on screens, hence it is possible that beliefs do not operate there. However, in many other cases animals obviously have dispositional mental states and they behave according to those dispositional states. For example, in various experiments in which we have multiple chimpanzees and perhaps human experimenters, the chimpanzee subjects do not *only* act on their perceptual experiences, since at many points in those experiments there are occlusions between them and their conspecifics and other relevant items (e.g., Hare et al. 2000). Therefore, we have to admit that they have perceptual *beliefs*, as opposed to perceptual *experiences* only.[9] One might object that the dispositions here might be purely behavioral, i.e., non-mental. This kind of countermove is almost always possible. I do not intend here to rule out the possibility that the relevant dispositions could be behavioral only. I have provided the case for the idea that at least some of those dispositions are mental, either perceptual or doxastic. More will be discussed on this in the next section.

3. From One's Mind to Others' Minds

Given this Prior-Lurz conception of primate beliefs, how should we proceed from here to a further question whether other primates, in particular chimpanzees, understand their conspecifics and perhaps human experimenters in terms of their understandings of others' mental states? Recall that we have confined ourselves to beliefs only; issues about other mental states, e.g., intention, perception, knowledge, have analogous but not exactly the same difficulties. In what follows I first outline the main problem briefly and provide some suggestions.

A natural starting point is the so-called "logical problem" (Lurz 2011: 25). It is logical rather than empirical, because it has an "in principle" flavor. In the case of humans, the situation is relatively simple because human subjects can give linguistic reports. Although the status of report is controversial, at least it is a *prima facie* way to know what human subjects have in mind. In the case of non-linguistic animals, however, the difficulty is more daunting, since they cannot report linguistically, and worse, we cannot give them linguistic descriptions about the experiments in question. As a result, we have to conduct experiments with behavioral cues. But if we do so, it seems impossible to tease apart behavior-reading and mind-reading

9 One important issue here is how different this picture is from the view of José Luis Bermúdez (2003). He starts with a "minimalist" account in which nothing is propositional-like, and raises various objections to it (34-63). Then he proposes a "success semantics" to strengthen the account (64-108). On the Prior-Lurz model, states of affairs are still structured in a significant sense, hence it is not clear that it will become the target of Bermúdez. However, here is not the place to go into his book-long project. What we have here should be compatible with the claim that the present account is not the whole story and various supplements are called for. What is more important for us to establish here is that propositional account does not work for the case of non-linguistic animals.

hypotheses. It seems that we will always have a "complementary behavior-reading hypothesis" (ibid.: 34). For example, when mind-reading theorists use "seeing" to capture chimpanzees' understanding of others, it seems that their behavior-reading opponents can always use "direct line of gaze" to replace the notion. That is, it seems unavoidable that every mental term has its behavioral counterpart, and current experimental protocols seem to be unable to decide between the two hypotheses.[10]

What we need are new experimental protocols that are sensitive to the differences of the two empirical hypotheses. Lurz offers three of them towards the end of his book. As he says in conclusion:

> Whether animals are capable of attributing beliefs...is very much an *open empirical* question...I have sketched three possible experimental protocols that can [potentially solve the problem]. The rest is now in the hands of empirical researchers – and, of course, the animals. (Lurz 2011: 193, my italics)

Rather than repeating Lurz's proposal here, in what follows I shall consider possible responses to the Prior-Lurz account of situational structure of primate beliefs. This is at least an indirect contribution to the debate since in the past many contributed to the debate in terms of the propositional model only. Before actually testing their theories with Lurz's protocols, it is important to see what different theorists might say about the situational model of primate beliefs. For mind-reading theorists, they might want to say that since chimpanzees have good perceptual working memory but not good at abstract processing, even if they are mind-readers to some extent, they are not able to think of others in abstract terms. For example, in cases where we have a subordinate and a dominant chimpanzee, with foods behind occlusions so that the subordinate one sees that the dominant one does not have direct line of gaze to the foods in question, one assumption could be that the subordinate might hold a belief relation regarding the situation in a possibly pictorial way, where the situation includes the dominant chimpanzee's mental states. On the other hand, the behavior-reading theorists might suggest that given the good perceptual working memory, it makes sense to say that chimpanzees can detect complex behavioral cues from others and then store those pieces of information in mind. The crucial thing for the behavior-reading theorists is to insist that *only* behavioral terms are involved in the interactions. As said above, the controversy needs to be solved by running new experimental protocols; before that, one interim conclusion can be drawn which is that both mind-reading and behavior-reading hypotheses could at a preliminary level accommodate the Prior-Lurz model of primate beliefs. More details need to be added once new experiments based on new protocols are done.

I now turn to consider a more radical position in conclusion of this paper. The position that the dichotomy between mind-reading and behavior-reading hypotheses, though seemingly intuitive, does not really make empirical sense.[11] The distinction is intuitively acceptable since one side is based on both mental terms and behavioral terms while the other is based on behavioral terms only. Povinelli's challenge is that although the two hypotheses are clearly distinct, no extant experimental paradigm can tease them apart. But the present skepticism attempts to go in the other direction: since no extant experimental paradigm can favor one hypothesis over the other, we should cast the very distinction into doubt. Perhaps the whole debate of animal mindreading is misguided by the distinction. The relevant discussions are substantial of course, but according to this position, we should conceptualize it in some other ways, if the distinction between the two hypotheses does not make good empirical sense.

It is interesting to try to turn the whole discussion around when we find ourselves stuck in quandaries for several decades. However, I do not find this particular proposal promising. First of all, there is a difference between relying on intuitions and on conceptual knowledge. Philosophers sometimes rely on intuitions in thinking about thought experiments and theories, but in the present case I submit that people accept the distinction between mind-reading and behavior-reading hypotheses *not* because of intuitions but because of *semantic differences*. One allows both behavior *and* mental terms but the other allows behavior terms only, period. This is not a use of intuition; this is our grasp of word meanings, including conjunction. To

10 The logical problem is often also referred as "Povinelli's challenge" (e.g., see his 1996). For detailed discussion of this problem, see Lurz (2011: ch.2), Cecilia Heyes (1998), and Call & Tomasello (2008).

11 This position was defended by Kyle Ferguson in personal communication. In what follows I will characterize the position in my own terms. Needless to say, he is not responsible for anything I say here.

account for our semantic grasping is difficult in philosophy of language but what we need here is not an account of it. What we need is only the relevant grasp of concepts invoked in those hypotheses. It is true that sometimes when theories become too complicated we lose our clear grasps of them and it might turn out that under careful reconstruction some theories are equivalent to each other. But this is not the case in this context, since both hypotheses are clear and simple enough. To insist that we have to operationalize first in order to maintain certain distinctions exemplifies an ideal but not realistic picture of science. Or even worse, it might embody a form of verificationism. Human capacities are quite limited, hence to think that nature's joints should be carved in terms of our operationalization abilities is to overestimate our position in nature. Operationalization should be a regulative ideal, not a starting point.

This completes my investigation in this paper. As a philosopher I do believe that cooperation between scientists and philosophers should be a standard enterprise and not a rare instance. Lurz has recently contributed to this interdisciplinary project to a great extent from the philosophical side; I hope in this paper I have made a rather humble but still visible contribution in the same vein.

References

Bermúdez, J. (2003). *Thinking without words*. Oxford: Oxford University Press.

Call, J. and Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later," *Trends in Cognitive Sciences*, 12(5), 187-192.

Davidson, D. (1967). Truth and meaning. Reprinted in his *Inquiries in truth and interpretation* (2001). Oxford: Oxford University Press.

Davidson, D. (1969). True to the facts. Reprinted in his *Inquiries in truth and interpretation* (2001). Oxford: Oxford University Press.

Davidson, D. (1975). Thought and talk. Reprinted in his *Inquiries in truth and interpretation* (2001). Oxford: Oxford University Press.

Davidson, D. (1982). Rational animals. Reprinted in his *Subjective, intersubjective, objective* (2001). Oxford: Oxford University Press.

Dennett, D. (1988). *The intentional stance*. Cambridge, MA: MIT Press.

Fodor, J. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.

Frege, G. (1892). On sense and reference. Reprinted in M. Beaney (ed.), *The Frege reader* (1997). Oxford: Wiley-Blackwell.

Giaquinto, M. (2011). *Visual thinking in mathematics: an epistemological study*. Oxford: Oxford University Press.

Hare, B. et al. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behavior*, 59, 771-785.

Hashiya, K. and Kojima, S. (2001). Acquisition of auditory-visual intermodal matching-to-sample by a chimpanzee (Pan troglodytes): comparison with visual-visual intramodal matching. *Animal Cognition*, 4, 231-239.

Heyes, C. (1998). Theory of mind in nonhuman primates. *Behavioral and Brain Sciences*, 21, 101-148.

Inoue, S. and Matsuzawa, T. (2007). Working memory of numerals in chimpanzees. *Current Biology*, 17(23), 1004-1005.

Jamieson, D. (2009). What do animals think? In R. Lurz (ed.), *The philosophy of animal minds*. Cambridge, UK: Cambridge University Press.

Lurz, R. (2011). *Mindreading animals*. Cambridge, MA: MIT Press.

McDowell, J. (1996). *Mind and world*. Cambridge, MA: Harvard University Press.

McDowell, J. (2007). What myth? Reprinted in his *The engaged intellect* (2009). Cambridge, MA: Harvard University Press.

Matsuzawa, T. (2009). Symbolic representation of number in chimpanzees. *Current Opinion in Neurobiology*, 19, 92-98.

Myowa-Yamakoshi, M. and Matsuzawa T. (1999). Factors influencing imitation of manipulatory actions in chimpanzees. *Journal of Comparative Psychology*, 113, 128-136.

Myowa-Yamakoshi, M. et al. (2004). Imitation in neonatal chimpanzees (Pan troglodytes). *Developmental Science*, 7, 437-442.

Povinelli, D. (1996). Chimpanzee theory of mind? The long road to strong inference. In P. Carruthers and P. Smith (eds.). *Theories of theories of mind*. Cambridge, UK: Cambridge University Press.

Premack, D. and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1, 515-526.

Prior, A. (1971). *Objects of thought*. Oxford: Oxford University Press.

Quine, W. V. (1960). *Word and object*. Cambridge, MA: MIT Press.

Schiffer, S. (2003). *The things we mean*. Oxford: Oxford University Press.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, 74(11), 1-29.

Tye, M. (2000). Consciousness, color, and content. Cambridge, MA: MIT Press.

Wood, J. et al. (2008). Free-ranging rhesus monkeys spontaneously individuate and enumerate small numbers of non-solid portions. *Cognition*, 106, 207-221.

Book Review **Open Access**

David Markwell

Moral Emotions: Reclaiming the Evidence of the Heart. By Anthony J. Steinbock

DOI 10.1515/pipjp-2016-0007

Moral Emotions: Reclaiming the Evidence of the Heart
By Anthony J. Steinbock
Northwestern University Press, 2014. Pp. xii + 339 ISBN: 9789810129566
Pbk $34.95

Anthony J. Steinbock's *Moral Emotions: Reclaiming the Evidence of the Heart*, published 2014 by Northwestern University attempts to construct an argument for the prominence of the emotions in intersubjective experience. Steinbock's stated aim is twofold: offer a more robust account of the human subject in which that subject is not hidebound to its 'perceptual and judicative dimensions, broadly defined (3),' and leverage this account of human subjectivity and the moral emotions to offer a possible solution to the 'problems associated with "modernity" and those encountered at the impasse of postmodernity (3).' He attempts this through a phenomenological approach to the explication of the moral emotions while using 'ordinary language philosophy' to provide him with 'a "leading clue" to the critical elucidation of the phenomena [that is, the moral emotions], (23).'

For Steinbock, the moral emotions are a distinct set of emotions which are fundamentally interpersonal. The moral emotions; pride, shame, guilt, repentance, hope, despair, trust, love, and humility are all 'essentially interpersonal (12)' and take place within the 'interpersonal nexus (12)' of experience. These moral emotions have a normative bearing on our 'moral praxis (13)' due to their arising from, and being lived within our interpersonal experience. 'The legitimacy of norms that emerge (e.g., through shame, guilt, trust, loving, etc.) originate from the experience in which they are given and the interpersonal register in which they are lived (13).'

Much in Steinbock's account warrants praise. His emphasis on, and attempt to reintroduce the emotions into the constitution of the subject, and the bearing this has on the subject as fundamentally interpersonal is welcome and an area where philosophy, specifically phenomenology, can continue to develop. His conception of the interpersonal and his interpersonal nexus in conjunction with our moral practice develops a thought-provoking basis from which to carve out an ethical position. He argues for a thin and nuanced conception of normativity; one that arises contingently out of the lived experience of the moral emotions and that is both constituted and regulated by them (that is, the moral emotions), without advocating moral prescriptivism. Steinbock's alignment of phenomenology and ordinary language philosophy is fitting given his subject matter and the personal nature of their experience, as well as respecting phenomenology's commitment to the evidentiary nature of first and second person descriptions.

Steinbock's distinction between the moral non-moral emotions and each one's respective relationship to the interpersonal nature of experience is central to his argument concerning the robustness of human subjectivity and intersubjectivity. However, it is here that his account begins to suffer. Steinbock intends this distinction to be extremely fine-grained, however he provides little argumentation or explication for how the moral emotions are necessarily interpersonal whereas the non-moral emotions are not. He excludes emotions such as disgust and wonder from the moral emotion because they 'may exclusively relate to an object without

***Corresponding author: David Markwell,** University College Dublin

an interpersonal dimension ever coming into play (14).' However, this is in direct contradiction to what Steinbock later states about love, which, for him, *does* number amongst the moral emotions. He says '[loving] can be oriented toward anything, from persons, to ideas, to a utensil, to the natural environment; we can love wisdom, beauty, truth, and we are oriented toward, and by, the love or hatred of these matters that exhibit these values (224).' For Steinbock, love, as a moral emotion, is fundamentally interpersonal no matter what one feels love towards (even if this happens to be an idea, object, or an environment), whereas disgust and wonder are not. He fails to fully explicate how these two types of emotions are distinct in relation to the interpersonal sphere. Due to the irrevocably interpersonal nature of experience I would argue that any emotion, including disgust and wonder, even towards an object, would necessarily have an interpersonal adumbration. This is not to say that these emotions need necessarily count as moral emotions or that Steinbock's account of moral emotions is invalid or inefficacious in shaping our practices in the moral sphere. Rather, Steinbock's distinction between the moral and non-moral emotions and each of their relationships to the interpersonal sphere of experience needs to be made sharper for the account to work properly.

This lack of clarity with regards to definitions appears again in Steinbock's discussion of guilt and his claim that there are three distinct types of demands that weigh on the subject; the epistemic, the aesthetic, and the moral (111-13). Steinbock again attempts to draw this distinction along the lines of the interpersonal which is so central to his argument. For him epistemic and aesthetic demands, while qualitatively distinct from each other, both rest on being context dependent (111 & 112). This, in and of itself, seems innocuous enough, except for the fact that when Steinbock turns towards an explication of the moral demands he states that what distinguishes the epistemic and the aesthetic demands from moral demands is that the latter 'occurs within an interpersonal framework (113),' instead of just being context dependent. Nowhere does Steinbock spell out explicitly what the difference between a context and an interpersonal framework is, instead relying on the reader's intuition to make this distinction for him. He separates these three types of demands without clearly explicating or arguing for the differences between a context and an interpersonal framework, while holding that all experience occurs as part of and within an interpersonal nexus. This is not to say that a context and an interpersonal framework are reducible to each other, but Steinbock is not clear on what exactly distinguishes these two, nor does he discuss the connections between the epistemic, aesthetic, and moral demands. A fuller account of all three of these demands, their connections, interconnections, and how they contribute to the constitution of the self, understood intersubjectively, is called for but not provided in Steinbock's text.

Overall Steinbock's book is well researched; drawing heavily on the phenomenological tradition, psychoanalysis, and the Abrahamic monotheistic traditions. The book's primary phenomenological inspirations are Husserl, Levinas, and Max Scheler; who receive support from some, perhaps, lesser well known thinkers such as Vladimir Jankélévitch and Eugène Minkowski, as well as contemporary researcher Dan Zahavi. Steinbock also draws on research conducted at the Phenomenology Research Center at Southern Illinois University, of which he is the head. A great deal of the 'ordinary language' data comes from the Research Center and one wishes that this was put to more use in the book.

Steinbock's approach to the moral emotions is one part of a larger phenomenological project he is undertaking to re-establish a more full-bodied notion of self and intersubjectivity and how these can proffer answers to the questions put to the subject by modernism and post-modernism. Whilst this in and of itself is not problematic, it becomes so in *Moral Emotions* because several key terms and concepts are introduced, but not clearly defined or argued for in the book. Terms such as 'Myself (12),' 'inter-Personal (13),' and 'vocational self (38)' are introduced and intended as ontologically distinct categories without an argument for how or why these terms should be understood differently from their mundane usage. Instead, Steinbock repeatedly refers the reader to his other works including *Home and Beyond: Generative Phenomenology after Husserl* (Northwestern University Press, 1995), *Phenomenology and Mysticism: The Verticality of Religious Experience* (Indiana University Press, 2007), and *Vocations and Exemplars: The Verticality of Moral Experience* (in preparation) not just for fuller discussions of these concepts, but for their very definitions. If Steinbock wishes these ideas to play a central role in his argument for the moral emotions and their contribution to the constitution of the self he needs to provide the definitions of, and arguments for, these concepts within this book, at least minimally.

Steinbock uses these distinct meanings of Myself, inter-Personal, and vocational self throughout the book to leverage his account of the moral emotions and their constitutive role on the subject as fundamentally intersubjective or interpersonal. It appears that Steinbock intends these terms to fulfil a similar role to the 'Other' and 'infinity' in a Levinasian sense. This leads to what is the most disconcerting aspect of Steinbock's account of the moral emotions; they appear to only make sense in, and be efficacious, within a profoundly religions framework. Steinbock's choice of emotions that count as moral emotions (pride, shame, guilt, repentance, love, etc.) all have long and storied religious histories, and are all exemplars of emotions in the Abrahamic monotheistic religious experience. Steinbock admits as such when he states:

> These moral emotions, which have emerged historically and creatively through personal as interpersonal acts, hold essentially for a set of "personal" traditions, for example, the Abrahamic tradition, Western modernity, and so. Indeed, it is also my contention that some emotions may not make sense within a different spiritual tradition, like Zen Buddhism (266-267).

Steinbock's cross-cultural critique appears to hold, what is problematic about his account is that *within* the Abrahamic monotheistic Western tradition it appears that all experiencers of the so-called moral emotions must participate in this tradition and or have a notion of 'Other' and 'infinity' similar to that of Levinas for the moral emotions to have an influence on one's moral practice. Steinbock's point about the contingency of the generation of the moral emotions is valid, and it cannot be denied that the history of the West is irrevocably structured by the Abrahamic monotheistic tradition. This does not mean that the emotions and the moral emotions can be efficacious in shaping our interpersonal lives only if one participates in the monotheistic traditions or has a notion of alterity as a form of infinite transcendence. Yet this appears to be what Steinbock's account suggests. There is no denying that the moral emotions have most certainly been generated within the Western Abrahamic monotheistic tradition. However, this does not mean that one needs to be a practicing member of one of the monotheistic faiths or have a notion of alterity as infinite transcendence for the moral emotions to have a bearing on one's understanding of the self and of intersubjectivity. I contend that emotions, both moral and non-moral, are facets of experience that *anyone* can undergo, and that they do have a bearing on one's conception of subjectivity and its relationship to intersubjectivity. If Steinbock wishes his account to answer questions raised by modernity and the impasses of post-modernity it needs to remove these religious strictures and be applicable to anyone raised in the Western metaphysical tradition.

This, however, is not a call for a cross-cultural or universal account of the moral emotions. Rather, for Steinbock's account to be properly phenomenological in relation to the moral emotions it must be able to account for the moral emotions as experienced by anyone *within* the historically contingent, experiential world he is investigating, and his account fails at this. Because of his choice of religious rhetoric and his use of the Levinasian notion of alterity as infinitely 'Other', Steinbock's account fails to capture, in any meaningful sense, the experience of the moral emotions.

This is unfortunate, as Steinbock's project of recovering a more robust notion of the self and of the irreducibly interpersonal nature of experience, and his attempt at providing a unique answer to the problems of modernity and post-modernity is vital. His attempt to bring the emotions back into our discourse on subjectivity and intersubjectivity and his thin notion of contingent, action guiding normativity that is constructed and regulated by the influences of the moral emotions are both extremely expedient concepts. Ultimately, to be able to fully defend Steinbock's view, he needs to clear up his distinctions between the moral emotions and the non-moral emotions and the connections between the two in the interpersonal sphere. In conjunction with this, more clarity about the connections and inter-connections between the epistemic, the aesthetic, and the moral demands and how all three of these demands are constitutive of the subject as intersubjective is needed. These issues arise out of an apparent contradiction in Steinbock's views regarding what does and does not count as moral emotion, and the connections between the subject and the interpersonal sphere. Steinbock also needs to offer a fuller account of how the moral emotions, although contingently generated within a history that is undeniably religious, can now be understood without an appeal to religious rhetoric or to a notion of 'otherness' and 'infinity' that many living in the post-modern era do not hold.